# The History of Information Processing

G. Jahns

Abstract. Life without information processing is virtually impossible. Every living organism has the ability to receive information, to process it and to react. Humans used technical means to communicate and to process information very early on. The development of these means first began slowly and sped up dramatically in the past few decades and now includes almost all areas of human dealings so that some scientists are already discussing the possibility that humans will become superfluous through their own development. In this chapter, the significant aspects of the development of information processing technology will be sketched and a short overview of its possible further development and its limits will be given.

**Keywords.** Information, Information processing, Numbers, Script, Computer, Processors, Memory, Networks, History.

#### 1.1.1 Introduction

Today, information processing permeates all areas of our daily life. The overview of its history can thus only be presented in a spotlight manner and should above all encourage readers to seek out more detailed literature. The Internet offers a further rich and easily accessible source of information.

Today the computer<sup>1</sup> is a synonym for information processing technology. Its development in no way followed a straight path. In today's computers different lines of development, which in the past had nothing in common, are united. Our computers originated from calculating machines. These machines originally functioned in a purely mechanical manner to link information in that the rules required and their storage were realized through one and the same components. Very early on it proved to be practical to separate these functions. While the linkage of information in the true calculator remained linked to a material realization (hardware), this was not the case for the rules (software).

<sup>&</sup>lt;sup>1</sup> The term *computer* first appeared in the literature in 1646 from Sir Thomas Browne, who meant the people who calculated the calendars to measure the passage of time. Up to the 1930s the meaning remained the same: persons with mathematical training who performed calculations in construction offices and observatories on the basis of formulas were known in Great Britain and the United States as "the computers."

Punch cards mark the beginning of the development of storage. Originally a control element for weaving looms and for automated music, they were for a short time of dominant significance for the storage of data and software in computer technology. They were soon replaced by magnetic memory technology. This technology, originally used in the record industry, was also only used for a limited time and was then replaced by other processes.

Another important development, which at the beginning had nothing to do with computer technology, was news transfer. Starting with the optical semaphores via electrical telegraphs, a worldwide telephone and news network was developed. This network was and is the prerequisite for the World Wide Web (web, or WWW) out of which a global computer with fully new qualities and totally new dimensions developed.

A further development line was wireless news transmission. For a long time, the main area of use was radio, television, and special applications. One of these special applications, namely radar technology, led to the development and the use of semi-conductors. The integrated circuits emerging from this technology are an essential part of every area of today's information processing. Apart from this, this development line melted into the previously mentioned news networks.

In contrast to the examples cited, the BUS (binary unit system) was developed for the transfer of data in computers, meaning from a central computer (or CPU, central processor unit) to the peripheral systems such as printer, memory, etc. Today it provides a basis for networks of technical systems in manufacturing, vehicles, etc. The LBS (agricultural BUS system) is one example. It conducts data transfer between very different groups of components and implements in agriculture.

The current development trends in information technology will be dealt with in subsequent chapters. Therefore, this contribution will concentrate only on the beginnings of information processing. And, finally, a perspective for further developments will be attempted.

Due to the nature of the subject, the following emphases emerge:

- observations on information and information processing,
- early aids in information processing,
- electromechanical and electronic calculators,
- the transmission and storage of information, and
- perspectives on the future development of information processing.

#### 1.1.2 Information and Information Processing

Life without information processing is not possible. Every living organism has the ability to receive information, to process it and to react with a single goal: survival. Through language, humans developed (between one and two million years ago) a unique possibility for communication, to exchange information. With the development of agriculture [1] about 13 000 years ago came high population density, with the consequent need for storage of supplies, trade, and crafts, and thus the need to administrate these factors. The development of numbers and writing must be viewed against

this background, because humans were for the first time in a position to store information and to communicate this over time and space.

# Writing and Numbers [2]

The discovery of writing without a model was apparently so difficult that it was only successful a few times in the history of mankind. It was certainly successful in about 3000 B.C. by the Babylonians and 600 B.C. by the Zapotecs in Mexico, possibly in Egypt in 3000 B.C. and in China around 1300 B.C. Naturally, the first efforts to write were based on numbers and substantives (bookkeeping for trade and administration). For many thousands of years, humans made aids for counting by notching pieces of wood and bones. These are the oldest forms of counting and sign-writing. This form can still be found in many parts of the world today. Roman numerals and old Chinese numbers are derived from this. Stones were also used for counting. For this reason the Latin word *calculus* (pebbles) is found in many languages as *Kalkül, calculus*, or *calcul*. A further, less common, form of counting and the storage of numbers is the knotting of a band.

The great achievement of the Sumerians was to print signs in clay instead of pebbles or notches. They used a wedge for one and a circle for ten, as well as an illustration of the product. In this way inventory lists for trade and taxes were made. Through the replacement of illustrations with substantives the first step in the development of writing was made.

Our current base ten system, related to our ten fingers, is in no way self-evident. In other cultures systems based on twelve, twenty, and sixty were developed. Our current degree and time division are based on the latter. Generally every natural number is suited as the basis for a counting system. Of particular significance was the development of the positional notation and the use of the zero. Both are attributed to the Babylonians. However the positional notation was apparently developed independently in several cultures. The "discovery" of the number zero is attributed to the Indians. Through this achievement the realm of natural numbers was increased.

The space value and the number zero made systematic writing and counting possible according to simple rules. With the blossoming of science in Central Europe in the 12<sup>th</sup> century, this form of presentation and the Arabic numerals expanded from Spain. However it took almost 300 years for these figures to fully replace the Roman numerals. This change caused a dramatic development in counting technologies, which serves as the basis for our current information processing.

# Concept of Information

Energy, matter, and information are the most important basic concepts in the natural sciences. In the broadest sense, the reception of information and its processing through to the generation of a reaction is "information processing." The reception takes place through receptors (sense organs/sensors) and as a rule includes preprocessing. Information processing in a narrower sense takes place via a special unit (brain/processor) and includes the linkages as well as the storage and transmission of information. The result is either sent out or causes a reaction (motor/actor).

Although the concept of *information*<sup>2</sup> is of central meaning (particularly for informatics, the science of the systematic processing of information) it is hardly specified. There is no commonly valid definition. Thus the only possibility is to characterize information by its features. Information

- can be presented via speech, signals, signs, etc.;
- can be stored, linked, transmitted, etc.;
- requires no fixed carriers, knows no original and can be copied any way;
- is without value (not to be confused with the value of its content);
- cannot age (although the content can certainly be outdated and the material carrying the information can age);
- can be combined or manipulated any way, and it is impossible to recognize this
  by the information itself, meaning that manipulations or falsifications of information or portions thereof are also information;
- serves to process information, meaning self-processing; and
- consists, as explained by Norbert Wiener, of syntax, semantic and pragmatic portions.

Signals are basic changes. A *date* is a signal presented by a sign. A *message* is a consequence of signals including their spatial and time organization. The simplest signal is an exchange between two objects, with the unit of measure being one *bit*. The creation of information is often falsely considered a sign of intelligence. In fact, intelligence is the reduction, the selection of information.

# 1.1.3 Early Aids in Information Processing [3]

If one speaks of information processing technology, one associates it today with technical aids, primarily computers. They are developed from calculation aids, which are in focus here (Figure 1 portrays this genealogy of information processing technology.) Humans began very early to create aids to process, store, and transfer information. Writing, pebbles, and knots as methods to store and transfer information have already been mentioned. The abacus and the counting table<sup>3</sup> were the first counting aids emerging from the stones or sand. The abacus has the advantage that its principle is independent of the number system. It relies on the relative position of the counting stones. It lost its significance in Europe with the change from the Roman to the Indo-Arabic counting system in the 16<sup>th</sup> century, but is still used as a systematic counting aid in many parts of the world, such as Russia and China.

With the discovery of the logarithm in 1614 by John Napier, the basis for the development of the slide rule was created. This tool was still being used until the emergence of the electronic calculator in the middle of the last century. Henry Briggs published the first logarithm table seven years later. Napier's idea, to take the already existing *Tabulae Pythagoria* from the Hindus and the Arabs and change it somewhat, cut

<sup>&</sup>lt;sup>2</sup> Information comes from the Latin *informare*, to give form, to educate someone through instruction.

<sup>&</sup>lt;sup>3</sup> Greek  $\alpha\beta\alpha\xi$  and Roman *abacus* mean table, tray, or round platter; in Semitic *abaq* means dust.

it into strips and transfer it to small sticks, was of high importance. With these "Napier's Bones" one could carry out all four types of basic calculation as well as squares and roots. Instead of calculating by ordering the "Napier's Bones" by hand, very soon refined mechanisms [3] were invented to mechanize this process, including sexagecimal systems for astronomic calculations.

Leonardo da Vinci had already developed a mechanical calculator but he never built it. The current state of knowledge indicates that the mechanical calculator by Professor Wilhelm Schickard of Tübingen was the first in the world (see Figure 2). He was inspired by Kepler's complaints about the time-consuming calculation of numbers that kept him from doing important things. Described by Schickard as a "calculating clock" and built in 1623 (the birth year of Blaise Pascal) this calculator basically consisted of two machines. Schickard used a wheelwork to add and subtract (see lower part of Figure 1) and used the "Napier's Bones" for multiplication and division (upper part of Figure 1 [4]). With Schikard's clock, it was for the first time possible to carry over tens. A bell signaled when the highest number, 999 999, was reached.

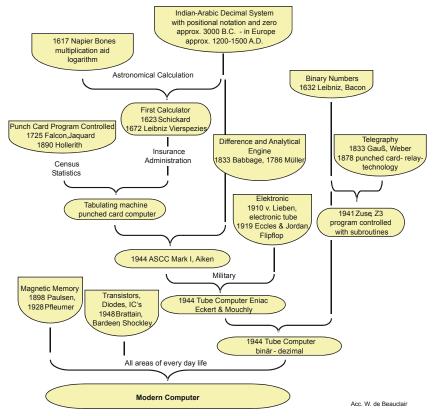


Figure 1. Genealogy of information processing technology [3].

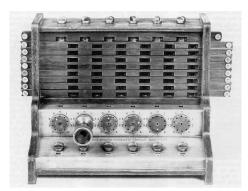


Figure 2. Replica of W. Schickard's calculator (1592-1635) [4].



Figure 3. Antikythera, ca. 80 B.C. [5].

Following Schikard, other constructive solutions for the carry-over of tens were proposed and realized. The best-known example is the eight-figure two-species calculator developed by Blaire Pascal in 1643 to ease the work of his father, a tax official. This calculator used a lever mechanism to carry out the ten carry-overs. In addition to the decade counting wheels this machine also included counting wheels for 20- and 12-divisions. Originals of the 50 machines built still exist.

A mechanical calculator with wheel works was developed well before Schikard. These astrolabes served to calculate the position of the stars and were actually more similar to a gear-work from a clock. One example of such an astronomical calculator of astonishing complexity was found in 1901 in a wrecked ship near the island Anti-kythera (see Figure 3). It was dated at about 80 B.C. [5]. Clocks as well as astrolabes are information processing machines. A presentation of their development would, however, be too extensive for this paper.

Although the art of clock making had already reached a high level of development in the 13<sup>th</sup> century, this was not advanced enough to build calculators at that time. And, thus, all inventors up into the 20<sup>th</sup> century complained about mechanical problems and a large number of projects, such as, for example, Babbage's, failed due to this problem and not because of their concept.

Gottfried Wilhelm Leibniz created a total of five different calculators beginning in 1672. These were the first functional four-species machines (Figure 4). Multiplication was carried out as continued addition and division as continued subtraction. Leibniz's calculators had a number of brilliant details that were used until the 20<sup>th</sup> century. These included an echelon roller, a sequential sled, and a spring ten carry-over. Of even more significance for the further development of information processing were Leibniz's contributions to the binary system. But Leibniz did not get to make a calculator according to this principle. This principle has only been applied since the beginning of the 20<sup>th</sup> century.



Figure 4. First functional four-species calculator built by G.W. Leibniz (1646-1716) [4].

In addition to many other inventors of calculators, the idea of the Italian Giovanni Poleni must also be mentioned, who used a sprocket wheel instead an echelon roller. Sprocket wheels and sliding rollers competed in very different creative constructions, for example in the inventions of Jacob Leupold $^4$  [6], P. M. Hahnor, J. H. Müller, and later F. Trinks, well into the  $20^{th}$  century.

The developments in the next 150 years were based to a large extent on the Leibniz ideas. Until the middle of the 19<sup>th</sup> century, only individual machines were built by inventors. The Frenchman Charles-Xavier Thomas was the first to start producing calculators on a larger scale in 1820. In 30 years, about 1500 calculators were built. It didn't take long for other European countries and the US to start building calculators. In Germany, for example, a truly functioning machine was produced on a large scale by the Brunsviga Co. in Braunschweig<sup>5</sup> at the end of the 19<sup>th</sup> century. This was based on the basic construction of the Swede Willgodt Theophil Odhner from Petersburg. Madas from Switzerland, Walther from Germany, and Burroughs from the US are other famous names of successful mechanical calculator manufacturers [4,7-10]. The era of mechanical calculators ended in the middle of the last century with the coming of the electronic table calculator.

# Program-Controlled Calculators [3]

Mathematical tables were used for astronomical and nautical calculations in the 18<sup>th</sup> century. These were calculated by hand and often faulty. The best way to create such a table without errors was to use the difference method, behind which nothing else is hidden than the fact that the n<sup>th</sup> derivation of a polynomial of order n is a constant. This basic thought was the basis underlying the "Difference Engine" by Babbage [11]. Its advantage is that only addition functions are necessary for multiplication.

<sup>&</sup>lt;sup>4</sup> Even more important than his mechanical calculator were the books he wrote between 1724 and 1728 in eight volumes with a total of 1518 pages, of which the most well known is the Theatrum Arithmetico-Geometricum [6]. Czar Peter I had it translated into Russian and it is said that James Watt learned German at an advanced age in order to read the books in their original language.

<sup>&</sup>lt;sup>5</sup> The collection of historical calculators from the Brunsviga Co. is now in the possession of the Braunschweig State Museum.

Charles Babbage (1792-1871) was far ahead of his time with his ideas and plans as he began developing the first of his three calculators in 1821. These machines had characteristics of today's computers and differ radically from all mechanical calculators so far. Thus, for example, his ideas included a central unit, a memory for interim and final results, a program control as well as extension possibilities in the form of input and output equipment such as a card reader, a card printer, copper punching equipment, etc. Unfortunately the realization of his plans were not possible due to the limited abilities of mechanical parts manufacturers of his time, not because the concept or details were wrong. It is not possible to speak about Babbage and his work without mentioning Augusta Ada Byron, later Countess of Lovelace, known as the first programmer of the world.

As so often in history, Babbage had a predecessor, without knowing it. In a book published by E. Klippstein in 1786 [12] about a newly discovered calculator, he reported on the invention of a machine using the principle of derivation, and printed the results by the engineer J.H. Müller of the Hessian Army. The machine used differences up to the third order. However, the machine was never built due to lack of funds.

#### Tabulation Machines and Punch Card Calculators

Babbage had already planned to control his calculator with punch cards. This idea to control machinery and equipment with punch cards can be traced back to the Frenchman Jean Baptiste Falcon (1728), who used small punched wooden boards to control looms. Jacque de Vaucanson (1709-1834) and Marie Jacquard (1752-1834) improved this invention by using punched metal plates, or punch band cards made of cardboard. This technology was used for weaving looms and also to a great extent for automatons, particularly musical automatons.

The breakthrough in this technology came with its use by Hermann Hollerith in the 11<sup>th</sup> US Census in 1890. Hollerith created a tabulator machine with a large number of detailed inventions comprised of punch cards, card punchers, card readers, sorting equipment, and electromagnetic counters. Through the electrical scanning of the punch card with contact pens he could process up to 1000 cards per hour. His machines were used in many European countries following their success in the US, and were continuously improved and their functions expanded by him and other inventors.

A merger of the Tabulating Machine Company for punch cards and machines, which was founded in 1896 by Hollerith, with two other companies spawned the International Business Machines Corporation (IBM) [13]. While punch cards and tapes initially were used for census and for organizational and administrational purposes, since the 1950s they were also used in the developing computer technology.

#### 1.1.4 Electromechanical and Electronic Calculators [14-17]

Mechanical calculators were only capable of carrying out the simplest of calculations. Even though their handling was constantly improved, simplified, and less sensitive to error, their functions were not basically changed. The German Hollerith Society made new contributions in 1935 with the IBM 601 and the D11, through merging of tabulating machines and calculators into punch card calculators. Around the same time that these electromechanical calculators were created, the first possibilities for elec-

tronic information processing appeared. The future of the electromechanical calculator linked to the names Konrad Zuse, Bell Telephone Laboratories, IBM, and H. A. Aiken, was limited and came to an end in the 1950s.

Motivated by boring, infinite calculations, Konrad Zuse, a civil engineer, constructed the first freely programmable computer. The fact that he had no knowledge of existing mechanical calculators turned out to be an advantage, since he took a totally new direction. He used, for example, the binary system, not the usual decimal number system, and he used a currently valid concept in his design resting on a control unit with memory and calculator. His first test model (V1, later Z1) was still purely mechanical, however it was freely programmable and could save 16 digits to 24 bits. The construction of Z1 and subsequent models took place in Zuse's private accommodations. The Z2 was originally planned to work with tubes. Since no tubes were available during the war, and Zuse's work received no further support, Zuse ultimately used relays, just as in the subsequent model Z3, which was completed in 1941. This was destroyed in an air attack on Zuse's home in 1944. The Z3 had a 64-word storage capacity; the length of a word was 22 bits: 14 for the mantissa, 7 for the exponent, and one for the sign. Zuse used a machine-internal presentation which is still in use today. The commands were read into a command register via programmed punched tapes, a punched film tape. A numerical keyboard and lamps served for input and output. The use of subprograms for recurring calculation operations was planned in his patent application (Z 26476) in 1941. The patent was denied due to lack of inventive step! The arithmetical unit consisted of an exponent and mantissa register as well as an ALU (arithmetic logic unit). The Z4, started during the war, was used until 1955 by the ETH Zurich, sold from there and used for another four years. Today it is in the collection of the German Museum in Munich.

Zuse was in no way alone with his ideas. Both Frenchmen R. L. A. Voltat (1931) and L. Couffignal (1936) used binary arithmetic. In London in 1936, E. W. Phillips presented a model with photocells that used an octal system, but was only planned for multiplication. Hungarian L. Kozma developed a calculating machine with relays, in Antwerp for Bell in 1937, which was lost in the confusion of war and forgotten. In 1939 in the US, the Bulgarian J. V. Atanasoff presented a binary prototype with vacuum tubes, which he completed in 1942 with C. E. Berry. Called the *ABC* (Atanasoff Berry Computer), it was a special computer for solving linear equations, but was not freely programmable.

The work of Georg R. Stibitz at Bell Telephone Laboratories to compute with complex numbers is also noteworthy. Bell needed calculations to develop filters and amplifiers for telecommunication technology. In the course of the war, Stibitz and Williams developed other models for military purposes of which the fifth model (1948) had essentially the same characteristics as the Z4. This further development was used for ballistic calculations. Howard A. Aiken started work in 1937 at Harvard University with the first considerations for the first mainframe computer, the Mark I, in 1944. The machine was 2.5 m in height, 14 m in width and weighed five tons. The calculation capability was impressive. The cost of half a million US dollars (two-thirds funded by IBM and one-third by the Navy) was equally impressive. Even though his

machine had little in common with Babbage's, Aiken was one of the few computer pioneers familiar with his work.

In 1938, Claude E. Shannon proved in his master's thesis that through the use of symbolic logic in relay machines, information can be processed with machines like any other material. He thus delivered the theoretical basis for that which Zuse and Stibitz had anticipated.

In contrast to other countries, the US military recognized the potential of computer technology and promoted its development accordingly. Thus contract was given to construct an "Electronic Numerical Integrator And Computer" (ENIAC) for ballistic calculations and to optimize the carpet of bombing. ENIAC was the electrical analogue to mechanical computers. It did not function in a binary manner and did not have many of the features of Zuse's and Atanasoff's computers. But through the use of tubes, ENIAC was revolutionarily fast and served as the bridge between the mechanical and electronic computers. The internal tact frequency was about 100 kHz, it had 20,000 tubes, weighed 30 tons and needed 174 kW. By the time it was completed it had cost half a million US dollars. The programming was a configuration and it took months because it took place via hardware through the re-plugging of cable linkages. For the calculation of ballistic tables this effort was only required for the initial programming; subsequently only a few switches needed to be changed.

It must also be mentioned that for ballistic calculations analogue mechanical computers were also developed. But they were soon replaced with analogue electronic computers. These were especially suited to solve differential equations. They lost their significance in the 1970s. Their use was a constant battle with corroded contacts and over amplification. Their optimal programming required extensive preparatory calculations for the correct scaling.

Before ENIAC was completed, the development of an electronic discrete variable computer (EDVC) began. This was controlled by punch cards, but had a central processor as well as program and data memory. It was used by John von Neumann to test the ability to produce the hydrogen bomb. In a publication from this time, von Neumann describes the principal design of computers and their functional components: calculator, control unit, memory, and input and output units. This *von Neumann architecture* is still valid today.

By this time, in the Z3 and in theoretical works (particularly those by von Neumann and Shannon), the step from calculator automatons to today's computers was made. What happened in the next 50 years was "only" a completion of their work, particularly through the use of new elements. Relays followed electron tubes, then semiconductors, and then integrated circuit boards. Today, these boards have become far more complex. In the framework of these developments, the pioneering individual geniuses have been replaced by staffs of engineers and natural scientists who work continuously on further development and improvement. This, for example, became apparent by the SAGE (Semi-Automatic Ground Environment) program, in which 4000 staff members participated. This group developed the Whirlwind computer, completed in 1951. It had a graphic surface that used the round cathode ray tube from radar tech-

nology, and had both computer graphics and a light pen. It served for air space monitoring.

The development of information processing in Russia and the East Bloc countries is only minimally documented and had less influence on today's information processing technology. This does not have to do with lack of genius on the part of the scientists and engineers there, but rather is the result of the economic and ideological situation during the Cold War.

#### Storage of Information

The storage of information is a key problem of information technology. Writing was the first aid created by humans to store and transfer information over space and time. In the 18<sup>th</sup> century, punch cards were used as a permanent retentive memory to control weaving looms and automatons and later in tabulator machines. In telegraphy, punch tapes were widely distributed and were used up until the middle of the last century for external storage of data and programs until they were replaced with magnetic data media. Since then, optical memories have emerged. In the multimedia and computer areas, a great need exists to store large amounts of data. The perspectives on this market support the financing of large research efforts and today intensive work is taking place to use molecular and atomic structures for memory systems.

Following the mechanical storage of information, relays and tube flip-flops were used as internal computer memory. In the search for fast memories, the acoustic running time in tubes filled with quicksilver, in wires, or in magnetostrictive material was tested. Through repeated use, the data could also be stored over a longer period of time. However, information stored in this manner could only be accessed at the end of each running time.

The first fast memory with free access was cathode ray storage. It was based on the fact that the electron ray creates not only a light spot where it touches the phosphorous surface but also deposits an electrostatic load, which lasts for about 0.2 seconds. This tube memory developed by Williams was used in many computers. Other electrostatic memories were the seletron tube by Jan Rahmann and neon lamps memory by Louis Couffignon (based on the fact that neon lamps need a high ignition current but low maintenance current). However, the practical use of this memory failed due to the various aging processes of neon lamps.

The first magnetic core memory was conceived at the end of the 1940s and was broadly used in the 1950s. Frederick Viehe is considered the inventor; he sold his successful 1947 patent application to IBM in 1956. With regard to reliability and access time, the magnetic core memory was a milestone in storage technology, but was soon replaced by semiconductor memories. The magnetic-drum memory by Billing and Booth was used above all for large amounts of data up until the mid-1900s. It had a high storage capacity but long access times. The same held true for the magnetic disk memories competing with the drum memory.

#### Man/Machine Interface

The link between humans and machines, the *man/machine interface*, has a great deal of significance in terms of efficiency and acceptance. This is true especially for inputs and outputs of computers.

At the time of an ENIAC or Mark, the programmer had to change plugs for cable connections. Later, the operators at the consoles needed to turn switches on and off, press buttons, observe displays, place punch cards in the machine and take them out—overall not a task for untrained people. The nearly graphic user interface of the Whirlwind computer was, in 1951, far ahead of its time. The military was especially interested in improving the interaction between humans and computers. In the framework of a project promoted by the US Defense Department, Douglas Engelbart applied for a patent for an X-Y position indicator for a monitor system—the mouse—in 1968. Today this is as natural as a keyboard.

Since we encounter information technology systems in all areas of life (today less than 3% of processors are used in PCs), it is necessary to adapt these systems to the abilities of humans and not the reverse. Systems and equipment that cannot be used intuitively and are not self-explanatory, but require intensive practice, are simply useless for ordinary purposes. This is not just a question of comfort, it is an indispensable quality and safety characteristic for both hardware and software.

#### 1.1.5 Information Transmission

The need to transmit information over large distances is certainly very long-standing. Fire, smoke signals, acoustic signals, runners, rider relays, and homing pigeons were the only means of transmitting information over long distances in antiquity.

# **Optical Telegraphs**

In 1836, the French abbot Claude Chappe succeeded in creating a functional optical system with semaphores (signal carriers) [18]. Chappe's semaphore or display telegraph consisted of a pole with a centrally rotatable bar, a regulator, at its ends. A short arm, the indicator, was attached at each end of the regulator. With such a display telegraph it was possible to present several hundred symbols. The Swede Abraham Niclas Edelcranz developed a similar device. It used ten hatches with which  $2^{10} = 1024$  symbols could be presented. An important prerequisite for both systems was the telescope, which was invented at the beginning of the  $17^{th}$  century. These optical systems showed characteristics that are used today: coding with code books to compress data, hand shaking, data packages, route encoding, and routines for error detection. The transfer performance was about 0.5 bits per second, or about 20 characters per minute. By 1852 the network of optical telegraphs covered 4800 km.

#### Electrical Telegraphs

As early as 1837, during the development of optical networks, the construction of the first electrical telegraph lines began. The path to electrical news transmission was a very long one. First efforts can be traced back to the Spaniard Campillo (1795). Knowledge about electricity was at the time very incomplete and so his first attempts

failed. In 1833 the first electromagnetic display telegraph was built. With electrical impulses and a mirror galvanometer, Carl Friedrich Gauss and Wilhelm Weber maintained a successful news channel for five years. With induction they steered a magnetic needle in two directions. They used a five-character code that can be re-found in the telegraph alphabet (CCITT nr. 2) and the five-hole punch or ticker tape. When the Leipzig Dresdner Railroad Company wanted to purchase such telegraphic equipment, the two scientists refused because in their opinion the electrical telegraph was not suited to practical use.

Cooke and Wheatstone successfully grasped the idea of a needle telegraph for the railroad. Brilliant improvements were made, such as those by Samuel Morse, who built a functional electromagnetic automatic writer and ultimately the Morse Code (1838), and those of Carl August von Steinheil, who used the earth as a return, made the transferred signal audible, and printed the signals. In 1844, the first commercial electrical telegraph line began in the US.

Due to a multiplicity of inventions and improvements, the development in many countries in Europe and the US was rather stormy after 1850. As early as 1852, a telegraph link between England and the Continent began operation. In 1858, a transatlantic connection succeeded for a few days; a first attempt in 1857 had failed. A permanent transatlantic connection was established in 1866. One of the most admirable technical achievements with the technology of that time was possible only with many additional inventions and technical novelties, as well as the determination of the driving force and financier Cyrus Field. The manufacture of an underwater cable was one of the greatest challenges. There was little success with hemp, wax, shellac, rubber, etc., before finally succeeding with gutta-percha for electrical insulation and water-proofing.

The telegraphs were permanently improved and thus by 1858 the automatic telegraph by Wheatstone could transfer up to 2000 characters per minute with punch tapes. In 1865 the International Telegraph Union (ITU) was founded; it became the current International Telecommunication Union. By 1901, the length of all telegraph lines was already increased to  $3.2 \times 10^6$  km, 80 times the circumference of the Earth.

#### **Telephone**

In Germany in 1870, Phillip Reis demonstrated the transfer of messages in natural speech via electric wires. Despite all of the respect that his presentation received, it was quickly dismissed as a technical toy. Three years later Alexander Graham Bell applied for his patent, just two hours before Elisha Gray. This patent and the founding of the Bell Telephone Company opened a new epoch in information transfer.

Until that time all forms of information transfer were targeted especially at trained users. The telephone made it possible for normal people to communicate over long distances for the first time. The development was stormy; as early as 1880, seven years after the original patent application, there were 30,000 telephones. The fact that cables already available from the telegraph system could be used was an advantage which should not be underestimated.

#### Wireless Transmission

The wireless transmission of information via radio telegraphs through to radio and radar technology required solid knowledge of the physical characteristics of electricity and could not be solved in a purely empirical manner. Heinrich Hertz ultimately succeeded in proving the existence of electromagnetic waves in 1887-1888, and many inventions and brilliant achievements had preceded him. As early as 1821, Thomas Seebeck was able to prove thermo-electricity. Building on this discovery, in 1826, Georg Simon Ohm discovered the law of electrical conductivity later named for him. Andre Marie Ampere discovered, in 1827, the attraction, rejection, and magnetic effects of electrical currents, explained magnetism through molecular currents, and created the first mathematically based electro-dynamic theory. In 1855, James Clerk Maxwell built on the clear concept of electrical and magnetic power lines and the proximity effects of Michael Faraday. He described the dynamic electromagnetic interaction in two partial differential equations of the second order, now called the Maxwell Equations. With proof of electromagnetic waves, a significant prerequisite for the wireless transmission was made. Building on the work of Tesla, in 1901 Guglielmo Marconi was successful in achieving a first transatlantic radio connection. With the invention of electron tubes, independently by both Robert von Lieben and Lee de Forest in 1906, the path was made free for the stormy development of wireless information transmission

#### BUS (Binary Unit System)

BUS technology holds a special place in data transmission. A BUS indicates data transmission between all units belonging to the system. Such a system can be a computer itself—early computers and the von Neumann concept comprehended such internal and peripheral units such as robots, production plants, vehicles, or tractor implement units. Generally, every unit at the BUS can address or receive information from all attached units. This requires a special administration, which is called the BUS protocol. This guaranties conflict-free BUS access, addressing, data transfer, error treatment, etc. A special BUS-controller performs this task.

At the beginning of the 80s the first electronics and microprocessors were used in tractors and implements. However, data exchange between these units was not possible (aside from the cost) because many hardware components (sensors, power supplies, cases, man/machine interfaces, etc.) were needed several times. These information islands caused functional restriction. To overcome these problems the concept of an agricultural BUS was first proposed in 1985 [19].

#### Via Radar to IC

The technologies used in radio and radar, particularly military technologies, played a very important role in the development of information processing. These uses stimulated many inventions and developments that served as the basis for today's information technology. The path leads to semiconductors and integrated circuits via radio and radar technology. In 1904, Christian Hülsmeyer had already received a patent for a process to "Remove metal objects with electrical waves and report this to an observer," called the "Tele-Mobiloscope." Industry showed no interest in this invention

and only a few scientists worked on the navigation of ships and airplanes with radio waves. In the scope of war preparations for the Second World War, the British military worked hard in the area of ship and plane navigation with radio waves. Thus, in February 1935 under the direction of Robert Wattson Watt, the first practical navigation of an airplane with radio waves was achieved. Further plans to use radio waves as an active weapon turned out to be not realizable.

In 1874, the German Karl Ferdinand Braun proved the rectifier effect of semiconductors. At the turn of the century this was used as crystal detectors in reception facilities for telegraph and radio, but these were soon replaced by electron valves. The latter were simultaneously rectifiers and amplifiers and much more reliable. However, they were too slow for radar technology with its high frequencies. Thus in the 1940s, crystal rectifiers and thus semiconductors returned to radio wave receivers. In 1947, following two years of research by Bell Laboratories, John Bardeen, together with Walter Brattain and William Shockley, succeeded in developing the first transistor: a germanium peak transistor with about 50-fold amplification. This was followed by the planar transistor. Germanium was quickly replaced with silicon, which made the manufacturing process much easier. Structures with various doping levels could be realized on one homogenous semiconductor plate. John St. Clair Kilby realized this idea. In 1958 he was able to show that it was possible to include all elements necessary for complex circuits—conductors, resistors, condensers, diodes, and transistors—on one small piece of semiconductor board. Thus the door was opened for the integrated circuit (IC), in which circuit paths and construction elements could be built on one wafer with lithographic processes.

#### Internet

The US Defense Department launched the Advanced Research Project Agency (ARPA) as an answer to the Sputnik-shock of 1958. One of its purposes was to develop a decentralized communication network, the ARPA Net. In 1972 this was presented to the public. Politicians and telephone companies showed little interest in the project. In 1973, Robert Metcalfe designed the Ethernet, with which 256 computers could be linked, within the scope of his dissertation at Harvard. In the aftermath, many networks were created that linked computers of different groups in universities, research facilities, and companies. Each of these networks was a stand-alone solution.

Ahead of this time, Ted Nelson described in 1974 a concept for non-sequential reading and writing, which he called "Hypertext." The idea became the conceptual basis for the page description language HTML (Hypertext Markup Language), which Tim Berners-Lee developed at CERN (European Organization for Nuclear Research) in 1989. It was his goal to link the different stand-alone networks together. For this he chose the name *World Wide Web*. The concept of addressability of an Internet server through the universal resource locator (URL) originated with him. Together with this address standard and the page description language HTML, the information transfer protocol HTTP (Hypertext Transfer Protocol) served as the basis for the WWW. This made it possible for text and graphic pages to appear the same regardless of the computer type of the sender and receiver. The success of the WWW rests in this universal-

ity. In 1991, CERN made the WWW software available at no cost, thus access was generally possible. It did, however, require the mastering of the protocol. This changed in 1994 with the freely available Netscape Navigator. Its functionality and its clearly designed user interface offered problem-free access to the Internet for everyone. In 1995, more than 45 million users used the Internet. In addition to functionality, this success was based on a broad distribution of computers in society and the worldwide availability of the telecommunication network.

# **Global Network Computers**

The Internet that links millions of computers is now a super computer of unimaginable computer capacity. In mid-1995, David Anderson of the University of California at Berkley developed a program<sup>6</sup> to use the computing capacity of the computers linked in the Internet at those times when they are not in use by their owners. Since the average workload of office and home computers is estimated at 10%, this presents a potential computer efficiency that places the current super computers far behind. If only 10 million computers participate, such a virtual computer would have a computing ability of a petaflop.<sup>7</sup>

# 1.1.6 Perspectives for the Future Development of Information Processing

In several passages in his book *Computers* [17], Christian Wurster showed how difficult it is to predict the future: In 1889, Charles Duell of the US Patent Agency stated his opinion that, "everything that can be invented has already been invented." In 1943, IBM estimated the world market for computers to be five computers. As recently as 1977, Kenneth Olsen of DEC stated that there is no reason for anyone in the world to have a computer in his or her home. Less than ten years later, the number of home computers reached into the millions. It was therefore all the more astonishing that Gordon Moore, one of the co-founders of Intel, in 1965 made the projection that computational ability (i.e., data density, originally the density of transistors) would double every 18 months, which has held true to date. A continuation of Moore's Law raises two questions:

- How long can an increase in capacity be continued; what are the physical limits?
- What does such a performance increase mean; what consequences can be expected for the mankind?

# Ultimate Physical Limits to Computation

An article by Seth Lloyd appeared under this title in the magazine *Nature* in 2000 [20]. In his considerations, he assumed that the computer is a physical system and thus that laws of physics determine its performance limits. He concentrated on the three basic problems of information processing, namely storage, processing, and the transfer

<sup>&</sup>lt;sup>6</sup> The program SETI (Search for Extraterrestrial Intelligence) was written to search for extraterrestrial life. In this task about 50 Gbyte data are studied daily for traces of suspicious communications.

<sup>7 10&</sup>lt;sup>15</sup> flops (floating point operations per second)

of information. His *ultimate laptop* has of one kilogram of material in a volume of one liter. Lloyd derived, from the Heisenberg uncertainty principle between energy and time, that calculation speed grows with available energy. Through the conversion of the entire kilogram of material into energy, a computational speed of  $5.4 \times 10^{50}$  operations per second would be possible. That would mean a tremendous increase in the current approximately  $10^{10}$  computational operations per second.

The second most important factor, after speed, is the memory capacity of the computer. Here he turned to thermodynamics because entropy is a limiting factor. The *ultimate laptop* comprising more or less only of rays of approximately  $10^8$  K would offer a storage capacity of  $10^{31}$  bits, in contrast to the  $10^{10}$  bits of today's laptops.

The possible consequences of ever more capable computers and robots are no longer just the fantasy of science fiction writers, but also the subject of serious scientific considerations. In April 2000, Bill Joy, the chief thinker for Sun Microsystems, caused an ethical debate on the responsibility of scientists and the future of mankind with his manifesto "Why the future doesn't need us" [21] to which many famous scientists contributed [22]. These included some who were previously critical of the future development of technology and its possible consequences, including above all Ray Kurzweil [23], Hans Moravec, Joseph Weizenbaum and many others. Hans Moravec, a scientist famous in the field of robot technology, stated that robots will be cleverer than humans in 30 to 40 years and speaks of a *post-biological* period. Joseph Weizenbaum sees no limits to how intelligent computers can be made, but in his opinion this intelligence differs from human intelligence.

# References<sup>8</sup>

- 1. Diamond, J. 1997. *Guns, Germs, and Steel: The Fate of Human Society.* New York, NY: W. W. Norton and Company.
- 2. Ifrath, G. 1981. Histoire Universelle des Chiffres. Paris, France: Seghers.
- 3. de Beauclair, W. 1968. *Rechnen mit Maschinen*. Braunschweig, Germany: Vieweg & Sohn.
- 4. Biegel, G. 1992. Braunschweigisches Landesmuseum. Informationen und Berichte, Nr.1.
- 5. Price, D. de Solla. 1974. Gears from the Greeks—The Antikythera Mechanism—

It is a weakness of the Internet that the Internet addresses, the URLs (uniform resource locators), are frequently valid only for a short time. They are often changed or deleted. Although the cited URLs have been tested before printing, we cannot foresee how long they will be available. Therefore, important scientific works published on the Internet must be permanently archived in libraries. For example, the CIGR EJournal is placed on CD and archived in several key libraries around the world. However, one strength of the Internet is that new references appear continuously. Therefore, it is worthwhile to search the Internet for topics of interest.

Another benefit is that you can find publications in your preferred language.

- A calendar computer from ca. 80 B.C. *Transactions of the American Philosophical Society* 64(7).
- 6. Leupold, J. 1727. *Theatrum Arithmetico-Geometricum das ist: Schauplatz der Rechen- und Meβ-Kunst*. Leipzig, Germany: Christoph Zunkel.
- 7. Mackay, R. 1997. The Walther Company—Historic recollection. Available at: http://www.xnumber.com/xnumber/walther.htm.
- 8. Hancock, M. 1998. Burroughs Adding Machine Company. Available at: http://www.xnumber.com/xnumber/hancock7.htm..
- 9. Mackay, R. 1998. Recollections on the Madas calculator. Available at: http://www.xnumber.com/xnumber/madas.htm..
- 10. Redin, J. 2001. A brief history of mechanical calculators.

  Part I available at: http://www.xnumber.com/xnumber/mechanical1.htm.

  Part II available at: http://www.xnumber.com/xnumber/mechanical2.htm.

  Part III available at: http://www.xnumber.com/xnumber/mechanical1.htm.
- 11. Dotzler, B., ed. 1996. Babbages Rechen-Automate. Springer.
- 12. Klippstein, P. E. 1786. J. H. Müller's Beschreibung seiner neu erfundenen Rechenmaschine, nach ihrer Gestalt, ihrem Gebrauch und Nutzen. Frankfurt und Mainz, Germany: Varrentrapp Sohn und Wenner. Available at: http://www.fbi.h-da.de/fileadmin/personal/h.schneider/public\_html/ba-wp-hi/arbeitsblaetter/mue ller-trans.pdf.
- 13. Winterstein, S. 1998. Von Hollerith zu IBM—Die geschichte der kommerziellen datenverarbeitung. Available at: http://stefan-winterstein.de/papers/hollerith-ibm/
- 14. Augarten, S. 1984. *BIT by BIT—An Illustrated History of Computers*. London, UK: Georg Allen & Unwin.
- 15. Brader, M., ed. 1994. A chronology of digital computing machines (to 1952). Available at: http://www.davros.org/misc/chronology.html
- 16. Williams, M. R. 1985. *A History of Computer Technology*. Englewood Cliffs, NJ: Prentice-Hall.
- 17. Wurster, C. 2001. Computers—Eine Illustrierte Geschichte. Köln, Germany: Taschen.
- 18. Holzmann, G. J., and B. Pehrson. 1995. *The Early History of Data Networks*. Los Alamitos, CA: IEEE Computer Society Press.
- Jahns, G., and H. Speckmann. 1984. Agricultural electronics on farm machinery needs standardized data transfer—A concept. ASAE Paper No. 84-1633. St Joseph, MI: ASAE.
- Lloyd, S. 2000. Ultimate physical limits to computation. Nature 406(8): 1047-1054
- 21. Joy, B. 2000. Why the future doesn't need us. Available at: http://www.wired.com/wired/archive/8.04/joy pr.html.
- 22. Heron, K. 2000. Rants & Raves: Why the future doesn't need us (Reactions on Bill Joy's article [21]). Available at: http://www.wired.com/wired/archive/8.07/rants\_pr.html.
- 23. Kurzweil, R. 1999. The Age of Spiritual Machines. London, UK: Viking Press.

# Hardware

# 2.1 Topics in Hardware Evolution

A. M. Saraiva, A. R. Hirakawa, and C. E. Cugnasca

Abstract. The astonishing evolution of computer systems in the last decades has two equally important components, hardware and software. Hardware has evolved dramatically and steadily, increasing in complexity, performance, quality, and reliability; decreasing in price; and spreading electronics applications to almost everything, including agriculture. Many technologies, components, and concepts that are important for the present and future of computer hardware have appeared and/or evolved. A few points related to these issues are presented on this section, which covers microelectronics trends, removable storage devices, dynamic RAM, display technologies, and new trends in computers applications, particularly the concept of "computers everywhere." It is a fascinating world of which this is a tiny sample.

**Keywords.** Microelectronics, Spintronics, Quantum computing, Removable storage devices, Dynamic RAM, Computer memory, Display, Pervasive computer, Wearable computer.

#### 2.1.1 Introduction

The changes that have occurred in microelectronics have paved the way to the overall development of information and communication technologies that have changed the world's face in a few decades. Earlier technologies never evolved as quickly have electronics. This frenetic pace, which makes things become obsolete in a few years if not months, involves a multi-billion dollar market worldwide and a variety of players: big and small companies, research centers, engineers, physicists, and others. This section aims at providing an overview of some topics on hardware evolution, past, present, and future.

# 2.1.2 Microelectronics Technology Evolution and Trends

Since the invention of the solid-state transistor at Bell Labs in 1947, much of the concepts and applications of electronics have changed. The development of new devices allowed the emergence of minicomputers in the 1960s, TV games, calculators, watches, automotive engine controls, personal computers and peripherals in the 1970s, and much more since then. The progress of technology ran at full power. The diffusion of junctions, silicon dioxides and wafers came in the 1950s. After that, the appearance

20 Chapter 2 Hardware

of field effect transistors (FET), metal-oxide-semiconductors (MOSFET), and large scale integration (LSI) came as natural progress [1].

The other driving force for progress was the evolution of production technology and the application of automation concepts, which increased productivity and reduced costs. The number of components per chip has dramatically increased and the size of the individual devices has decreased.

#### Moore's Law

In 1965, Gordon E. Moore speculated that by 1975 it would be possible to squeeze as many as 65,000 components on a single silicon chip occupying an area of only about one-fourth of a square inch. His reasoning was a log-linear relationship between device complexity and time, which showed that the relation of the computational power available at a particular price doubled every 18 months [2]. This was an empirical assertion, although it was surprisingly based on only three data points. He redrew the plot from 1975 onwards with a less-steep slope reflecting a slowdown in rate, but still behaving in a log-linear fashion. Shortly thereafter, someone (not Moore), named this curve *Moore's Law*. Officially, Moore's Law states that circuit density or capacity of semiconductors doubles every 18 months or quadruples every three years. It even appears in mathematical form:

Circuits per chip on a given year =  $2^{(\text{year-}1975)/1.5} \times (\text{circuits per chip in } 1975)$ 

In this period CMOS circuits have taken the biggest share and will continue to increase it in ASIC (application specific integrated circuits) and ASSP (application specific standard parts) applications because they fully benefit from the progress made in the mainstream digital area in terms of integration density, performance, maturity for production, and cost optimization. The integration density now available with CMOS allows to develop complete systems-on-a-chip including highly precise analogue blocks, RF front-ends, and sensors.

In the future, deep-submicron effects and the limitation on signal swing for analogue circuits due to reduced supply voltages will result in the use of technologies some generations behind the leading edge. This trend is not only due to technical reasons, but also due to increasing costs. On the economic side, before the end of the decade it is probable that transition to latest generation technology will no longer be commercially attractive for many applications.

# **Spintronics**

Spintronics, or spin electronics, refers to the study of the role played by the electron (and more generally nuclear) spin in solid-state physics, and possible devices that specifically exploit spin properties instead of or in addition to charge degrees of freedom. Spin relaxation and spin transport in metals and semiconductors are of fundamental research interest. The prototype device that is already in use in industry as a read head and a memory-storage cell is the *giant magnetoresistive* (GMR) sandwich structure, which consists of alternating ferromagnetic and nonmagnetic metal layers. Depending on the relative orientation of the magnetizations in the magnetic layers, the device resistance changes from small (parallel magnetizations) to large (antiparallel magneti-

zations). This change in resistance (also called *magnetoresistance*) is used to sense changes in magnetic fields. Recent efforts in GMR technology have also involved magnetic tunnel junction devices where the tunneling current depends on spin orientations of the electrodes.

Current efforts in designing and manufacturing spintronic devices involve two different approaches. The first is perfecting the existing GMR-based technology by either developing new materials with larger spin polarization of electrons or making improvements or variations in the existing devices that allow for better spin filtering. The second effort, which is more radical, focuses on finding novel ways of both generation and utilization of spin-polarized currents. These include investigation of spin transport in semiconductors and looking for ways in which semiconductors can function as spin polarizers and spin valves. The importance of this effort lies in the fact that the existing metal-based devices do not amplify signals (although they are successful switches or valves), whereas semiconductor-based spintronic devices could, in principle, provide amplification and serve, in general, as multi-functional devices. In addition to the near-term studies of various spin transistors and spin transport properties of semiconductors, a long-term and ambitious subfield of spintronics is the application of electron and nuclear spins to quantum information processing and quantum computation. It has long been pointed out that quantum mechanics may provide great advantages over classical physics in physical computation. However, the real boom started after the advent of Shor's factorization algorithm and quantum error correction schemes [9].

# Quantum Computing

Quantum computing is a new computing paradigm in which information is assigned to a quantum particle. The fundamental unit of quantum information is called a *quantum bit* or *qubit*. Considering the spin of an electron, a property that can be imagined as the spin of a top with its axis pointing either up or down, the up or down spin can correspond to 0 or 1. A qubit can be both 0 and 1 at the same time by the ghostly dual existence, known as a superposition of states, which depends on how an electron is placed. Carrying out a calculation using the electron, and performing it simultaneously on both 0 and 1, allows two calculations for the price of one. With more qubits, it could be seen that the representing power increases exponentially.

In order to perform computations with electrons, it is necessary to implement logic operations. One of the most important elements is the controlled NOT gate, similar to a controllable inverter. In such an element, the state of one qubit determines whether the final state of a second qubit will be inverted by a series of RF pulses [4].

There are many efforts to develop all bases of the new paradigm but considerable progress will be required before anything concrete becomes possible.

# 2.1.3 Removable Storage Devices

In the last few years, a whole new class of storage devices has evolved and become very popular, almost a household item: removable storage drives. They have become more necessary not only to expand hard disk storage capacity, but for backing up the ever-growing amount of multimedia data stored. Transporting data between two computers, sharing data between users, storing software and information that is seldom

accessed, and securing classified information also demands new technologies for storing data. Removable storage devices and their media fall into four broad categories: magnetic, optical, magneto-optical (MO), and solid-state. Tape, floppy, and hard drives use magnetic fields to store data. Optical drives like CDs (compact disc) and DVDs (digital video/versatile disc) use a laser to read and record information. Magneto-optical devices are a hybrid incorporating both magnetic and optical technologies. Flash memory cards use solid-state technology. Due to the variety of details of the underlying technologies, interfaces, capacities, and products, a complete look at all of them is far beyond the scope of this section, but an overview of the main characteristics of those four categories is presented.

#### Magnetic Storage Technology

The largest category of storage devices uses magnetic media, including floppy and superfloppy disks, hard disks, and tape. The disk or tape has a micro-thin layer of magnetic particles on its surface, which can be polarized in two modes, north and south (0 and 1), by a magnetic field, which is created by the read/write inductive head of the drive. They can be erased and reused many times, and are reasonably inexpensive and easy to use.

# Magneto-Optical Technology

Magneto-optical technology drives employ a laser to read data on the disk, while additionally needing a magnetic field to write data. A disk will be exposed to a magnet on the label side and to the light (laser beam) on the opposite side. The disks have a special alloy layer with the property of reflecting laser light at slightly different angles depending on which way it is magnetized, and data can be stored on it as north and south magnetic spots.

Magnetization occurs only at a temperature around 200°C and that is achieved with a laser beam. Once heated, the magnetic particles can be polarized by a magnetic field generated by the read/write head to store 0 or 1. As the laser is very accurate, only specific regions of magnetic particles are heated, which allows higher information density, more than other pure magnetic devices. Information is read using a less powerful laser, making use of the Kerr effect, in which the polarity of the reflected light is altered depending on the orientation of the magnetic particles. This fairly cheap media is extremely stable at room temperature, since polarization only occurs at high temperatures, and is rated with an average life of 30 years, far more than magnetic media. However, its performance tends to be poorer.

Another MO technology is *LIMDOW* (light intensity modulated direct overwrite) which, instead of using a magnetic head in the drive to make the changes, has the magnets built into the disk itself. So, the disk has two magnetic layers just behind the reflective writing surface. Depending on the temperature to which the write surface is heated it can take magnetism from one of those magnetic layers. The write process occurs in a single step, thus decreasing write times and improving performance. MO disks are not proprietary, being available from many different storage media manufacturers, in several different capacities and at competitive costs.

Optical super density (OSD) technology is a new technology that aims at developing a high capacity (40 GB or more) removable MO drive. It is based on a few innovative technologies which allow: the lens to be positioned much closer to the recording surface; the use of a higher numerical aperture lens, resulting in much higher data densities; the use of independent read/write heads on both sides of the media allowing access to both sides of the disk simultaneously, doubling the capacity and allowing data rates comparable with hard disk products.

# Magnetoresistive (MR) Technology

In a *magnetoresistive (MR)* drive the read head has a thin strip of material that changes its resistance in the presence of magnetic fields. The head write element, a tiny electro-magnet, works in the conventional fashion by magnetizing the particles coating the disc. This task division makes the heads capable of reading and writing at higher data densities and is expected to allow increasing areal densities.

#### Optical Technology

Optical technology has the *compact disc (CD)* as the most popular example. It can store 783 MB on a small surface that is very inexpensive to manufacture. The CD surface is basically an aluminum mirror with billions of tiny bumps that represent the information. In order to read that information, a precise laser beam is used. It is reflected at the disk surface and read by an optoelectronic sensor, which detects the changes in reflectivity due to the bumps. The mirror is built with an aluminum layer that coats an injection-molded piece of clear plastic. However, the conventional CD is not a removable storage, since it cannot be written by the end user.

In a *CD-recordable (CD-R)* the aluminum layer of a normal CD is replaced with an organic dye compound, which is normally reflective, but when the laser focuses on a spot and heats it to a certain temperature, the dye is darkened or burned. For the CD reader, each burned spot works as the bumps on a normal CD. It does not allow rewriting the data, since the dye cannot be changed back to its reflective state. *CD-rewritables (CD-RWs)* overcome this problem by using the phase change characteristic of a particular metal compound that changes its reflectivity depending on whether it is crystallized or not. After being heated to a certain temperature, it crystallizes as it cools and becomes very reflective, but when heated to a higher temperature, it does not crystallize when it cools and loses its reflectivity. MiniDiscs and DVDs employ similar approaches.

The *blue laser* represents a chance for achieving a significant increase in capacity in the near future with optical storage technology. Its shorter wavelength allows focusing smaller areas, and so putting more information in the same area. However, until now, some of the problems encountered are related to the power, the size, and the price of laser required. *Blu-ray disc* is a proposed format expected to allow capacities of up to 50 GB on dual-layer discs.

Fluorescent discs use revolutionary technology that promises to deliver capacities of 140 GB and above on a single multilayer disc. On conventional optical disc drive technology, signal quality degrades rapidly with the number of recording layers, mainly because of optical interference between the probing laser beam and the re-

flected signal, which are of the same wavelength, and because of the nature of the highly coherent reflected signal used. As a result, it is only possible to have two recording layers. With the use of fluorescent readout systems, signal degradation is lower and many more layers (around 100) can be used on a CD. The key aspect is a transparent organic substance in which fluorescence can be triggered by a laser beam long enough for it to be detected by a standard photoreceiver. This allows writing information on each level of a stack of transparent layers simultaneously, resulting in very high storage capacities and very good performance. The technology will allow the development of a *ClearCard* of 16 cm<sup>2</sup> with 50 layers, which provides a capacity of 1 TB and reading speeds of the order of 1 GBps.

#### Solid State Technology

Flash memory is the solid-state technology for removable storage. It is based on a matrix of transistors which can be polarized to retain 0 or 1 even after they are detached from a power supply, i.e., it is a non-volatile memory. Because flash memories have no moving parts, they are much more rugged than other technologies, making them ideal for portable computing. They are manufactured in small packages and draw minimal power, but their cost per megabyte and the relatively small capacity (up to now) restrict their use to applications as removable storage for small devices, such as digital cameras, PDAs, pocket MP3 players, and other types of mobile computing. For these devices, it has become the de facto storage medium.

There are several types of flash memory available today for a variety of applications and devices. They were developed by different manufacturers and differ in interface (logical and physical), size, and capacity. *SmartMedia* is very thin with dimensions less than half of that of a typical business card. It is often used in digital cameras. It can also be used on computers via a PC card adapter or a floppy disk adapter, which incorporates the function of an IDE interface. It is available in sizes up to 512 MB. *CompactFlash* cards are thicker than SmartMedia, because they include the controller circuitry of the IDE interface; they are available in up to 8 GB. *Memory stick* cards are used in digital cameras, music players, and notebook computers. They are about the same size as a stick of chewing gum and are available in up to 2 GB. *DataFlash* is a flash memory packaged in a Type II PC card (PCMCIA). Its extra size allows for higher capacities of up to 1 GB. *Secure Digital (SD)* and *MultiMedia (MM)* cards are very compact (24 mm × 32 mm). They are available up to 2 GB. SD cards also allow the secure exchange of data, which makes them suitable for financial transactions and protection of copyrighted material.

Removable storage devices cover a whole spectrum in terms of capacity, performance, ruggedness, price, and compatibility. Faster than one can notice, capacity is increasing, performance is improving, and size and price per mega- (or giga-) byte is decreasing. The technologies shown above and the ones that are still to be developed will push that change.

#### 2.1.4 Dynamic RAM Memory Technologies

Memories play a very important role in a computer system. They must match processor speed in order not to decrease overall system performance that, however, depends on many other aspects. Ideally, an engineer would wish to use the greatest possible amount of the fastest memory with the lowest power consumption. However, most real computer systems, such as personal computers, use memory architecture based on a very small capacity of high speed, high cost, high power consumption static RAM memory, which will constitute the cache, and a comparatively huge amount of lower speed, lower cost, and lower power consumption *dynamic RAM (DRAM)* memory, also called *main memory*. DRAM companies are constantly trying to improve memory technologies with respect to those characteristics (capacity, speed, and power consumption).

In the next paragraphs, the main DRAM technologies available are presented, starting with the older ones for comparison. From the many different technology types of DRAM available, some are similar in their essence, but the latest present radical changes in their core. They differ mostly in how they are organized and accessed. The basic protocol for memory access implies in the processor sending an address to the memory—the address that needs to be accessed for reading or writing the data—and then performing the read or write operation.

Conventional DRAM, which has been used since the first PCs, is said to be asynchronous, because it is not synchronized to the system (PC) clock signal. The address is divided in two parts, the row address and the column address (since the memory cells are arranged in an array). A memory access is begun by sending the row address, followed by the column address; a certain period of time later the memory value appears on the data bus. There is no direct relation of these signals to the system clock. Conventional DRAM was replaced years ago by fast page mode (FPM) memory, which is slightly faster. It sends the row address just once for many accesses to memory in locations near each other, improving access time. The next generation of asynchronous DRAM is called extended data out (EDO) memory or hyper page mode DRAM. It is slightly faster than FPM memory due to modified timing circuits such that one access to the memory can begin before the last one has finished. Burst EDO (BEDO) memory is also an evolutionary improvement in conventional DRAM. BEDO combines EDO memory with pipelining technology and special latches to allow for much faster access time. Although it allows the use of much higher memory bus speeds at very little additional manufacturing cost to the producer, it has not succeeded commercially and was soon defeated by SDRAM.

The synchronous DRAM (SDRAM) has its signals synchronized to the system clock so that its timing is better controlled. It is much faster than asynchronous DRAM and can really improve the system performance. It has internal performance improvements, such as internal interleaving, which allows half the module to begin an access while the other half is still finishing the previous one. Double data rate SDRAM (DDR SDRAM or simply DDR) doubles the bandwidth of the memory, compared to SDRAM, by transferring data twice per cycle, on both the rising and falling edges of the clock signal, instead of only on one signal transition.

Direct Rambus DRAM (DRDRAM or just Rambus) is referred to as being a revolutionary design. It is based on an internal bus called the Direct Rambus Channel, a high-speed 16-bit bus running at a clock rate of 400 MHz, and data transfers on both the rising and falling edges of the clock, resulting in an effective bandwidth of 1.6 GB/s. Although it reduces the bus from its conventional 64-bit to 16-bit, the channel is now capable of running at much higher speeds increasing the overall bandwidth. It has attracted a lot of attention due to Intel's decision to use it. Besides the technical concerns about it being or not being the best solution, one problem is the need to pay licensing fees since it is proprietary.

Synchronous-Link DRAM or SLDRAM is a new technology under development by the SLDRAM Consortium to be the next generation PC memory standard. It is an evolutionary design that does not demand a radical change in architecture. It is a 64-bit bus running at a 200 MHz clock frequency. The effective speed is doubled by also making data transfers on both clock edges, yielding a bandwidth of 3.2 GB/s. It is an open standard, so no license fees are required.

#### Trends

History reminds us that the best technical solution does not always win, and commercial aspects are frequently decisive in determining the standard of the next generation. This may be now the case for DRAM technology in the next few years. Although it is going to influence the choices the user will have, it is convenient to point out that DRAM is only a part in overall computer performance. In many cases, cache memory masks the improvement in DRAM efficiency, and the system chipset restricts the types of memory that can be used in a computer. Besides that, newer technologies are often more expensive and their cost-effectiveness must be compared to older technologies that are cheaper. Often, *more* system memory is more important to performance than *better* system memory.

#### 2.1.5 Display Technology

The *cathode ray tube (CRT)* is still a formidable screen technology. It is a cheap-to-make monitor capable of excellent performance, producing stable images in true colors at high display resolutions.

However, CRTs have disadvantages such as high power consumption, large size, misfocus, misconvergence and color variations across the screen, clunky high-voltage electric circuits, and strong magnetic fields which create harmful electromagnetic radiation. Those are some arguments for research and development of new display technologies. One of them makes use of a dot array of (LEDs).

# Light-Emitting Diodes (LEDs)

The basic *light-emitting diode (LED)* is a solid-state device containing a chemical compound that emits light when energy levels change in the semiconductor diode. The specific wavelength of the light depends on the difference in energy levels as well as on the type of semiconductor material used to form the LED chip.

LEDs are highly efficient and have extremely fast switching times. They come in two varieties: devices that use inorganic light-emitting materials and those using or-

ganic light-emitting materials (OLEDs). The OLED cell structure consists of a stack of thin organic layers sandwiched between a transparent anode and a metallic cathode. When an appropriate voltage (typically a few volts) is applied to the cell, the injected positive and negative charges recombine in the emissive layer to produce light.

#### Liquid Crystal Displays (LCDs)

Since 1971, *liquid crystal displays (LCDs)* have been used in a variety of fields, including miniature televisions, digital still and video cameras, and monitors. Liquid crystals were first discovered in the late 19<sup>th</sup> century and they are almost transparent substances, exhibiting some of the properties of both solid and liquid matter. In the 1960s, it was discovered that charging liquid crystals with electricity changed their molecular alignment and, consequently, the way light passed through them. Three display techniques deriving from LCD are described below.

DSTN displays—A normal passive matrix LCD comprises a number of layers. The first is a sheet of glass coated with metal oxide. It operates as a grid of row and column electrodes, which pass the current, needed to activate the screen elements, and is called STN (super-twist nematic). DSTN (dual scan STN) is an STN with separated row and column scans, which improve its velocity. The first color LCD displays simply use additional red, green, and blue colored filters over three separate LCD elements to create a single multicolored pixel.

TFT displays—Thin film transistor (TFT) screens (also known as active matrix) use an extra matrix of transistors connected to the LCD panel—one transistor for each color of each pixel. These transistors drive the pixels, eliminating at a stroke the problems of ghosting and slow response speed. TFT screens can be made thinner, therefore lighter, and refresh rates now approach those of CRTs as the current runs about ten times faster than on a DSTN screen.

Polysilicon panels—Polysilicon (p-Si) TFT displays use a crystalline structure formed at temperatures of about 450°C. A major attraction of p-Si technology is that the increased efficiency of the transistors allows the driver circuitry and peripheral electronics to be made an integral part of the display. This considerably reduces the number of components for an individual display.

#### Plasma Displays

Plasma display panels (PDPs) can be viewed as a matrix of tiny fluorescent tubes that are controlled in a sophisticated fashion. Each pixel, or cell, comprises a small capacitor with three electrodes. An electrical discharge across the electrodes causes the rare gases that fill them to be converted to plasma form as it ionizes. Once energized, the plasma cells release ultraviolet light, which then strikes and excites red, green, and blue phosphors along the face of each pixel, causing them to glow.

Manufacturing is simpler than for LCDs and costs are similar to CRTs at the same volume. However, the ultimate limitation of the plasma screen has proved to be pixel size. At present, manufacturers cannot see how to get pixels smaller than 0.3 mm, even in the long run.

ALiS, for alternate lighting of surfaces, is a PDP that overcomes the low-resolution restrictions. The technique uses interlaced rather than progressive scans, where the sustain electrodes are arranged at identical intervals and the spaces between them are used as display lines. Therefore, the resolution can be doubled compared with a conventional display having the same number of electrodes.

The plasma addressed liquid crystal display (PALCD) is a peculiar hybrid of PDP and LCD. Rather than use the ionization effect of the contained gas for the production of an image, PALCD replaces the active matrix design of TFT LCDs with a grid of anodes and cathodes that use the plasma discharge to activate LCD screen elements.

#### Surface-Conduction Electron-Emitter Display (SED)

The *surface-conduction electron-emitter display (SED)* is a flat-panel, high-resolution display currently under development by Canon and Toshiba. It is expected to overcome plasma display technology. For a similar screen size SED consumes only about half of the power of CRT displays or a third of the power of PDPs, due to the efficiency of 5 lm/W or higher to convert electrical energy to light.

An SED consists of a glass plate mounted with electron emitters, similar to a CRT's electron gun, in an amount equal to the number of pixels on the display. Positioned next to it is another glass substrate coated with a fluorescent substance like phosphor. Between the two glass plates there is vacuum. The key to the electron emitters at the heart of the SED is an extremely narrow slit (several nanometers wide) made from ultrafine-particle film between the two electric poles. Electrons are emitted from one side of the slit when 16 to 18 V of potential difference is applied. Some of these electrons are scattered at the other side of the slit and accelerated by another potential difference (approximately 10 kV) applied between the glass plates, and they collide with the fluorescent-coated glass plate, causing light to be emitted. SEDs can provide dynamic color expressions, a sharp picture, and faster video response than LCDs and PDPs.

#### Field Emission Displays (FED)

Instead of using a single bulky tube as do CRTs, *field emission displays (FEDs)* use tiny "mini tubes" for each pixel, and the display can be built in approximately the same size as an LCD screen. Each red, green, and blue sub-pixel is effectively a miniature vacuum tube with sharp cathode points, or nanocones, at its rear. Since FEDs produce light only from the "on" pixels, power consumption is dependent on the display content. That is better than the always-on LCD backlight.

One of FED's implementations is called the ThinCRT. Beams of electrons are fired from negatively charged electrodes (cathodes) through an evacuated glass tube. The electrons strike phosphors at the front of the tube, causing them to glow and create a high-resolution picture. While conventional CRTs consist of a large bell-shaped tube, a ThinCRT uses a flat tube merely 3.5 mm thin.

# HAD (Holographic Autostereoscopic Display) Technology

Holographic autostereoscopic display (HAD) is a simple conversion of LCD technology, replacing the LCD's backlight with an HOE (holographic optical element).

This is divided into two sets of horizontal bands that correspond to each eye. The result is that the left eye sees one image and the right eye sees another one, thereby achieving a 3D effect.

# 2.1.6 Computers Everywhere

The great progress verified in the last years in communication technology has been influencing the lives of many people, allowing access to information and on-line services almost anywhere and at any time. However, the great revolution has not yet arrived. Specialists have foreseen that soon computers will be embedded in practically all devices, with easy and fast Internet connection, allowing new applications.

The following sections present several technologies that contribute to computers being present everywhere.

#### **Embedded Systems**

Nowadays embedded computer systems are ubiquitous: a great many products include electronic systems based on one or more microprocessors to perform a specific dedicated application. These include cellular phones, PDAs, toys, home appliances, industrial equipment, automotive and aircraft devices, and many other everyday products [7]. Such *embedded systems (ESs)* nowadays are responsible for the vast majority of the microprocessors sold.

The main characteristics of ESs are their hardware and software dedicated to specific functions, and their interfaces that can be quite different from that of a typical computer. The human interface may be simple. Wireless communication resources and other forms of interaction with the external world (various sensors and output devices) are frequent. The tendency is for specialized equipment to replace the conventional and multifunctional microcomputers in many applications. Nevertheless, ESs normally have tight constraints concerning some features, like size and weight, power consumption, cost, and functionality. On the other hand, their performance and real-time capabilities are proportional to their cost.

The demand for more complex equipment is growing, due to applications such as multimedia and embedded Java virtual machines. So ESs must be more powerful, with higher processing capabilities, with more memory and network connection resources.

# Pervasive Computing

Current communication resources, especially wireless ones, have been intensifying computer ubiquity, due to the great mobility allowed. Equipment with wireless LANs (WLAN) can be used in a dynamic way, and the only requirement for a network connection is the proximity of a piece of equipment to another one [12,13].

Wireless communication allows the implementation of ad hoc networks, which group several mobile nodes that are moving around a place temporarily, establishing an ephemeral network. One of the most interesting ad hoc networks applications is *pervasive computing*, which provides distributed and invisible computational environments. Pervasive computing allows establishing *personal area networks (PANs)* that in the future will invade homes, buildings, and roads, allowing wide connectivity.

Specialists estimate at least five years for many of the speculations to begin to come true, since they still depend on the evolution of several technologies, such as the popularization of high-speed communication. New applications will be possible, like info-house home appliances that contact the manufacturer in case of fault or before a fault occurs, and the manufacturer can even repair it remotely. Another example is a refrigerator that automatically orders foods from the supermarket.

#### **Portable Computers**

The most common portable computers are *personal digital assistants (PDAs)*. They are an evolution of daytimers, organizers, and small computers, used mainly to look up and record personal information such as appointments, contacts, and notes [10,12]. They are small in size, portable, and easy to use. PDA screens are LCDs, monochrome or colored around 16.5 cm (6.5 inches) diagonally. Data input is based on the *stylus* (a stick similar to a pen) and a touch screen with handwriting recognition software, or on typing each letter at onscreen keyboard. Some PDAs provide an external small keyboard. PDA memory capacity varies from 2 to 64 Mbytes, and is used to store the operating system, programs, and data. Some PDAs allow expansion memory cards.

Several operating systems are available for PDAs, including Palm OS from Palm Inc., Windows Mobile from Microsoft, Linux, and Epoc from Psion. PDAs can be connected to PCs via serial cable (RS 232 or USB) or infrared port, in order to share data (synchronization). Some PC sync software support popular applications like Outlook and Lotus Organizer. Modem can be used to remote sync and access the Internet (with special browsers); emails can also be accessed (both send and receive).

Some PDAs allow listening to MP3 music or hearing an audio book. There are many other accessories that have been integrated with PDAs, such as scanners, wireless communication, GPS receivers, cellular phones, and digital cameras.

They have reasonable processing and storage capabilities, robustness, easy interface, low power consumption, low cost, and powerful programming tools, encouraging new applications. They have been used in restaurants, museums, stores, agricultural vehicles, and many other places.

# Wearable Computers

Most computers do not interact with their users all day long. Notebooks and, more recently, PDAs have some mobility, but they still have a restricted use. *Wearable computing (WC)* changes this situation, allowing the user to wear a computer like a piece of clothing, glasses, or a watch, and interact with it most of the time in a proactive way, acting as an always-on, aware, and smart assistant [6,8,11].

To achieve this status, WCs must have some features (such as low weight) to allow mobility, sensors to collect information about the user's surrounding environment, wireless communication tools, and a friendly interface requiring the minimal use of hands. Specialists foresee that soon many devices—pagers, cellular phones, notebooks, PDAs, music players, and radios—will be merging in WCs that will be worn by most people.

#### 2.1.7 Conclusion

The recent evolution of hardware has probably no parallel on the history of technology. Every year new concepts are developed, new materials appear, and new limits are surpassed. There is much effort behind that evolution. There is also a billion-dollar market. The physical limits of microelectronics, which were once thought to be a problem, will be overcome by new developments and alternative technologies. The incredible pace of that evolution shows no sign of slowing down. The future may show that the "computers everywhere" concept will come true, probably not the way it is envisioned today, but maybe more pervasively than we might expect or desire.

# References

- 1. Seeger, K. 1997. Semiconductor Physics and Introduction. 6<sup>th</sup> ed. Springer-Verlag.
- 2. Schaller, B. 1996. The Origin, Nature, and Implications of "Moore's Law". In *The Benchmark of Progress in Semiconductor Electronics*. Available at: http://mason.gmu.edu/~rschalle/moorelaw.html. Accessed 3 March 2005.
- 3. Jurvetson, S. 2001. Transcending Moore's Law. *Tech News CNET.com*. Available at: http://news.com.com/2010-1071-281576.html. Accessed 3 March 2005.
- 4. Mullins, J. 2001. The topsy turvy world of quantum computing. *IEEE Spectrum* (2): 42-49.
- 5. Pinnel, M. R. 2000. *Display Technology Overview*. San Jose, CA: United States Display Consortium.
- 6. Rhodes, B. 1997. The wearable remembrance agent: A system for augmented memory. In *Proc. of 1<sup>st</sup> International Symposium on Wearable Computers*. 123-128.
- 7. Koopman Jr., P. D. 1996. Embedded system design issues (The rest of the story). In *Proc. of International Conference on Computer Design* (ICCD 96). Austin, TX: EUA.
- 8. 1997. *1<sup>st</sup> International Symposium on Wearable Computers*. Cambridge, MA: IEEE Computer Society.
- 9. Zorpette, G. 2001. The quest of spin transistor. *IEEE Spectrum*, 30-35.
- 10. The Palm Computing Platform—An Overview. Available at: http://www. wirelessdevnet.com/channels/pda/training/palmoverview.html. Accessed 3 March 2005.
- 11. PCWorld.com. Wearable PCs offer function, not fashion. 1999. PC World Communications, Inc. Available at: http://www.pcworld.com/news/article/0,aid, 14535,00.asp. Accessed 3 March 2005.
- 12. Hansmann, U., L. Merk, M. S. Nikolous, and T. Stober. 2003. *Pervasive Computing*. 2<sup>nd</sup> ed. Springer Professional Computing.
- 13. Weiser, M. 1991. The computer for the 21<sup>st</sup> century. *Scientific American* 265(3): 94-104.

# 2.2 Sensors

R. Isermann

Abstract. Sensors and measurement equipment are of fundamental importance for all technical products and processes. They are used to indicate some variables of the processes, to monitor the internal state and the basis for manual control, automatic feedforward and feedback control, supervision, and optimization. After considering the elements of a measurement system and a classification of measured quantities, some sensor properties and signal types are discussed. This is followed by brief descriptions of sensors that are required frequently in industrial, mechatronic, and agricultural systems. This includes measurement of displacement, velocity, acceleration, vibration, force, pressure, torque, temperature, and flow. Finally, A/D conversion, electromagnetic compatibility, and integrated and intelligent sensors are discussed.

**Keywords.** Sensors, Measurement, Signals, Transducers, Amplifiers, Displacement, Velocity, Acceleration, Vibration, Oscillation, Force, Pressure, Torque, Temperature, Flow, EMC, Integrated and intelligent sensors.

**Remark.** This is a shortened version of Isermann, R. 2002. *Mechatronic Systems*. Springer, London.

#### 2.2.1 Introduction

Sensors and the associated measuring systems provide the required measurable information about any process in technical systems. They represent an essential link between the process and the information processing part, i.e., microcomputers (see Figure 1). Sensors that measure mechanical or thermal quantities and transform them into electrical signals are of special importance for mechatronic systems and precision agriculture. This section gives a brief overview of some of the characteristic features, signal types, and measurement principles. A more detailed description of the broad field of metrology is given in, e.g., [1-10]. Mass flow sensors for solid goods are not treated in this section. For related measuring principles and, in particular, applications in yield mapping for combine harvesting, cf. [11-13].

# 2.2.2 Measuring Systems

The purpose of a measuring system is to observe and quantify a variable physical quantity (called the *measurand*) and to process the information obtained. The first element of this system is the *sensor* or *sensing element* (this term is increasingly used

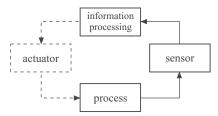


Figure 1. Sensors as links between a process and the information processing unit.

2.2 Sensors 33

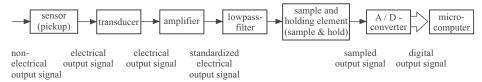


Figure 2. A measuring system.

instead of *pickup*). Its primary function is to detect the measurand and transform it into a suitable signal (Figure 2). Mechatronic systems generally rely on sensors with an electrical output signal. The characteristics of the output signal depends on the measurement principle of the sensor. Transducers and amplifiers transform the electric sensor output/signal into a standardized electrical signal, e.g., 0 to 20 mA or 4 to 20 mA or 0 to 10 V, which is more suitable for further processing. If high frequency disturbances contaminate the usable signal, a low-pass filter is applied in order to decrease the influence. A sample-and-hold device and an analog-to-digital converter are necessary if the sensor-signal is to be processed by a microcomputer.

Consumer goods and low-cost appliances do not require high precision measurement and a modular arrangement of the measuring system. Therefore, simplifications may be made in order to reduce costs, e.g., by omitting the generation of standardized electrical signals.

#### 2.2.3 Classification of Sensors

Because of the broad spectrum of metrology it is difficult to classify sensors and the corresponding signal processing devices. A survey entitled "Technical Sensors" compiled in 1983 [14] proposed a hierarchical division consisting of five levels and 75 subdivisions. Important features for the classification of sensors are: measured quantity, sensor-principle, manufacturing technology, signal types and interfaces, fields of application, properties and features, quality class, and cost.

Table 1 gives an overview of a classification of some important measurands. A rough classification might be: mechanical quantities, thermal/caloric quantities, electrical quantities, and chemical and physical quantities.

	•			
Class		Measuring Quantity		
Mechanical quantities	Geometrical quantities	Displacement, angle, level, gradient		
	Kinematic quantities	Speed, acceleration, oscillation, flow		
	Stress quantities	Force, pressure, torque		
	Material characteristics	Mass, density, viscosity		
	Acoustic quantities	Sound velocity, sound pressure, sound frequency		
Thermal quantities	Temperature	Temperature of contact, temperature of radiation		
Electrical quantities	Electrical state variable	Voltage, current, electrical power		
	Electrical parameter	Resistance, impedance, capacity, inductance		
	Field variable	Magnetic field, electrical field		
Chemical and physical	Concentration	Ph-value, humidity, heat conduction		
quantities	Size of particle	Content of suspended matter, content of dust		
	Kind of molecule	Gas molecules, fluid molecules, rigid body molecules		
	Optical quantities	Intensity, wavelength, color		

Table 1. Survey of the classification of important measuring quantities.

The following sections will deal with properties and features of sensors, as well as different kinds of signal types. In addition the principles of some sensors with an electrical output signal will be described.

# 2.2.4 Sensor Properties

The transformation of non-electrical quantities into electrical ones depends on physical or chemical effects, which may be divided into main and side effects. The *main effect* is responsible for generating the desired measuring signal, e.g., the electrical voltage of a piezoelectric pressure sensor. However, disturbing *side effects*, e.g., the influence of temperature changes, are frequently superimposed. The design process for sensors needs to take these side effects (sometimes called *cross sensitivity*) into account. Their influence should have only little effect or has to be compensated by appropriate measures.

The most important criteria for evaluating sensors are static behavior, dynamic behavior, quality class and measuring range, overload capacity, compatibility with associated components, environmental influences, and reliability.

A sensor's *static behavior* is described by the characteristics of the sensor. It defines the sensitivity of a sensor, i.e. the ratio of the change of the electrical output signal to the change of the measured variable. Other important properties of a sensor are linearity, hysteresis, and repeatability (reproducibility).

The *dynamic behavior* is described by a sensor's frequency response or simple characteristic values, e.g., cut-off frequencies or time constants. The sensor dynamics has to be adjusted to the process and the measuring task.

The *quality class* gives a basic measure about a sensor's accuracy. It is the percentage maximum error of a measurement with reference to the full scale. Applications for consumer goods don't need a high accuracy (2% to 5% are sufficient). Industrial applications, on the other hand, require a much higher precision (0.05% to 1%). Equipment for high precision measurements, e.g., calibration and test equipment, have to meet very strict requirements. The *measuring range* describes the range in which the sensor's specifications are met.

The *overload capacity* specifies the range in which a sensor may be operated without changes in the sensor's characteristics or damage to itself. Typical overload capacities are between 200% and 500%.

A sensor's *compatibility* depends on the output signal type (see next section). *Environmental influences*, e.g., temperature, acceleration, corrosion, contamination, and wear and tear, are especially important.

The *reliability* of a sensor is described by characteristic parameters, e.g. the *mean time to failures* (MTTF in [h]) or its reciprocal value, the *mean failure rate* (in [h<sup>-1</sup>]).

# 2.2.5 Signal Types, Transducers, and Measuring Amplifiers

The type of signal supplied by the sensor depends on both the measuring principle and on the associated signal transmission and signal processing devices. Signal types may be subdivided into the following categories: amplitude modulated signals, frequency modulated signals, and digital signals.

2.2 Sensors 35

	Signal Type					
Properties	Amplitude Modulated	Frequency Modulated	Digital			
Static accuracy	Large	Large	Limited by word length			
Dynamic behavior	Very fast	Limited through transducer	Limited through sampling			
Noise sensitivity	Medium/large	Small	Small			
Galvanic separation	Costly	Simple (transducer)	Simple (optical coupling)			
Interfacing to a digital computer	Analog-to-digital converter	Simple (frequency counter)	Simple			
Computational operation	Very limited	Limited	Simple, if microcomputer			

Table 2. Some properties of signal types for measuring signals.

Amplitude modulated signals are characterized by a proportional relationship between the signal amplitude and the measured quantity. If the signal frequency is proportional to the measured quantity the signal is called *frequency modulated signal*. Digital signals encode a measured quantity using serial or parallel binary signals. Table 2 describes some of the properties of these signal types [2].

*Transducers* convert the amplitude modulated signal into another appropriate electrical signal. Examples of transducer circuits without amplification are the voltage-current transducer with precision resistor, voltage divider and current divider, resistance-current transducer, and compensation network for measuring voltage, current, or resistance (resistance bridge).

Measuring amplifiers raise low power sensor output signals to a higher power level or generate more powerful standardized signals (0 to 10 V, 0 to 20 mA). High power sensor output signals are needed for associated components of the measuring chain, e.g., transmission links, filters, and displays. Measuring amplifiers often consist of operational amplifiers made out of resistors and transistors in the form of analog integrated circuits. Operational amplifiers possess high gains which may vary considerably due to aging and dependence on temperature. Without additional circuitry, operational amplifiers may only be used as zero gains for comparator or compensator circuits. By adding a negative feedback, the gain of the entire circuit depends mainly on the resistors of the negative feedback for the case of high feed forward gains of the operational amplifier. Measuring amplifiers with negative feedback are divided into four basic types of circuits: voltage amplifiers, voltage amplifiers with current output, current amplifiers, and current amplifiers with voltage output.

The following sections give a short description of some important sensor principles. For more detailed descriptions refer to the references given in the bibliography.

# 2.2.6 Displacement Measurement

# Resistive Sensing Elements

Resistive sensing elements exploit the proportional relationship between the length of a wire or film resistor and its electrical resistance (Table 3). They are potentiometers made of electrically conductive plastics or metal wire. Potentiometers are

•	resistive sensors	inductive sensors	capacitive sensors	strain gauge	code sensor	incremental sensors	Hall sensor
sensor principle (example)		$V_{R_0}$	x d x x A x x + 8	$R_1$ $R_2$ $R_3$ $R_3$ $R_3$	Gray Code i: 123456	light sensor	Hall-IC V <sub>s</sub>
material	metal, semiconductors, conductive plastics	ferromagnetic metal	capacitor	metal, semiconductor	optical encoders	glass, metals	Hall semiconductor
output signal	analog voltage	analog voltage	analog voltage	analog voltage	binary signal	binary signal	binary signal
measurement range	1cm2m 300° (angular displacement)	+100μm+50cm	0.1cm10cm			10mm3m	360°
maximum sensitivity	0.2V/° or 2V/cm	0.1V/cm 40mV/μm		$k = \frac{\Delta R/R}{\Delta l/l} = 2 \text{ (metal)}$ $k = 100 \text{ (semiconductor)}$	4096 pulses/rev		4000 pulses per revolution
accuracy, resolution	max. 40 μm or 0.1°	0.1 μm	<0.1 nm		1 LSB	0.1 μm 0.00005°	10 <sup>-5</sup> revs
temperature range	-50°C+250°C	-40°C+100°C	up to 800°C	-270°C+1000°C	-50°C+100°C	0°C+50°C	-200°C150°C

Table 3. Displacement sensors (linear)

2.2 Sensors 37

wired as voltage dividers and offered as linear or rotational sensing elements, in the latter form as multiplex potentiometers (e.g., 10 revolutions). The measuring range of linear sensing elements starts at a few millimeters and goes up to about two meters. Encapsulation of the sensor housing enables the deployment of the sensor in rough environments. Safety rails ensure a movement of the brush free of lateral forces. Sensing elements made out of electrically conductive plastics have a very high resolution, e.g. measurement of 100 mm with a resolution of 0.01 mm. However, a high accuracy requires a very precise voltage source.

## **Inductive Sensing Elements**

Inductive sensing elements rely on the dependence of the change of the self- and mutual inductance on the element's position. The inductance of coil arrays is changed by variation of the air gap. A lattice network consisting of differential coils ensures an almost linear characteristic.

Differential transformers exploit the relationship of the mutual inductance between the primary and secondary coil and a displacement of the iron core. The primary coil is subjected to a carrier frequency. The difference in the voltage of the secondary coil acts as a displacement dependent output signal. Inductive sensors are non-contact sensing elements. Their measuring range starts at a few millimeters and goes up to about one meter. Other designs are displacement angle sensing elements.

## Capacitive Sensing Elements

A change in plate distance, plate area, or of the dielectric material between plates influences the capacitance of a capacitor. The signal processing circuits consist of AC lattice networks (capacitive bridges). They have to be operated with a high carrier frequency (0.5 to 1 MHz) because of small capacitances

### Strain Gauges

Strain gauges transform small linear deformations into electrical signals. They are based on the effect that a change of the length of an electrical conductor results in a change of the electrical resistance. If one expands a wire of length L by  $\Delta L$  the electrical resistance of the wire changes due to changes of the specific electrical resistance (because of structural deformations), the length, and the cross section area.

Metal wire and film strain gauges consist of thin constantan wires or films. Changes of the length and cross section area lead to a change of the electrical resistance, while the specific electrical resistance is not affected. For *semiconductor* strain gauges the main effect is the change of the specific resistance due to the structural deformation by elongation. Semiconductor strain gauges are much more sensitive to elongation (about 40 to 80 times more sensitive than constantan) but have non-linear characteristics at large elongations and are much more expensive than metal strain gauges. Embedded between thin films, strain gauges are pasted directly on the measuring object. This could directly be the constructive component whose elongation is to be measured. Strain gauges in conjunction with special spring elements and diaphragms serve as force, torque, and pressure sensors. The change of the electrical resistance is evaluated with lattice networks (bridge circuits). Temperature compensation is often provided.

#### Encoders

Encoders use code rulers or code discs on which the discrete displacement data is encoded. The allocation is absolute because they do not need an external reference. Unit-distance codes, e.g., Gray code, are often used for coding. The sampling is performed optically. In order to distinguish  $2^n$  different discrete positions one needs n sampling tracks. This makes this kind of sensing element relatively complicated. Coded sensing elements are mainly used in industrial metrology, e.g., for numerically controlled machine tools and robots.

## **Incremental Sensing Elements**

Incremental position and angle sensors count the number of so-called notches or slots, i.e. *increments*, relative to an initial point. Sampling is performed either by optical (e.g., diodes) or inductive methods resulting in pulse trains, which are counted. The initial point can be chosen arbitrarily. If a failure (e.g., a power failure) occurs the initial point gets lost and must therefore be reset by moving to a reference position. A fault while counting the increments influences all following readings. Additional circuitry with two samplers per scale enables detection of movement direction and a pulse multiplication.

The sensor housing often contains a pulse shaping circuitry. Incremental sensors are mainly used in industrial metrology, e.g., manufacturing. Linear scales with a range of up to 3 m and a graduation of 1  $\mu$ m are available. Shaft encoders with up to 36 000 graduations are used for precision applications. Using pulse multiplication (interpolation) a resolution of about 0.00005° is achievable.

### Hall Effect Sensors

If a voltage is applied to a conductor or semiconductor located in a magnetic field perpendicular (right-angled with the current flow) to the applied voltage, the Hall voltage is generated. This is perpendicular to both the current flow and the magnetic field. The dependence of the Hall voltage on the magnetization is now used for proximity or position measurement. If the semiconductor is a silicon Hall-plate the voltage has to be amplified. Integrated Hall ICs exist which incorporate amplification, stabilization, and temperature compensation. A rotational position sensor now consists of a permanent magnet and a soft-magnetic tooth wheel moving through a gap between the magnet and the Hall IC (e.g., bipolar technology). Hence, by interruption of the magnetic field a pulse train is generated whose frequency is proportional to the rotational velocity. The Hall IC requires a supply voltage (e.g., 12 V). This type of sensor is used, for example, for rotor position sensing of brushless DC motors and for ignition triggering of SI engines [4].

### **Other Measuring Methods**

Ultrasonic distance sensors are used for level measurement (bulk material, fluids) or as distance meters (park assistance for an automobile). They rely on the time interval between transmission and echo-return of the ultrasonic signal. Laser interferometers are based on the phase comparison of coherent light and are used for contactless precision displacement measurement.

2.2 Sensors 39

All of the above mentioned measuring methods may also be used for the measurement of force or pressure. They detect displacement of springs or membranes and transform it into an electrical signal.

## 2.2.7 Velocity Measurement

One possibility of measuring velocities is to differentiate the signal of displacement or angular sensors. However, this has the disadvantage of amplifying the noise relative to the usable signal. Therefore direct measuring methods for velocity measuring are more suitable. Angular velocity sensors are of especially practical significance. Translational velocities are often converted into rotational velocities for measurement purposes (i.e., a speedometer).

## Active Electrodynamic Sensing Elements

Active electrodynamic sensing elements operate like generators. If N electrical conductors move through a magnetic field with magnetic flux  $\Phi$  this induces a voltage

$$V = -\frac{N d\Phi}{dt}$$

in the conductors according to the law of induction. One can distinguish two types:

- (1) Sensors for translation—A permanent magnet moves inside a coil and induces a voltage which is proportional to the velocity.
- (2) Sensors for rotation—AC generators use a permanent magnetic rotor and a stator winding. In order to measure the rotational speed both the output signal's frequency and voltage may be used. A linear voltage characteristic results. DC tacho generators consist of a commutator and a coil rotating in a constant magnetic field derived from a permanent magnet. The generated voltages are then proportional to the rotational speed. The polarity of the voltage depends on the sense of rotation. This sort of sensing element has a linear characteristic but a rippled voltage signal because of the commutation.

## **Incremental Inductive Sensors**

These sensing elements correspond to incremental displacement sensors. A ferromagnetic ring gear with rectangular teeth passes an inductive sensor. This sensor consists of a bar magnet with a soft-magnetic pole pin and an induction coil. The voltage in the coil is proportional to the periodic variation in the magnetic flux. The output signal therefore is a pulse sequence and its frequency is related to the rotational speed. A frequency-voltage converter transforms the pulse frequency into a voltage. Discrete evaluation is performed by counting the number of pulses during a certain time span or by measuring the time interval between two pulses. The number of pulses per rotation depends on the specific application and ranges from 1 to several 1000 pulses. This sensor is for example used for engine crankshafts and wheels for ABS (antilock braking system) functions. It does not need an electrical power supply.

### Other Methods

Another method to measure translational velocity exploits *Doppler's effect*. A velocity-dependent frequency shift between a transmitted and a reflected signal by a

moving object occurs. Well-known examples for these kinds of sensors are the Doppler radar (traffic radar, speed measurement for agricultural machines) and the laser Doppler for high precision, contactless measurement (at high cost).

Yet another method utilizes *cross correlation* between two stochastic or periodic signals. Two identical sensing elements are placed at distance l and the measurement of the time delay  $\tau$  between the transmitted and received signal enables the determination of the velocity  $v = l/\tau$ . This is, for example, used with optical sensors for rough surfaces or for fluids.

### 2.2.8 Acceleration Measurement

Measurement of acceleration is frequently based on force measurement using the relationship between the acceleration a of a mass m and the inertial force F: a = F/m. For direct measurement of the force F, piezoelectric force sensors may be used. Due to the large spring stiffness and small masses of these transducers it is possible to achieve high natural frequencies (100 kHz).

Spring-mass systems consist of a seismic mass connected to the sensor casing with springs and dampers. The acceleration is determined by measuring the displacement of the spring (e.g., inductively). There are many different kinds of acceleration sensors that are capable of measuring accelerations of, e.g.,  $10^{-6}$  g for inertial navigation purposes or up to  $10^{5}$  g for measuring explosions. Masses, springs, and dampers are chosen such that high natural frequencies (15 Hz to over 100 kHz) are achieved. The utilizable measurement frequency reaches about half of the natural frequency. Sensors for angular acceleration use equivalent arrangements.

Acceleration may also be determined by differentiating the signal of a velocity sensor once or by differentiating the signal of a displacement sensor twice. However, this leads to an increase in high frequency noise and makes low-pass filtering inevitable.

### 2.2.9 Vibration and Oscillation Measurement

The measurement of *relative vibrations* is based on the displacement between two reference points, which is measured by displacement sensors. For measuring *absolute vibrations*, the missing second reference point has to be replaced by a seismic mass. Measurement of the oscillation amplitude requires a large seismic mass and a small spring constant of the suspension (small natural frequency). This leads to a motionless seismic mass and an oscillating sensor housing. A displacement sensor (e.g., inductive sensor) measures the displacement between the seismic mass and the sensor housing. The same applies to electrodynamic velocity sensors which measure the vibration velocity. Vibration acceleration is measured with acceleration sensors calibrated for high natural frequencies.

#### 2.2.10 Force and Pressure Measurement

Measurement of pressure and force is performed indirectly by measuring the spring or diaphragm deflection with displacement sensors like those described in Section 2.2.6 (especially strain gauges, used for what are called *load cells*, and inductive sensing elements).

2.2 Sensors 41

Piezosensing elements are of special importance for pressure and force gauging. Piezoelectric sensing elements exploit the piezoelectric effect: A displacement results in an electrical charge at the surface of a crystal lattice. The arising displacement is very small (a few  $\mu$ m). The electrical charge charges the artificial capacitance (consisting of the sensing element, wire, and amplifier input). The resulting voltage V decays with the time constant T = RL. This is the reason why piezoelectric sensing elements are suitable for dynamic measurement only.

The subsequent measuring amplifier has to have a very high input resistance ( $R > 10^{13} \Omega$ ) in order to obtain a large time constant. Charge amplifiers are used with time constants of up to several hours. The maximum measuring frequency is about 100 kHz.

*Piezoresistive* sensing elements exploit the piezoresistive effect. In this case a crystal subjected to mechanical forces changes its electrical resistance due to a dislocation in the crystal lattice structure. This allows also static measurements.

## 2.2.11 Torque Measurement

Torque is measured by gauging the torsion of a shaft section using angular, displacement, or elongation sensors. For this purpose special *torquemeters* may be attached to the shaft using flanges with or without bearings. Another possibility is to base the measurement of torque on the torsion of the shaft due to a load. The signal transmission depends whether the shaft is rotating or not. If the shaft is rotating the revolving sensors transmit their signals to the stationary electronic signal processing unit via a slip ring or without direct electrical contact, e.g., through inductive coupling.

Measurements of a shaft's torsion can be obtained by employing either of the two following methods. The first method uses wire strain gauges which are placed on the shaft with an inclination of 45° to its longitudinal axis, interconnected to a Wheatstone network (see Table 4). The second method measures the change of permeability by measuring the voltage induced in coils. Both measurement principles may be applied directly to the shaft or in conjunction with special torquemeters. In many cases *torque-gauge heads* are necessary which are easily incorporable, do not require much space, and do not introduce too much elasticity. Additional requirements include the possibility to connect a sensor via, e.g., a flange to the shaft or integration of the sensor into pulleys.

The rotational angle between two twisted discs or axial dislocation of discs due to kinematic transmission may be measured inductively. Another measurement principle uses disc- or sleeve-like parts that consist of electrically conductive and non-conductive zones. Twisting of the shaft leads to a shift of the zones against each other and results in a change of the eddy current. This yields in a change of the impedance of a stationary measuring coil. Additional possibilities for measuring torque include the use of surface-resonators and piezoelectric sensors which are placed into the force flux.

42 Chapter 2 Hardware

Tuble in Force, torque, and pressure sensors.											
	Sensor Principle (example)	Material	Output Signal	Measure- ment Range	Sensi- tivity	Tempera- ture					
Piezoelectric force sensor		Piezo- electric material	Analog voltage	1 N to 1mN	125 v/kN	-80°C to +150°C					
Torque measurement using strain gauges	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Strain gauges on surface	Analog voltage	0.05 Nm to 50 kNm		+10°C to +60°C					
Force spring deflection sensor	-00000-	Spring in a casing	Analog displacement			-40°C to +60°C					
Pressure diaphragm sensor	<i>p</i> <sub>2</sub> <i>p</i> <sub>1</sub> <i>p</i> <sub>2</sub> <i>p</i> <sub>3</sub> <i>p</i> <sub>4</sub> <i>p</i> <sub>5</sub> <i>p</i> <sub>7</sub> <i>p</i> <sub>7</sub> <i>p</i> <sub>8</sub> <i>p</i> <sub>7</sub> <i>p</i> <sub>8</sub> <i>p</i> <sub>7</sub> <i>p</i> <sub>8</sub> <i>p</i>	Flexible diaphragm	Analog displacement	0.1 bar to 10000 bar		-25°C to +100°C					

Table 4. Force, torque, and pressure sensors.

# 2.2.12 Temperature Measurement

#### Resistance Thermometers

Passive resistance thermometers exploit the sensitivity of electrical resistors to temperature variation (compare Table 5). Metal resistance thermometers consist of nickel or platinum wire wound in the form of a free spiral around thin mica or laminated paper stripes or embedded in glass. The resistance-temperature-sensitivity is 0.358  $\Omega/K$  for platinum (Pt100) and 0.612  $\Omega/K$  for nickel (Ni100). The nominal resistance in each case is 100  $\Omega$  and the maximum measuring temperature is 250°C for Ni and 850°C for Pt.

A thermistor is a resistance thermometer consisting of a metal-oxide semiconductor material. It is about ten times as sensitive as metal resistance thermometers, has a strongly nonlinear characteristic and is less accurate. Commercial forms are very small (< 0.5 mm) and thus possess a low heat capacity. This makes thermistors suitable for measuring surface temperatures and for measurement of dynamic processes. If the semiconductor possesses a negative temperature coefficient (electrical conductivity increases if temperature rises) it is called an *NTC thermistor*, while a semiconductor

Table 5. Temperature sensors. Measurement Output Sensi-Accuracy, Signal tivity Resolution Sensor Principle Material Range Resist--200°C Pt ance to thermo-Metal +1000°C 0.3% meters resistor -250°C Ni to to +1000°C 0.25% of  $(=/\sim) V_0$ -40°C voltage measured NTC temperto Semi-+850°C ature conductor -200°C resistor PTC to +850°C Thermocold junction 0.25%. metal A copper couples -180°C Analog 53μV/K 0.75% of hot Fe-Cu measured junction +760°C temper-

with a positive temperature coefficient (electrical conductivity decreases if temperature rises) is called a PTC thermistor. These metallic resistors can be produced as thin-film or thick-film sensors and are then integrated upon a single substrate wafer with neutral trimming resistors for precision manufacturing [4]. Maximum temperatures range from 100°C to 1000°C for NTC thermistors and from -10°C to 500°C for PTC thermistors.

copper

ature

## **Thermocouples**

metal B

The arrangement of two lengths of dissimilar wire, insulated from each other but joined at one end, is known as a thermocouple [1]. Thermocouples are active temperature sensors. Exposing the junction of the two metals to heat generates an EMF (electromotive force) that depends on the temperature at the junction. This is called the Seebeck effect. When the junction between two dissimilar metals is heated or cooled relative to a second reference junction, the resulting overall voltage is a function of the difference temperature of the two junctions. The performance of a thermocouple is usually specified in relation to a reference temperature of 0°C (ice water). In many practical arrangements the reference junction is located in a controlled environment with non-zero reference temperature.

The temperature-voltage characteristic is non-linear but a linearization is possible for a wide range of operation. The advantages of thermocouples are their small dimensions, which lead to small measuring points, and that they do not need a power supply. However, they have low sensitivities and the output signal level is small. Thermocouples enclosed in protective tubes with a diameter of 0.25 mm to 3 mm have a wide measuring range of 220°C to 2400°C. Due to the small heat capacities of such sensors it is possible to measure even rapid temperature changes.

#### 2.2.13 Flow Measurement

The flow  $\dot{q}$  results from the ratio of a fluidic quantity  $\Delta q$  (liquid, gas) per time interval  $\Delta t$  flowing through an area A with mean velocity v as

$$\dot{q} = \frac{\Delta q}{\Delta t}$$

If the quantity is a volume one obtains a volumetric flow as

$$\dot{V} = \frac{\Delta V}{\Delta t} = A V \left[ \frac{m^3}{s} \right]$$

and if it is a mass, the mass flow of

$$\dot{m} = \frac{\Delta m}{\Delta t} = \dot{V} \rho \left[ \frac{\text{kg}}{\text{s}} \right]$$

results, with the mass density  $\rho$  [kg/m<sup>3</sup>].

One distinguishes volumetric flow measurement and mass flow measurement. In the first case the volume is measured as fluid parts or as flow speed and in the second case as mass of fluid parts. However, if the density  $\rho$  is known both flow measures can be calculated from each other. There exists a large variety of flow measurement principles, see e.g. [7,9,10,16,17]. The following measurement principles describe briefly some common used flow meters and are restricted to single phase flows of gases and liquids. Table 6 gives a summary of the discussed flowrate meters.

## Positive Displacement Flowmeter

Positive displacement flowmeters measure the flow by momentarily entrapping a part of the fluid into a compartment of known volume. Because of the displacement from the inlet side to the outlet side the total volume is measured by counting the number of compartments and the volumetric flowrate by the ratio of compartments volume per time interval. For liquids most common types are piston, sliding vane, biand tri-rotor, and disc-type flowmeters; for gases, bellows, diaphragm, or roots flowmeters. Table 6 shows an oval gear flowmeter (piston-type). Positive displacement flowmeters have a high accuracy, a wide measurement range, can be used for liquids with wide range of viscosity and density, and are insensitive to the flow velocity profile, but have moving parts, are sensitive to solid particles, and need maintenance.

#### **Turbine Flowmeters**

Turbine flowmeters have a multibladed rotor that rotates with a speed proportional to the flowrate through a section. The speed of the rotor is, e.g., measured by the frequency of an alternating voltage, induced by permanent magnet on the rotor or outside in a coil mounted at the housing. These flowmeters measure the average speed, and as the pipe diameter is known, the volumetric flowrate can be determined. Turbine flowmeters have a high accuracy, can be used for liquids and gases, have a wide range, but wear because of the moving parts and need maintenance and recalibration. Other types that use the same measurement principle use radial-vaned impellers, single and multi-jet propellers, pelton wheels, or paddle wheels.

Table 6. Flowrate sensors for liquids

mass flowmeter	hot film	Hot-film Sensor Temperature Sensor	>	>	platinum hot-film	bridge voltage	1:50	± 2 %		wide range high dynamics small size
volumetric flowmeter	Coriolis		>	<i>&gt;</i>	U-shaped, vibrating tube	twisted U-tube	1:50 1:100	± 0.25 %		wide range high accuracy expensive
	ultrasonic	Sensor 2	>	<b>/</b>		travel times Doppler frequency	1:20 $1:100$	± 2 ± 5 %		robust wide range expensive
	electromagnetic	Sensor Magnetic Core	>	<i>&gt;</i>	AC & DC	induced coil voltage	1:10 (1:30) 0.3 10 m/s	± 0.25 ± 5 %		conductive liquids robust limited range
	variable area		>	<i>&gt;</i>	conictube and float	float position	1:10	± 2 ± 4 %		directly readable robust laboratories
	differential pressure		>	<i>/</i>	orifice Venturi-tube nozzle	differential pressure	1:3	$\pm$ 1 $\pm$ 2 %		pressure drop limited range maintenance
,	turbine flow meter		>	<b>/</b>	multibladed rotor	voltage frequency	1:10 $1:1000$	± 0.1 ± 1 %		high accuracy wide range maintenance
classification displacement flowmeter	oval gear		>		rotating components	voltage pulses	1:20	$\pm~0.5~~\pm~1\%$		high accuracy wide range maintenance
classification	example	sensor principle (example)	liquids	gases	components	output signal	measurement range	accuracy %	max. temp.	remarks

## Differential Pressure

Differential pressure flowmeters measure the pressure drop across a restriction. This restriction can be an orifice plate, a nozzle, or Venturi tube. The measurement principle is based on the Bernoulli equation for a stationary flow without friction which relates the pressure  $p_1$  and  $p_2$  at two different cross-sections  $A_1$  and  $A_2$  of a conic pipe to the velocities  $v_1$  and  $v_2$ , as

$$\rho \frac{v_1^2}{2} + p_1 = \rho \frac{v_2^2}{2} + p_2$$

The conservation of mass yields

$$v_1 A_1 \rho = v_2 A_2 \rho$$

and from both equations it follows

$$\dot{V} = v_1 A_1 = \frac{A_2}{\sqrt{1 - \left(\frac{A_2}{A_1}\right)^2}} \sqrt{\frac{2}{\rho} (p_1 - p_2)}$$

Because of non-ideal fluids with friction, turbulent flow, compressibility, and contradiction of the flow at the orifice some correction factors are required and one finally uses the following equation in case of an orifice.

$$\dot{V} = \alpha A_0 \sqrt{\frac{2}{\rho}(p_1 - p_2)}$$

$$\alpha = \frac{C\varepsilon}{\sqrt{1-\beta^2}}$$

where C is a discharge coefficient,  $\varepsilon$  the expansibility factor, and  $\beta = d/D$  the ratio of the diameter d of the orifice and the upstream diameter D of the pipe. C and  $\varepsilon$  are documented in tables of international standards, such as ISO 5167-1 (EU) or API 2530 (US). Installation requirements have to be followed in order to have a well developed turbulent flow at the orifice.

The pressure difference  $\Delta p = p_1 - p_2$  before and after the orifice or between the inlet or smallest section of a Venturi tube is measured by a difference pressure sensor. Differential pressure sensors are widely used in industrial plants, where a measurement accuracy of about  $\pm 2\%$  is sufficient. However, they need maintenance because of contamination by dirt and wear of the orifice. Because the flowrate is proportional to  $\sqrt{\Delta\rho}$  the measurement range is rather limited, to about 1:3.

#### Variable Area Flowmeters

Variable flowmeters consist of a conic tube and float (or bob) which change the fluidic area by moving up or down. The differential pressure is approximately constant.

2.2 Sensors 47

The most known type is the taper tube and float flowmeter also known as a *rotameter*. The position of the float is a function of the flowrate. If the tube is of glass or clear plastic, the position of the float is visible and a calibrated scale shows directly the flow without additional power supply.

For higher pressure and temperatures the tube is of stainless steel. A magnetic coupling between the float and an outside pointer may indicate the position. The variable flowmeter is especially attractive for laboratory applications and is also frequently used in industry together with electric limit switches.

## Electromagnetic Flowmeters

The underlying principle of electromagnetic flowmeters is Faraday's law of electromagnetic induction. If a conductive fluid moves perpendicular to a magnetic field, a voltage, which is proportional to the velocity, is induced across the conduction. The voltage is measured using electrodes in the pipe wall. The electrodes have mostly to be isolated from the metal pipe wall by a nonconductive liner. Coils on the outside of the pipe generate a magnetic field across the fluid with AC or pulsed DC excitation voltages. This flowmeter has no moving parts, no obstruction of the pipe is necessary, is relatively easy to install, and has high accuracy, but it has limited measurement range (1:10 or larger) and needs fluids with a conductivity greater than 1 to 5  $\mu S/m$ . It is not suitable for gases or steam.

### Ultrasonic Flowmeters

Ultrasonic flowmeters use the fact that the propagation velocity of an acoustic wave and the flow are added vectorially. In the case of the *transit time flowmeter*, the difference in the transit times between two ultrasonic pulses transmitted upstream and downstream are measured across the flow. Two transducers, transmitters (piezoelectric) and receivers, are mounted on both sides of a pipe with a certain inclination angle for the ultrasonic waves passing through the fluid. The time difference for a pulse of ultrasound (500 kHz to 10 MHz) to travel between these transducers upstream and downstream is proportional to the fluid velocity. Because the difference of travel time is in the range of 100 ns, digital signal evaluation with high solution counters is required. It can be used for liquids and gases, has a wide measurement range and a high accuracy.

The ultrasonic *Doppler flowmeter* uses the Doppler shift effect. The Doppler frequency is the difference between the transmitted and received signals, which is proportionally to the velocity of the liquid. However, this requires that a part of the ultrasonic waves are reflected by acoustic discontinuities such as particles, bubbles, or turbulent eddies. This flowmeter is sensitive to changes in the velocity profile and spatial distribution of discontinuities. Therefore, the accuracy is about  $\pm 5\%$ . It cannot be used for gases or clean liquids.

### Coriolis Flowmeter

This flow meter makes use of the Coriolis effect. The fluid passes through a vibrating tube resulting in a Coriolis force which is proportional to the mass flow. The fluid is accelerated as it moves towards the point of maximum vibration and decelerates by

48 Chapter 2 Hardware

moving further. A U-type of tube is therefore twisted. The measured twisting of the tube is then a measure for the mass flowrate of the fluid. Changes in density, viscosity, velocity profile, etc., do not effect the calibration. Coriolis flowmeters are until now limited to tubes of 150 mm diameter. They can also be used for two-phase flows. Advantages are the high accuracy and the wide range. Disadvantages are the high cost and, for larger flowrates, bulky construction.

#### Thermal Mass Flowsensors

Thermal flowmeters measure, for example, the effect of a flowing fluid on a hot wire or hot film. If the displacement of temperature profile around a heater is measured, it is called a *calorimetric sensor*.

Hot-wire and hot-film sensors contain a sensor element that has a highly temperature-dependent resistance, such as platinum. For protection, hot-film sensors are coated with alumina if used in a gas, or with quartz if used in a liquid. The hot film (or wire) is arranged in a Wheatstone bridge providing a constant current through the sensor element. A flow in the measurement channel cools the hot film (or wire), decreases in resistance and unbalances the bridge, generating an output related to the mass flow.

An alternative is a constant temperature bridge. A flow cools the hot film and its resistance decreases and unbalances the bridge. A differential amplifier balances the bridge with a feedback voltage. The output voltage is proportional to the square root of the flowrate. Hot-film sensors may have a wide measurement range and good accuracy, can be built as small integrated elements, and have a small time constant. They are, for example, used in combustion engines for air flow measurement.

### 2.2.14 Analog to Digital Conversion

The discussed sensing elements transform a physical quantity into a change of electrical resistance, capacitance, or inductance. DC or AC lattice networks transform these changes of electrical properties into a voltage signal and amplify it. Other sensing elements produce low power voltage or current signals which need to be amplified by special circuits, e.g., voltage or current amplifiers, electrometer, or charge amplifiers.

If a microcomputer is used for data acquisition purposes, analog-to-digital conversion is necessary. A low-pass filter limits the measurement signal bandwidth if necessary in order to achieve compliance with the sampling theorem. The filter is followed by a sample-and-hold device and finally the analog-to-digital converter (ADC). The precision of the ADC is chosen depending on the application. Standard ADCs have an 8-, 10-, or 12-bit precision while high precision applications need 16-bit. Converters that use time interval or frequency as an intermediate quantity (charge-balancing or dual-slope converters) integrate the input voltage over a fixed input sample time (the measuring interval). An example is the digital voltmeter. This kind of conversion method is very precise but the measured quantity may not change rapidly. Often low-pass filters and sample-and-hold devices are not used in this conversion process. ADC that use the principle of compensation, e.g., successive-approximation converters, operate by comparing the input voltage with the output of a digital-to-analog converter (DAC). The ADC employs the DAC in a feedback loop [1]. In conjunction with sample-and-hold devices this type of converter makes conversion rates of up to 1 MHz

2.2 Sensors 49

possible. Parallel ADC (flash converters) reach even higher conversion rates (up to 100 MHz with 10-bit precision).

## 2.2.15 Electromagnetic Compatibility (EMC)

The surroundings in which a sensor system is applied influences the selection of a sensor type. The environmental conditions are of mechanical (vibration, shock), thermal, and chemical (water, salt, oil, solvents) nature. Electromagnetic radiation from the surrounding environment poses another important influence on sensor systems. A sensor system's property to remain neutral to this influence is called *electromagnetic* compatibility (EMC). The sources of this sort of interference are manifold and cover a wide range of frequencies, e.g., from  $16^2/3$  Hz (power supply of trains) to several GHz (radar installations). The power supplies of trains, electrical substations, transmission lines, radio, television and communication transmissions, radar installations, welding tools, lightning, etc. all emit electromagnetic energy that influences a system from a distance. Other interferences are due to wire-bound influences from the power supply. Examples of this are peak loads of other electric consumers, commutator sparking of electric motors, variations of the power supply level, e.g., in a 12 V vehicle electrical system, or spikes and collapses in the power supply due to a breakdown of other systems. In addition to this, electric and electronic devices influence each other. This near field influence has to be taken into account, too. This is especially the case if the devices are mounted in a confined space. Examples of near field influence include crosstalk between wires in a cable tree, emission of interference by electric drives, the clock rate of microprocessors and other digital devices, thyristor circuits, and ignition systems. In addition to these effects one has to take into account static charges, accidental earths and handling errors, e.g., faulty connections or short circuits. There are a lot of problems of this sort especially in automotive applications.

Suppression or reduction of the interference emission at its source is an important countermeasure. There are several methods for achieving this, such as appropriate housing or use of interference suppressor coils in supply lines. Limits for noise field intensities are given in the appropriate regulations.

Sufficient space between devices, especially between wires, helps to reduce the noise level. Instrumentation and power cables should always be installed separately. The use of radio shielding (metal casing) is one possibility to shield devices and components. Transmission lines configured as twisted pairs are less susceptible to inductive influences and may be shielded from capacitive (and high-frequency magnetic) influences. A proper connection of the shielding to earth is essential. An effective method for avoiding problems with EMC is reducing the length of wires needed for interconnecting devices. This is achievable by integrating sensor, measuring amplifier, and signal conditioning devices into a single unit. Using signal transmission that is safe from interference (use of high signal levels, current transmission 4 to 20 mA, encoded transmission with error detection) further improves on the EMC. Optical transmission using fiber cables in conjunction with measuring principles insensitive to electromagnetic influences (optical, digital) leads to even less susceptibility to interference.

50 Chapter 2 Hardware

Examining the EMC properties of parts of a system or of an entire system (EMC/EMP tests) is very expensive and difficult. A complete examination of all interactions in a complex system is usually impossible. Therefore it is necessary to consider EMC aspects early, during the design of a system or component. Both emission of interference and susceptibility to interference of all components has to be minimized.

# 2.2.16 Integrated and Intelligent Sensors

Sensors are designed with two goals in mind. The desired signal should be the dominating signal a sensor measures and the measured signal must correspond to the actual physical quantity unambiguously. In reality these goals cannot be fulfilled entirely. Sensors are subjected to side effects, e.g., cross-sensitivity, perturbations, nonlinear transmission (non-linear characteristics, hysteresis, responsiveness, null drift), drift, aging, slow dynamic behavior, and individual manufacturing tolerances. Disregarding these non-ideal properties during signal processing leads to faulty measurements. This is why steps have to be considered to compensate for some of these side effects even when using analog evaluation circuits. These steps include the use of filters, signal-differences of two identical sensors, or special circuits to suppress the null drift [15]. However, many additional and new possibilities arise with the use of *digital signal processing*.

Figure 3a shows a conventional measuring chain with associated analog-to-digital converter and a microcomputer or microcontroller. If, for example, the non-linear characteristics of the sensor do not change over time it is possible to linearize or adjust them by using the microcomputer. This enables calibration of each individual sensor during its manufacturing and decreases the necessary measures on the analog part of the sensor. The use of microcontrollers makes frequency-modulated and incremental sensors economical because of the built-in counters which are able to measure frequencies easily.

Further improvement is possible by *integration* of sensor, signal processing, ADC, and micro-computer with bus interface into one single unit as shown in Figure 3b. This integration (possibly onto one single chip) has several important advantages: reduction of costs for large scale manufacturing, reduction of space requirements, higher precision, and decrease of susceptibility to noise. Because of this integration, however, the requirements for robustness and reliability increase because a sensor is often subject to a rough environment.

Integrated sensor components allow the realization of additional functions. This leads to what are called *smart sensors* or *intelligent sensors*. One example is the use of a second sensor to measure a parasitic quantity, e.g., temperature, and using this measurement to compensate an unwanted side effect. Other examples result from the use of special algorithms built into the microcomputer. These algorithms serve as noise filters, for compensation and linearization purposes, for compensating hysteresis effects (due to magnetic properties, friction, responsiveness), for compensation of self-

2.2 Sensors 51

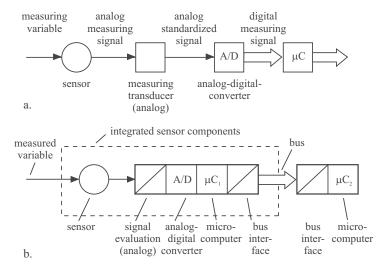


Figure 3. Integration of sensor technology. (a) Conventional measuring chain with digital processing unit. (b) Integrated sensor components with digital processing unit.

calibration for all algorithms during manufacturing and maintenance and even fault detection and fault diagnosis schemes. The fact that all of the algorithms may be programmed individually for each sensor is also very important. Manufacturing tolerances for the sensors need not be as high as in the conventional case. One way of realizing the digital processing chip is an application-specific integrated circuit or ASIC. Tränkler [2] describes the development of intelligent sensors.

It is interesting to see that sensor technology development follows a path similar to that of mechatronic system development, i.e., reducing the requirements on the sensor elements and shifting certain functions into microelectronics by even including more intelligent functions.

Further possibilities arise with the use of multi-sensor technology, i.e., a combination of similar or different types of sensors, and with the many developments now emerging through micromechanics.

### References

- Jones, B. E. 1977. Instrumentation, Measurement and Feedback. New York, NY: McGraw-Hill.
- 2. Tränkler, H. R. 1992. *Taschenbuch der Meßtechnik*. München, Germany: Oldenbourg.
- 3. Beckwith, T. G., R. D. Marangoni, and J. H. Lienhard. 1995. *Mechanical Measurement*. 5<sup>th</sup> ed. Addison Wesley Publishing Company.
- 4. Bauer, H., ed. 1996. *Automotive Handbook*. 4<sup>th</sup> ed. Warrendale, PA: Bosch, Stuttgart and SAE.
- 5. Christiansen, D., ed. 1996. *Electronics Engineers' Handbook*. 4<sup>th</sup> ed. New York, NY: McGraw Hill.

- 6. Jurgen, R. K. 1997. Sensors and Transducers. Warrendale, PA: SAE.
- 7. Webster, J. G. 1999. *The Measurement, Instrumentation and Sensors Handbook*. Boca Raton, FL: CRC Press.
- 8. Whitaker, J. C. 2002. *Electronic Systems Maintenance Handbook*. Boca Raton, FL: CRC Press.
- 9. Profos, P., and T. Pfeifer. 2002. *Handbuch der industriellen Meßtechnik*. München, Germany: Oldenbourg.
- 10. Czichos, H., ed. 2004. *Hütte—Die Grundlage der Ingenieurwissenschaften*. Berlin, Germany: Springer.
- 11. Missotten, B. 1998. Measurement systems for the mapping and the evaluation of crop production performance. PhD diss. K.U. Leuven.
- 12. Reyns, P., B. Missotten, H. Ramon, and J. De Baerdemaeker. 2002. A review of combine sensors for precision farming. *Precision Agriculture* 3: 169-182.
- 13. Strubbe, G. 1997. Mechanics of friction compensation in mass flow measurement of bulk solids. PhD diss. K.U. Leuven.
- 14. Technische Sensoren. 1983. Forschungsheft 104. Frankfurt, Germany: Forschungskuratorium Maschinenbau.
- 15. Tränkler, H. R., and Böttcher. 1992. Information processing in sensing devices and microsystems. *IFAC-Symposium on Intelligent Components*, 20-22. Malaga (Spanien), Mai 1992.
- 16. Miller, R. W. 1996. *Flow Measurement Engineering Handbook*. 3<sup>rd</sup> ed. New York, NY: McGraw-Hill.
- 17. Baker, R. C. 2000. *Flow Measurement Handbook*. Cambridge, UK: Cambridge University Press.

Y. Li

Abstract. A biosensor basically consists of a biosensing material and a transducer and can be used for detection of biological and chemical agents. Biosensing materials, including enzymes, antibodies, nucleic acid probes, cells, tissues, and organelles, are able to selectively recognize target analytes and transducers, including electrochemical, optical, piezoelectric, thermal, and magnetic devices, can quantitatively monitor the biochemical reaction. Biosensors, integrated with new technologies in molecular biology, microfluidics, and nanomaterials, have applications in agricultural production, food processing, and environmental monitoring for rapid, specific, sensitive, inexpensive, in-field, on-line and/or real-time detection of pesticides, antibiotics, pathogens, toxins, proteins, nutrilites, odors, microbes, and more in plants, animals, foods, soil, air, and water.

**Keywords.** Biosensors, Biosensing materials, Transducers, Biological agents, Chemical agents.

#### 2.3.1 Introduction

Biosensors have been studied and developed for the past forty years, and they have drawn more attention in recent years for rapid detection of biological and chemical agents in agricultural production, food processing, and environmental monitoring as well as clinical diagnostics, pharmaceutical tests, bioprocessing, biowarfare, and anti-bioterrorism. In this section we will briefly review the history and development of biosensors and their definitions, and classify biosensors based on their biosensing materials and transducing methods. We will summarize the applications of biosensors in agriculture, food, and the environment, and finally discuss some commercial biosensor products.

#### 2.3.2 Definition of Biosensors

The term *biosensor* is used in diverse ways, but, generally, a biosensor should selectively, rapidly, and continuously respond to a biological or chemical agent, and should have these features:

- It includes a bioactive substance, or a biosensing material;
- This material is able to specifically recognize the species of interest, or an analyte; and
- The biosensing material is in close contact with a transducing device.

In general, a biosensor is a device or instrument comprising a biological sensing material coupled to a chemical or physical transducer which converts a biological, chemical, or biochemical signal into a quantifiable and processable electrical signal [1-4].

As shown in Figure 1, biosensing materials used in a biosensor include enzymes, antibodies, nucleic acids, whole cells, receptors, tissues, organelles, and more. A

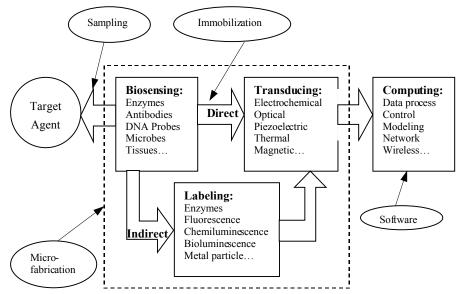


Figure 1. Typical structure and components of a biosensor.

54 Chapter 2 Hardware

transducer for a biosensor may be electrochemical (including voltammetric, amperometric, potentiometric, conductive, capacitive, impedance), optical (including absorption, surface plasmon resonance, chemiluminescence, bioluminescence, fluorescence, optical fiber), piezoelectric (quartz crystal microbalance, surface acoustic wave), calorimetric, magnetic, and others. When the biosensing material can provide detectable signals to a transducer, it can be directly connected to the transducer to form a label-free biosensor. For some applications, labels may be required to amplify biological signals, and they can be enzymes, fluorescence, chemiluminescence, bioluminescence, and metal particles, specifically nanoparticles. A computing unit usually is required for data acquisition and control, database and modeling uses, a network connection, and wireless communication. It is necessary to employ effective immobilization techniques to link the biosensing materials to transducers. The sampling process is critical in separation and concentration of target analytes to get specific and detectable signals. Advanced microfabrication and software are required to approach to a miniaturized and automatic biosensor.

A biosensor can be a system covering sampling, sensing, transducing, and computing. Biosensor technology is based on an interdisciplinary approach encompassing biology, chemistry, and engineering.

## 2.3.3 History of Biosensors

The first biosensor was described by Clark and Lyons [5] in their research on an enzyme-electrode, in which an oxido-reductase enzyme was held next to a platinum electrode in a membrane sandwich (Figure 2). The platinum cathode polarized at -0.7 V responded to the peroxide produced by the enzyme reaction with a substrate. The primary target substrate for this system was glucose. The glucose oxidation reaction, catalyzed by glucose oxidase, can be expressed as:

$$glucose + O_2 + H_2O \xrightarrow{glucose \ oxidase} gluconic \ acid + H_2O_2$$
 (1)

At the electrode:

$$O_2 + 2e^{-} + 2H^{+} \longrightarrow H_2O_2$$
 (2)

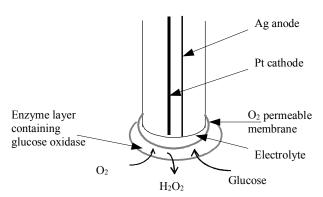


Figure 2. Schematic diagram of a Clark enzyme electrode for glucose detection.

The voltage applied between the platinum and silver electrodes is sufficient to reduce the oxygen, and the electrical current, which is proportional to the oxygen concentration, can be measured. The concentration of glucose is then proportional to the decrease in the current. The electrode is covered with an oxygen permeable membrane, such as polythene or cellophane. A layer of enzyme (glucose oxidase) is placed over this membrane and held in place with a second membrane such as acetate.

This led to the first biosensor product of Yellow Springs Instruments (Model 23YSI, Yellow Springs, OH), which appeared on the market in 1974. The same principle and design were then applied to many other oxygen-mediated oxido-reductase enzyme biosensors. At that time, a biosensor was usually called an *enzyme electrode* or *bioelectrode*, because an enzyme was commonly used as the biosensing material, or biorecognition element, by which a specific biological or chemical event could be detected either directly or indirectly through an enzymatic reaction.

The next major innovation in the development of biosensors was the potentiometric urea electrode reported by Guilbault and Montalvo [6], in which the biorecognition molecule is retained in the vicinity of the base sensor behind a dialysis membrane. This type of electrode was classified as the first generation of biosensors [7].

For the second generation of biosensors, immobilization of biosensing materials was achieved using cross-linking reagents or bifunctional reagents at a modified transducer interface or incorporating them into a polymer matrix on the transducer surface. Typically, the transducer surface was modified by chemicals followed by attachment of the biosensing material. The ELISA (enzyme-linked immunosorbent assay) electrodes belong to this group.

For the third generation of biosensors, the biomolecule becomes an integral part of the biosensing material, such as in SPR (surface plasmon resonance) biosensors. The fourth generation of biosensors may be expected to have more features with MEMS/NEMS/BioNEMS (micro-, nano- or bionano-electromechanical systems), nanotechnology and biotechnology.

Biotechnological advances, including better understanding of biomolecules and biomolecular interactions, make biosensor technology encompass molecular recognition and amplification in many biochemical reactions by combining them with signal processing and transmission capabilities of electronics and optical technology. With contributions from scientists and engineers in different areas, the biosensor has emerged into a new multidisciplinary area, including chemistry and biochemistry, physics, biology, computers and engineering, with great opportunities for developing innovative biosensors. Engineers play the principal role in design and fabrication of biosensors as well as in biosensor applications.

The major application of biosensors was first focused on clinical diagnosis. The glucose sensors are the most successful and widely used biosensors even today. Then, biosensor technology was applied to the analysis of blood samples, diagnosis of infection diseases and drug screening. Further, biosensors have been adopted for detection of biological and chemical agents in agricultural production, food analysis, and envi-

ronmental monitoring as well as in mining, bioprocessing, biowarfare, and homeland security. The research on biosensors has been conducted intensively in many disciplines of both science and engineering for the past decade. After more than 40 years of development, biosensors have become more powerful devices and instruments for monitoring bio-specific interactions and detecting biological and chemical agents in all areas.

Many review articles [8-16] and some books [1,17-27] have presented in detail the definition, history, and development of biosensor technology and its applications and commercial products.

### 2.3.4 Classification of Biosensors

Biosensing technologies consist of biosensing materials, transducing devices, and immobilization methods, and they are developed through multidisciplinary research in biology, chemistry, and engineering. Therefore, biosensors can be classified in different ways based on either biosensing materials used or transducing devices applied.

### Classification Based on Biosensing Materials

The biosensing materials used in biosensors can be first generally divided into three distinct groups based on their mechanisms: *biocatalytic*, *bioaffinity*, and *microbebased* [12]. The biocatalytic group includes enzymes, the bioaffinity group comprises antibodies, receptors, and nucleic acids, while the microbe-based group contains microorganisms, cells, organelles, and tissues.

Enzyme Sensors

Enzymes are molecules that determine the pattern of chemical transformations, and they are the catalysts of biological systems. They also mediate the transformation of different forms of energy. As Stryer [28] described, the most important characteristics of enzymes are their immense catalytic power to accelerate reactions by factors of at least a million and their high specificity both in the reaction catalyzed and in their choice of reactants. Furthermore, the catalytic actions of many enzymes are regulated. Nearly all known enzymes are proteins that are highly effective in catalyzing diverse chemical reactions. Enzymes accelerate reactions by stabilizing transition states and formation of an enzyme-substrate complex is the first step in enzymatic catalysis.

Enzyme catalysis mechanism can be explained as follows:

$$S + E \xrightarrow[k_{-1}]{k_1} ES \xrightarrow{k_2} E + P \tag{3}$$

where S = substrate

E = enzyme

ES =enzyme-substrate complex

P = product

 $k_1$  = rate of enzyme-substrate complex formation

 $k_{-1}$  = rate of enzyme-substrate complex dissociation

 $k_2$  = rate of the dissociation of the enzyme-substrate complex to products

Enzymes are the most commonly used biosensing material in biosensors. For example, enzyme biosensor for measurement of glucose is the most extensively studied and the most commercially developed one. The original form involved the oxidation

of glucose by oxygen, catalyzed by glucose oxidase, to give gluconic acid and hydrogen peroxide (Equation 1).

More than 2,500 enzymes have been identified in nature and most of them are now commercially available. Three major enzymes among more than 20 different enzymes commonly used in biosensors are alkaline phosphatase (ALP), horseradish peroxidase (HRP), and  $E.\ coli\ \beta$ –D-galactosidase (BG).

Immobilization methods used for enzyme biosensors can be classified into two major groups: (1) adsorption or physical entrapment of the enzyme through van der Waals forces, ionic binding, or diffusion barriers; and (2) covalent binding of the enzyme to the transducer through the reaction between functional groups of the protein and the support materials [29]. However, there is often a loss of activity when enzymes are immobilized on the surface of a transducer. As proteins, they are denatured by known factors such as temperature, acids, bases, organic solvents, detergents, salts, and others, weakening their catalytic activities.

In enzyme sensors, enzymes are commonly used with amperometric, potentiometric, chemiluminescence, and thermal transducers. For example, an enzyme-based manometric biosensor was designed for on-line measurement of milk urea [30], an enzyme amperometric biosensor was reported for determination of carbohydrates in foods [31], and an ISFET based enzyme biosensor was used for analysis of potato glycoalkaloids [32].

#### Immuno-Sensors

Antibodies represent one of the major classes of proteins. They constitute about 20% of the total plasma protein and are collectively called immunoglobulins (Ig), which are a group of glycoproteins present in the serum and tissue fluids of all mammals. They are produced in response to specific antigens and counteract their effects by neutralizing toxins, agglutinating bacteria or cells, and precipitating soluble antigens. The functions of antibodies that mediate humoral immunity can be divided into two groups: (1) specific binding to pathogens or to toxins, and (2) interactions with cellular or molecular components of the host's immune system [33].

IgG is the predominant immunoglobulin in normal human serum accounting for 70% to 75% of the total immunoglobulin pool. IgG is the only immunoglobulin capable of crossing the placenta in humans. IgG is distributed evenly between the intra-and extravascular pools, is the major antibody of secondary immuno-response, and the exclusive anti-toxin class. Because of its relative abundance and excellent specificity toward antigens, IgG is the principle antibody used in immunosensors as well as immunological research and diagnostic products.

In the basic four-chain structure of an immunoglobulin, two identical light (L) polypeptide chains and two identical heavy (H) polypeptide chains are linked together by disulfide bonds. The antigen binding site is at the N-terminal end of the molecule. The arms of these Y-shaped molecules have a great degree of flexibility and can operate independently. Heterogeneity in antibody molecules is due to isotypic (different heavy and light chain classes and subclasses), allotypic (variation mostly in the constant region), or idiotypic (variation only in the variable region) variations and is ge-

netically controlled [34]. The primary function of the antibody is to bind the antigen. In addition to their direct neutralizing effects in such case (e.g., on bacterial toxin or viral penetration of cells), they perform several effector functions. Organisms develop antibodies (Ab) that can bind with an invading antigen (Ag) and remove it from harm.

$$Ab + Ag \longrightarrow Ab \cdot Ag$$
 (4)

Antibodies bind more powerfully and specifically to the corresponding antigen than enzymes do to their substrates. They can be very specific to different strains of the same species, or even different serotypes of the same strain. They are ultra-sensitive, although they do not have the catalytic activity of enzymes. The antibody can be directly immobilized to the surface of an electrode or an optical waveguide in a transducer to provide detectable signal when they bind the analyte. In this case, they are called primary or capture antibodies. They are also used for labeling when they carry labels such as radioisotopes, enzymes, red cells, fluorescent probes, chemiluminescent probes, metal tags, or nanoparticles. Then they are named as the secondary or detection antibodies.

Immunoassays are defined as techniques based on the reaction between an antigen and an antibody for measuring the concentration of either reactant in solution. The immunoassays can be homogeneous or heterogeneous. A homogeneous system does not require separation of free and bound antigens, and just relies on the alteration of the electroactivity or function of the label on formation of antibody-antigen complex. A more sensitive approach is the heterogeneous assay format in which there is a separation step.

As illustrated in Figure 3, there are three basic formats of immunoassays, which are also used in design of immunosensors:

- Direct immunoassay, in which antibodies (or antigens) are immobilized on a transducer surface, and the analyte in a sample binds to the immobilized antibodies:
- 2. Sandwich immunoassay, in which the antibodies are immobilized on a transducer surface, the analyte in a sample binds the immobilized antibodies (the primary or capture antibodies), and then the labeled secondary antibodies (or detection antibodies) bind the analyte; and
- Competitive immunoassay, in which the antibodies are immobilized on a transducer surface, and the analyte in a sample and the labeled analyte analog together bind the immobilized antibodies.

The response signal increases in the direct and sandwich formats, but decreases in the competitive format as the concentration of the analyte increases. The sandwich format of the immunoassays can be used only for macromolecular antigens with at least two epitopes for the two binding processes needed to form the sandwich complex.

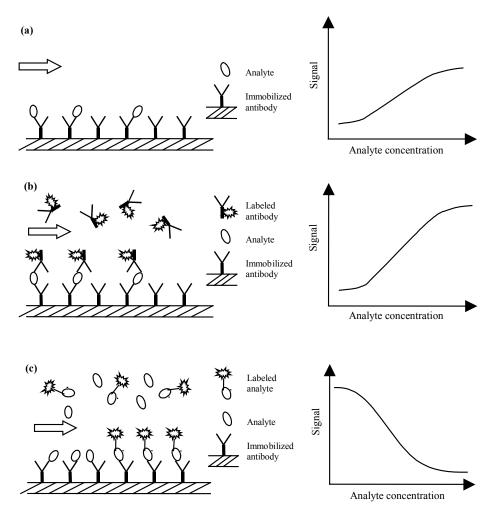


Figure 3. Different formats and corresponding response signals of immunoassays: (a) direct assay, (b) sandwich assay, and (c) competitive assay.

Antibodies, as either primary or secondary antibodies in a biosensor, are commonly coupled with electrochemical, optical, and piezoelectric transducers. Examples include an antibody-based optical biosensor that was developed for on-line measurement of progesterone in bovine milk [35, 36], a label-free QCM immunosensor that was studied for rapid detection of *E. coli* O157:H7 and *Salmonella typhimurium* in food samples [37], and an optical immunosensor that was reported for detection of antibiotics and pesticide residues in water [38].

#### Nucleic Acid Probe Sensors

All nucleic acids (DNA or RNA) are composed of a sugar or derivative of a sugar, phosphoric acid, and a base, and found in cell nuclei. The nucleotide/nucleoside family contains two main branches: those based on the sugar ribose and those based on 2-deoxyribose. Both branches of mononucleotides are strong acids due to a phosphate acid group that has pKa values between 1 and 6. Mono-, bi- and tri-phosphates all occur. In particular the derivatives of the base adenosine (AMP, ADP, and ATP) are important as phosphate transfer agents. The tri- and bi-phosphates also perform the functions of a covalently bound carrier for specific building blocks [18]. A single-strand nucleic acid molecule is able to recognize and bind (hybridize) to its complementary partner in a sample, which can be used in a biosensor as gene probes. Therefore, a nucleic acid probe is a segment of nucleic acid that specifically recognizes and binds to a nucleic acid target, depending upon the formation of stable hydrogen bonds between the two nucleic acid strands. The length of DNA probes can be from tens to several thousands of nucleotides, but usually lengths of less than 30 nucleotides are used in the design of DNA probes [39].

Very recently, *aptamers*, which are specific nucleic acids selected from random sequence pools, have been used as biosensing material in biosensors. Aptamers are able to bind non-nucleic acid targets, such as small molecules or proteins with high affinity [40]. Aptamers can also be selected against target haptens, such as toxins or prions. Aptamers are now considered a valid alternative to antibodies or other bio-mimetic receptors in the development of biosensors for applications based on molecular recognition.

Nucleic acid probes can be used to examine the genetic makeup of an individual and reveal the presence of genes or mutant genes associated with genetically determined diseases. They can also be used for determination of pathogenic bacteria or viruses in samples of water, foods, plants, or animals.

It is important to have samples treated for nucleic acid extraction and nucleic acid sequence amplification. Detergents and NaOH are used to lyse cells and denature proteins as well as double-strand nucleic acids to make the single-strand target DNAs available for hybridization. Theoretically, DNA probes can detect nucleic acids at picogram level, but this is not sensitive enough in many applications such as for the detection of pathogens in foods or environment. Thus, the target nucleic acids need to be amplified to reach a detectable level. One way is to detect naturally amplified target, ribosomal RNA target molecules [41], and another way is the PCR (polymerase chain reaction) amplification method [42].

Usually, either the targets or the probes are immobilized on the surface of a transducer using nitrocellulose, nylon, polyvinylidene difluoride, or synthesizing oligonucleotides with amino, thil, or bitin groups, and then hybridizing with the probes or targets under a controlled condition. The DNA probe may be labeled with an enzyme, fluorescence, or a hapten. A DNA probe biosensor may detect as low as 1 pg of target DNA.

DNA, RNA, and aptamer probes are typically associated with electrochemical, optical, piezoelectric, and magnetic transducers in a biosensor. For example, aptamers were labeled with fluorescence and used in a chip-based biosensor for simultaneous detection and quantification of individual proteins in complex biological mixtures

[43], a nanoparticle-based DNA biosensor was reported for visual detection of genetically modified organisms in main transgenic crops [44], and a microsystem-based DNA-probe biosensor was developed for detection of viable *E. coli* in food and water monitoring [45].

### Microbe-Based or Cell-Based Sensors

Microorganisms are used as indicators for chemical components and other living things. The first microbial sensor was developed by Davis [46] for determination of ethanol using cells of *Acetobacter xylynum*. A microbial biosensor consists of immobilized viable microbial cells in conjunction with a transducer based on the respiratory and metabolic functions of the cell. The analyte to be monitored can be either a substrate or an inhibitor of these processes. Therefore, microbial biosensors are classified as sensing either respiration activity or electrochemically active metabolites [47].

Figure 4 shows the principle of cell-based biosensors of the respiration activity type. Changes in respiration activity of microorganisms caused by assimilation are detected by an oxygen electrode, and then from these changes, the substrate concentration can be estimated based on the current measurement ( $I_0$  -  $I_s$ ). Aerobic microorganisms are used in this type of biosensor. A microbial biosensor is dipped into buffer solution saturated with oxygen. Upon the addition of a substrate, the respiration activity of microorganisms is increased, which causes a decrease in oxygen concentration near the membrane. The electrochemically active metabolite type of microbial biosensors detect electrochemically active metabolites, such as  $H_2$ ,  $CO_2$ ,  $NH_3$ , and organic acids, which are secreted from microorganisms. Not limited to aerobes, this type of cell-based biosensor can use anaerobic microorganisms. Bioluminescence-based microbial biosensors may be developed using genetically engineered microorganisms constructed by fusing the *lux* gene with an inducible gene promoter for toxicity and bioavailability testing [48].

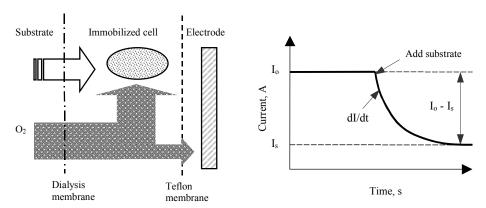


Figure 4. Principle of the respiratory activity type of cell-based biosensors. (left) Schematic drawing of the biosensing concept, and (right) change in the current measurement, where I<sub>0</sub> is initial current and I<sub>S</sub> is the current after a substrate is added.

Microbes or viable cells are mainly used with an amperometric, potentiometric, or impedance transducer to form a cell-based biosensor. Cell-based biosensors mainly are developed for BOD measurement in environmental monitoring and detection of pathogens and toxins in food safety. For example, a microbial fuel cell based biosensor was used to continuous determination of BOD [49], a glass microchip and cultured cell-based biosensor was studied for detection of lipopolysaccharide [50], and an optical microbial biosensor for detection of methyl parathion pesticide using *Flavobacterium* sp. whole cells on glass fiber filters [51]. Several review articles describe whole-cell based biosensors and their potential applications in cell biology, toxicology, pharmacology, and environmental measurements [52-54].

# Tissue-Based and Organelle-Based Sensors

Tissue materials from plant and animal sources have been used as biosensing materials in biosensors. The first tissue-based sensor was developed by Rechnitz [55] for the determination of arginine. It used a thin slice of bovine liver and an aliquot of the enzyme urease. Subcellular organelles such as membranes, respiratory chain, chloroplasts, mitochondria, and microsomes that carry out essential cell functions have been used in some biosensors to recognize specific analytes. The tissue- or organelle-based biosensors are able to provide higher stability than enzyme biosensors, but usually need longer detection time and do not have higher specificity. Enzyme inhibitors, activators, and stabilizing agents are used to improve the selectivity and lifetime of tissue-and organelle-based biosensors.

Wijesuriya and Rechnitz [56] presented a broad review of biosensors based on plant and animal tissues for the detection of various important analytes including drugs, hormones, toxicants, neurotransmitters and amino acids. Receptors isolated from animals and plants are immobilized on transducers including ISFETs (ion-selective field effect transistors), electric capacitors, and optical fibers for biosensors to obtain short response time, a high sensitivity, a wide range of linear response, and inherent selectivity. In recent reports, for example, *Malva vulgaris* tissue homogenate was used as biosensing material in an enzyme amperometric biosensor for detection of sulfite in food samples [57], porcine kidney was used in a chemiluminescence biosensor to detect lactic acid [58], and mushroom tissue homogenate was immobilized onto the electrode of an amperometric biosensor to determine phenolic compounds [59].

### Classification Based on Transducing Methods

Based on the formats of interfaces between biosensing materials and transducing devices, biosensors can be divided into two general categories:

- Direct biosensors (or label-free biosensors), in which physical and chemical signals directly represent the existence and amount of an analyte (e.g., glucose, bacteria, ammonia), using any of the transducers, mostly enzyme electrodes, impedance, optical fiber, surface plasmon resonance (SPR), surface acoustic waveguide (SAW), or QCM transducers.
- *Indirect biosensors* (or *labeled biosensors*), in which the chemical reaction caused by the existence of analytes can be detected through the labels (e.g., enzymes, fluorescence, metal particles) that amplify biochemical signals, using

various transducers including electrochemical, impedance, optical, field-effect transistor (FET), QCM, calorimetric, and magnetic transducers.

Clearly, the same biosensor can be used either directly or indirectly for different applications. Based on the different transducing devices used, we can now further classify biosensors into electrochemical, optical, piezoelectric, thermal and magnetic biosensors.

#### Electrochemical Biosensors

Electrochemical biosensors are the oldest and best-developed biosensors compared to other types. In the early stages, biosensors were mainly enzyme electrodes developed exclusively for the clinical analysis of glucose. Then the enzyme-linked immunoelectrochemical (IEC) assay was developed by Heinemann et al. [60] to enhance the sensitivity of electrochemical biosensors. The effect of the electrical properties of a buffer (caused by enzymatic reaction or Ab-Ag interaction) can be measured by various electrochemical methods. Recent research in electrochemical biosensors has focused on the improvement of electrode design (e.g., miniaturization, efficient electron transfer, nanomaterials, and better immobilization procedures). Electrochemical biosensors can be further divided into amperometric/voltammetric, potentiometric, and conductivity/capacitance/impedance biosensors, each discussed below.

Amperometric/Voltammetric Biosensors—The amperometric and voltammetric biosensors are characterized by their current-potential relationship with the electrochemical system. Amperometric sensors can be viewed as a subclass of voltammetric sensors. In amperometric sensors, a fixed potential is applied to the electrochemical cell, and then a corresponding current is obtained due to a reduction or oxidation reaction. However, a voltammetric sensor can operate in other modes such as linear or cyclic voltammetric modes. Consequently, the respective current and potential response for each mode will be different.

Generally, amperometric biosensors measure the concentration-dependent current through an electrochemical electrode coated with biologically active material. Amperometric transduction is based on the oxidation and reduction of an electroactive species on an electrode surface. The relationship between the electrical current and the concentration of the analyte can be expressed by the Cottrell equation:

$$i = n F A C_0 \left[ D/(\pi t) \right]^{1/2}$$
 (5)

where i = the current to be measured

n = the number of electrons transferred

F = the Faraday constant, 96 487 C per equivalent

A = the area of electrode

 $C_0$  = the concentration of the analyte

D = the diffusion constant

t = the time elapsed since the potential was applied

A lot of work has been done to improve the electron transduction. The fourth generation of amperometric biosensors have been improved by transducing the biological reaction by the oxidation or reduction at the electrode surface; using mediator molecules to transfer the electrons from the enzyme to the electrode, after it reduces or oxi-

64 Chapter 2 Hardware

dizes the substrate; modifying the electrode surface by the addition of molecules, allowing direct oxidation or reaction of the enzyme at the electrode; and applying nanotubes/ nanowires/nanofibers to make the electrode become micro- or nanometers in its size or use interdigitated array microelectrodes.

An amperometric transducer can be coupled with any of the biosensing materials including enzymes, antibodies, DNA-probes, whole cells, and tissues. For example, a redox hidrogel-based amperometric bienzyme biosensors was developed for fish freshness monitoring [61], a bienzyme electrochemical biosensor coupled with immunomagnetic separation was investigated for rapid detection of *Escherichia coli* O157:H7 in food samples [62], and a biosensor based on self-assembling acetylcholinesterase on carbon nanotubes was reported for flow injection/amperometric detection of organophosphate pesticides [63].

Potentiometric Biosensors—Potentiometric measurements involve a non-faradic electrode process that has no net current flow and operates on the principle of an accumulation of charge density at an electrode surface, resulting in the development of a significant potential at that electrode. Potentiometric biosensors, by using suitable bioreceptors and compatible transducers, monitor the changes in electrical potential brought on by the binding of an ion to an ionophore. Potentiometric detection measures the potential across an electrochemical cell containing the biological sensing element, usually by measuring the activity of either a product or a reactant in the electrochemical reaction. The measured potential is given by the Nernst equation:

$$E = E_0 + [RT/(nF)] \ln a \tag{6}$$

where E = the potential to be measured in V

 $E_0$  = the standard potential for  $a = 1 \text{ mol } 1^{-1}$ 

R =the gas constant

T = the temperature in K

F = the Faraday constant

n = the electron transfer number

a = the relative activity of the ion of interest

There are various configurations in which enzymes can be used with potentiometric electrodes for analytical purposes. The field-effect transistor (FET) was proposed by Bergveld [64], which is very suitable for performing an unlabelled immunoassay. The small dimensions of FETs as well as the built-in impedance conversion are very attractive features. There are four types of FETs used for biosensing purposes [18]:

- ISFETs (ion-selective field-effect transistors), which respond to ions in solution;
- ENFETs (enzyme field-effect transistors), in which immobilized enzymes are
  used to measure enzyme substrates or species that are coupled to an enzyme reaction;
- *IMFETs* (*immuno-field-effect transistors*), which generate charge separation via antibody-antigen interaction; and
- SGFETs (suspended-gate field-effect transistors), which are based on the changes in work function and dipole orientation resulting from the interaction of the biosensing material with various gases.

In some cases, to prevent interference by ammonia or carbon dioxide already present in the sample (one only wants to detect the substance produced by the enzyme), the ENFET is accompanied by an ISFET for that species, the signal of the latter being subtracted from that of the former.

Another group of potentiometric biosensors is based on the light-addressable potentiometric sensor (LAPS), which are able to detect relatively large particles [65,66]. Both FET and LAPS transducers can be coupled with any of the biosensing materials to become a potentiometric biosensor. Recently, for example, a LAPS biosensor was used with streptavidin-coated magnetic beads for detecting *E. coli* O157:H7 in foods [67], an anion-selective LAPS was reported for the determination of nitrate and sulphate ions [68], and an enzyme-potentiometric biosensor was developed for direct determination of cyanides in food plants [69]. More information on amperometric and potentiometric biosensors can be found in several review articles [70-72].

Conductance/Capacitance/Impedance Biosensors—Conductance, capacitance, and impedance biosensors measure different changes of an electrical field. Those changes can be overall electrical conductivity of the solution or medium and capacity alteration due to the immobilized layer on the electrode surface, which also can be reflected in impedimetric response. The early conductance/impedance biosensors were based on measuring the conductivity changes in the medium caused by target analytes. However, as the resistance of a solution is determined by migration of all ions that are present, conductance measurements are generally considered to be relatively non-specific. This problem is overcome by monitoring the changes in conductance produced by the catalytic action of enzymes immobilized on or in a planar microelectronic conductance cell [73]. Many enzymes catalyze reactions, which leads to overall changes in solution conductance and thereby display great potential as sensing elements in conductometric biosensors.

Conductance and capacitance biosensors are really the simple versions of impedance biosensors. The relationship between impedance and resistance, capacitance, and inductance is expressed as:

$$Z = R + j X = R + j (X_L - X_C)$$
 (7)

where Z = impedance, which is a complex quantity

R = resistance

X = reactance

 $X_C$  = capacitive reactance =  $(2\pi f C)^{-1}$ , where f = frequency and C = capacitance

 $X_L$  = inductive reactance

j = imaginary unit

Since usually the inductance can be ignored in an electrochemical analysis system, the impedance in Equation 7 contains only resistance, R, and capacitive reactance,  $X_C$ .

Because of impedance microbiology, impedance biosensors have found their applications more to the detection of pathogenic bacteria, specifically viable bacteria. Several types of impedance biosensors are designed in couple with microfluidics and interdigitated array microelectrodes for rapid detection of *Listeria monocytogenes* and *E. coli* O157:H7 [74-77]. A portable impedance based biosensor, which contained a vari-

ety of disposable analyte-specific sensor modules, was developed by Louie et al. [78] for field use. A capacitive field-effect biosensor and a conductive tyrosinase biosensor are reported for directly detecting pesticide residues [79,80]. Some review articles are also available for further information on impedance biosensors [81,82].

## Optical Biosensors

Optical methods are among the oldest and best-established techniques for sensing biological and chemical analytes. A variety of optical techniques have been used to construct biosensors. A typical optical biosensor consists of a light source, a number of optical components to generate a light beam with specific characteristics and to direct this light to a modulating agent, a modified sensing head (optic fiber or crystals coated with antibodies, dyes and proteins, and modified by functional chemical groups), and a photodetector. Similar to our earlier discussion, there are two formats for optical biosensors: direct and indirect detection of the target analyte. In the direct format, the analyte directly affects the optical properties of a waveguide, such as evanescent waves (electromagnetic waves generated in the medium outside the optical waveguide when light is reflected from within) or surface plasmon resonance (an evanescent wave in a thin film deposited on a waveguide surface). In the indirect format, optical labels such as fluorescence, metal particles, or nanoparticles are used to generate optical signals proportional to the target analyte. Optical biosensors may be designed based on optical phenomenon, including adsorption, fluorescence, phosphorescence, polarization, rotation, interference, or on non-linear phenomena such as harmonic generation [12, 83]. Several types of optical biosensors including absorption and reflection, luminescence, fluorescence, SPR, optical fiber, and others are summarized below.

Absorption and Reflection Biosensors—The optical response in absorption is based on the Lambert-Beer Law (Equation 8) that characterizes the intensity of transmitted light though a uniform medium as a function of the incident light when the optical properties are affected by the chemical concentration. The absorbance, A, is expressed as:

$$A = -\log(I/I_0) = \epsilon C l \tag{8}$$

where I = the intensity of transmitted light

 $I_0$  = the intensity of incident light

 $\varepsilon$  = the extinction coefficient in M<sup>-1</sup> cm<sup>-1</sup>

C = the concentration of analyte in M

l = the path length of light through the medium in cm

A spectrum of light is absorbed in a range of wavelengths, which is different for each chemical species. The signal of light intensity for a specific chemical species usually is compared to a reference signal that is measured only with the background. Usually, absorption measurement is adopted in a biosensor with sandwich format in which the enzyme labels are used to react with a substrate to change the optical absorbance of the medium. However, the absorption sensors may be disturbed by light scattering, external light, internal light reflection from the surface of sample chamber, and the light penetration limit. These may be overcome by measuring the light reflected back from a surface or from a deeper layer of a medium. The change in the intensity of reflected light is proportional to physical or chemical event associated with the target

analyte. In most applications, absorption and reflections methods are coupled with optic fiber, fluorescent capillary-fill devices (FCFD), fluorescence, or total internal reflection fluorescence (TIRF), which will be briefly discussed later.

Absorption/reflection transducers are used with enzymes, antibodies and DNA/RNA probes in a biosensor. Recently, an absorption biosensor based on enzyme immunoassay was developed for detection of penicillin residues in meat and dairy products [84], an absorbance biosensor coupled with immunomagnetic separation was used for detection of *Escherichia coli* O157:H7 in foods [85], and a disposable receptorbased absorption biosensor was reported for detection of nitrate in water samples [86].

Luminescence Biosensors—Luminescence occurs when molecules emit light while they return to the ground state after being excited. Chemiluminescence and bioluminescence have been used in different types of biosensors, such as enzyme, antibody, DNA-probe, and cell-based biosensors. The light emission of chemiluminescence is determined by chemical reaction, whereas bioluminescence is from living organisms (bacteria, fish, insects, and fungi) with reactions catalyzed by enzymes. In a luminescence biosensor, a sensing layer with immobilized biosensing materials is able to recognize specific target analyte in a sample, and after light emission starts the light is transmitted through a waveguide to a photo detector.

A luminescent enzymatic sensor is usually able to sensitively measure ATP, NAD(P)H, or  $H_2O_2$ , and other analytes can be detected by more complex systems involving auxiliary enzymes working sequentially with the luminescent system [87]. For example, to detect  $H_2O_2$ , the chemiluminescent compounds mostly used are luminol (5-amino-2,3-dihydro-1,4 phthalazinedione) and related hydrazides. Luminol-mediated chemiluminescence in the presence of horseradish peroxidase (HRP, EC 1.11.1.7) is able to detect hydrogen peroxide as follows:

$$2H_2O_2 + Luminol + OH^- \xrightarrow{HRP}$$

$$3 - aminophthalate + N_2 + 3H_2O + hv (\lambda_{max} = 430 nm)$$
(9)

The  $\rm H_2O_2$  measurement can be done within a minute using peroxidase in a sensing membrane with a detection limit of  $10^{-8}$  M, and similarly the detection limit can be as low as  $10^{-11}$  and  $10^{-9}$  M for ATP and NADH, respectively [88]. Since chemiluminescent signals decay very rapidly, the substrate needs to be rapidly mixed with chemiluminescent reagent, and the photodetector should have a suitable bandwidth to record the signal.

Luminescence biosensors are designed more often using enzymes, antibodies, viable microbes, engineered cells, or a combination of two of them as biosensing materials. They have been used for detection of pesticides, pathogens, and food components. For example, a bioluminescence and cell-based sensor was investigated for detection of toxicities of hydrogen peroxide, phenol, and mitomycin C in water samples [89], an immunomagnetic chemiluminescence fiber-optic biosensor was developed for rapid detection of *E. coli* O157:H7 in ground beef, chicken carcasses, and lettuce samples [90], and an ATP-bioluminescence biosensor was used to bioenergetically confirm viable pathogens in foods [91].

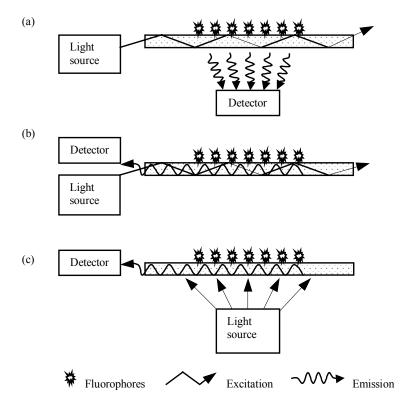


Figure 5. Three formats of fluorescent biosensors with use of the evanescent wave:
(a) excitation through the waveguide, (b) excitation and emission through the waveguide,
and (c) collection of the emission through the waveguide.

Fluorescence Biosensors—Fluorescence occurs in certain molecules, such as fluorescent dyes, fluorophores, and fluorochromes, when an external light source is applied. The emission signal is typically weaker and with a longer wavelength than the excitation light. The fluorescence response or emission occurs instantaneously with the light source, or excitation. In most fluorescence biosensors, the combination of biosensing materials with an evanescent wave at a surface provides the specificity. As described by Liley [83], there are mainly three formats in using evanescent wave to detect surface-bound fluorescent labels: excite the fluorophores by using the evanescent wave (Figure 5a), excite the fluorescence and collect the emission through the waveguide (Figure 5b), and collect the emitted light by the optical waveguide (Figure 5c).

Fluorescence biosensors are able to measure surface-specific binding events in real time. Waveguides can be made of materials that have desired optical properties and are easily modified for immobilization of biosensing materials. Sensor design is adaptable owing to the wide variety of visible and near-IR light sources and detectors. Additionally, the systems described in detail here, fiber optic and planar array fluores-

cence sensors that utilize the evanescent wave for excitation of fluorescently tagged reporters, gain improved discrimination of specific binding from non-specific adsorption of sample components [92]. However, the biointeractions outside of the evanescent field are not detectable since they are not covered by the evanescent field. This may make it very difficult to detect large targets such as cells. Furthermore, below a critical flow rate, mass transport may limit the binding of target analyte to the immobilized biosensing materials. Fluorescence transducers are widely used with all types of biosensing materials to form a fluorescence biosensor. Recently, based on fluorescence labeled antibodies and total internal reflection fluorescence, a portable optical immunosensor was developed for multiple analytes including pesticides monitored in water pollution control [93]. Green fluorescent protein mutants were used in a fluorescence biosensor for detection of bacterial endotoxin [94], and quantum dots were used as fluorescent labels in an immunosensor for quantitative detection of Salmonella typhimurium in chicken carcass wash water [95]. Several review articles covered the background, development, and applications of fluorescence biosensors including evanescent wave fluorescence biosensors [92], fluorescence-based glucose biosensor [96], and live-cell fluorescent biosensors [97].

SPR Biosensors—The surface plasmon resonance (SPR) transduction is widely used as an analytical tool for measuring small changes in the optical refractive index (RI) of thin biological films on a sensor interface. SPR is the result of optical excitation of a surface plasmon wave along the interface between a highly conductive metal and a dielectric material. The conditions for excitation are determined by the permeability of the metal and the sample materials, as well as the wavelength and angle of incident light (Figure 6). The resonance angle is sensitive to changes in refractive index and dielectric constant at the interface up to a distance of 800 nm from the actual metal surface, with an exponential fall in sensitivity with distance from the surface, which makes the SPR based biosensor work better on small particles [98]. The change

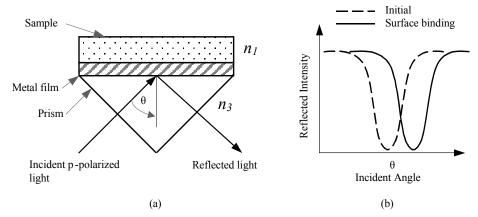


Figure 6. Principles of SPR biosensors. (a) Three-layer geometry for exciting surface plasmons.

A surface plasmon wave is excited on the metal interface. (b) SPR response.

70 Chapter 2 Hardware

in RI is related to the on-going biochemical reaction occurring on the sensor surface. Therefore, an antigen-antibody system and a DNA complimentary fragment system are employed to construct an SPR-based biosensor. The changes on the sensor surface can be quantitatively defined in several ways by measuring the changes in the wave length of the incident light, the incident angle, the intensity of reflection, and the reflected phase as the refractive index of sample changes, while keeping other parameters fixed.

The advantages of SPR biosensors are that SPR allows the measurement of the biochemical interactions in real time with a high degree of sensitivity; and that no labeling of the analytes is necessary for their detection. However, SPR, just like most direct-detection methods that rely on surface adsorption, cannot distinguish between specifically and non-specifically adsorbed molecules of similar size [99].

Currently, commercially available SPR-based biosensor systems developed by Biacore (Uppsala, Sweden) show considerable promise in enabling determination of varieties of biological and chemical agents in different samples. Other companies such as Texas Instruments (Dallas, TX) and Nycomed Amersham (Buckinghamshire, UK) provide SPR sensors that can be easily modified into biosensors by coating biosensing materials onto the waveguide. SPR-transducing technology has been used with many different biosensing materials in design of SPR optical biosensors. For example, a miniaturized surface plasmon resonance biosensor was reported for detection of *E. coli* O157:H7 [100], and an SPR biosensor was developed for the simultaneous determination of thiamphenicol, florefenicol, florefenicol amine, and chloramphenicol residues in shrimps [101].

Optical Fiber Biosensors—An optical fiber is used as a plain transducer to guide light to a remote sample and return light from the sample to the detection system. Changes in the intrinsic optical properties of the medium itself are sensed by an external spectrophotometer. An indicator or chemical reagent placed inside or on a polymeric support near the tip of the optical fiber is used as a mediator to generate a detectable optical signal. The advantages of using optic fibers as a transducing method are the easy use in field- and real-time detection, convenience and flexibility, potentially long interaction length, and low cost.

One type of fiber optic biosensor is the planar waveguide optical biosensor. This biosensor is based on the detection of evanescent effects. Evanescent waves (EW) are set up when light traveling through a wave guide, in direct contact with a solution, undergoes total internal reflection at the wave guide surface. The EW propagates into the solution and decays exponentially with distance from the surface-solution interface. The binding of labeled antibodies within the evanescent field can be related to the concentration of sample antigen. The change measured may be absorbency, fluorescence, or light scattering.

Fiber optic biosensors have shown very successful applications in food processing and environment monitoring and they work very well for both pesticide residues and bacteria cells. The narrow penetration depth of fluorescent light consequently requires high capture efficiency of target bacteria. However, the contacting area of an optic fiber (usually very small) and the complexity of a food sample make no guarantee of

good capture efficiency. Another barrier that should be overcome is the regeneration of the fiber tip in practical applications.

Other Optical Biosensors—The fluorescent capillary-fill device (FCFD) developed by Robinson et al. [102] was a typical construction format of the EW based biosensors. This device consists of two glass plates separated by a narrow gap of about 100 µm. The lower plate acts as an optical wave guide and contains on its surface an immobilized layer of antibodies. This biosensor benefits from the capillary fill system, which delivers a highly reproducible volume of samples with easy use. Since all the reagents are contained within the FCFD device, a user just needs to dip the biosensor into the sample to incubate it with the reagent. The FCFD is used to assay both small and large analytes for a wide range of sample matrices without any need for sample pretreatment.

Microcontact printing provides an alternative for antibody patterning processes, which results in a novel optical biosensor. An antibody-grating pattern was stamped on a silicon surface. The antibody grating alone produces insignificant optical diffraction, but upon immuno-capture of analytes, the optical phase change produces a diffraction pattern. This technique eliminated much of the surface modifications and the secondary immunochemical or enzyme-linked steps that are common in immunoassay. Another type of optical biosensor is the resonant mirror (RM) biosensor that is based on total internal reflection. It combines the enhanced sensitivity of waveguide device with the facile construction and operation of an SPR biosensor [103].

Total internal reflection fluorescence (TIRF) is used with planer and fiber optic waveguides as transducers in optical biosensors. Light is propagated down a waveguide that generates an evanescent wave on the surface of the optically denser medium of the waveguide and the adjacent, less optically dense, medium. The signal of the standing wave decreases exponentially with the distance to the lower refractive index material. Biosensing materials can be immobilized on the side of the waveguide and the fluorescence excited within the evanescent field can be collected outside the waveguide. TIRF is used in FCFD and resonance mirror devices.

Books about optical and fluorescence biosensors include [104] and [105].

### Piezoelectric Biosensors

Piezoelectric biosensors have been developed based on established theories in electricity, mass, and viscoelasticity, using commercially available instruments such as the quartz crystal microbalance. Piezoelectric sensors show their advantages over other sensors in terms of sensitivity, versatility, low cost, and simplicity, and they are label free.

A typical piezoelectric sensing head consists of a quartz crystal wafer and two excitation electrodes plated on opposite sides of the crystal (Figure 7). The wafer is cut from a natural or synthetic crystal of quartz. The electromechanical coupling and stresses resulting from an applied electric field depend on the crystal symmetry, cut angle, and electrode configuration. Different modes of electromechanical coupling lead to different types of acoustic waves including thickness shear mode (TSM), surface acoustic wave (SAW), shear horizontal (SH) SAW, SH acoustic plate mode (APM), and flexural plate wave (FPW).

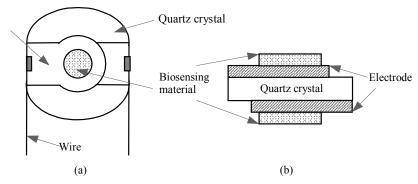


Figure 7. Structure of a piezoelectric crystal sensing head used in a QCM biosensor, (a) top view and (b) cross-section view.

Piezoelectric biosensors are principally based on the measurement of changes in resonant frequency of a piezoelectric crystal as a result of mass changes on the crystal surface (usually caused by biochemical interaction, such as an antibody-antigen reaction or a DNA fragment and its complimentary sequence). There are two main types of piezoelectric devices: the quartz crystal microbalance (QCM) and the surface acoustic wave (SAW) device. Intensive research on piezoelectric biosensors indicates the potential of these systems as low-cost sensors for one-step detection of biochemical reaction

The QCM, which operates at frequencies below 15 MHz, has been used to study the surface modification of the piezoelectric crystal and subsequent detection of the antibody of interest. The relationship between frequency shift and surface mass change is given by the Sauerbrey equation:

$$\Delta f = -2.3 \times 10^6 f_0^2 \, \Delta M/A \tag{10}$$

where  $\Delta f$  = the change in fundamental frequency of the coated crystal in Hz

 $f_0$  = the resonant frequency of the crystal in MHz

A =the coated area in cm<sup>2</sup>

 $\Delta M$  = the mass deposited in g

For a QCM made from a 10 MHz AT-cut quartz crystal, its sensitivity can be 4 ng cm<sup>-2</sup> Hz<sup>-1</sup>. These devices can operate in liquid phase with a frequency determination limit of 0.01 to 1 Hz, and the detection limit of mass bound to the electrode surface is  $10^{-10}$  to  $10^{-12}$  g.

The SAW device, which operates at a frequency normally above 100 MHz, also applies for biosensors. More complex equations have been derived for different types of SAW devices. Figure 8 shows the basic structure of a SAW biosensor. These devices offer higher sensitivity than the QCM because mass sensitivity is directly related to operating frequency. The SAW device has shown high sensitivity for detection of peptides, DNA sequences, pathogens, and pesticides. However, piezoelectric biosensors have their limitations in detection limits and reusable electrodes

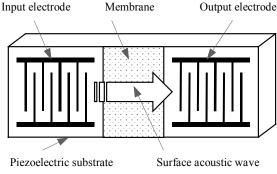


Figure 8. Principle of the SAW biosensor.

One important feature of piezoelectric sensors is that they can be designed as label-free immunosensors. The immunosensors taking advantage of the antibody-antigen affinity reaction are among the most promising biosensors due to their high specificity and versatility. Conventional immunosensors generally involve the formation of a sandwich immuno-complex consisting of the immobilized primary antibody, captured target analyte, and labeled secondary antibody followed by an optical or electrochemical measurement to detect the label directly or indirectly. Piezoelectric immunosensors do not need a labeled antibody and are thus much simpler and easier in operation than sandwich immunosensors.

The first piezoelectric immunosensor was reported by Shons et al. [106], who modified a quartz crystal with BSA (bovine serum albumin) and used it to detect anti-BSA antibodies. Since this, numerous piezoelectric immunosensors have been reported for the detection of various analytes from small molecules to biological macromolecules to whole viruses and cells. In short, a piezoelectric immunosensor is made by immobilizing a specific antibody/antigen on the surface of AT-cut PQC. When the immunosensing surface is exposed to a sample solution, a binding reaction occurs between the immobilized antibody/antigen and its complementary part (target analyte). The binding event is monitored *in situ* by QCM based on the change of surface mass loading and/or other properties such as viscoelasticity, and thus the target species is quantitatively detected.

Piezoelectric genosensors are fabricated by immobilizing a single stranded DNA/RNA probe on the PQC surface. Specific hybridization between the immobilized DNA/RNA probe and its complementary strand in the sample causes a change in the resonant frequency of the QCM. Various methods have been used for the immobilization of DNA probes onto the QCM surface. Among these, the SAM method is most commonly used because it offers an ordered, stable, and convenient immobilization. Thiolated oligonucleotides can directly form an SAM on the gold surface of the QCM electrode via the Au-thiolate bond.

QCM transducers are typically used with antibodies and DNA/RNA probes to form biosensors. Recently, a QCM immunosensor was studied for screening tests of porcine reproductive and respiratory syndrome virus infection in pigs [107], a langasite pure

shear horizontal SAW biosensor was reported for detection of *E. coli* O157:H7 [108], and a QCM immunosensor with simultaneous measurement of resonant frequency and motional resistance was used for detection of *Salmonella* in poultry products [109]. More information on piezoelectric biosensors can be found in [13,110-115].

#### Thermal Biosensors

Thermal biosensors also are called *calorimetric biosensors*. They are developed by incorporating a biosensing material (enzyme, organelle, microorganism, plant or animal cell, or tissue) with a physical transducer such as a thermometer, thermopile, or thermistor. As shown in Table 1, in analytical solution calorimetry the instruments are classified into heat conduction, isoperibol calorimetry, and isothermal calorimetry [116]. The enthalpy change for biochemical reactions is in the range of 25 to 100 kJ mol<sup>-1</sup> [117]. For example, the reaction with glucose catalyzed by glucose oxidase, or with urea catalyzed by urease, releases 80 and 49 kJ/mol, respectively. Generally, isothermal conditions are assumed in the interpretation of biosensor signals. Either the thermal mass of the biosensor is so large that heat can be rapidly dissipated, or the temperature of the entire device is regulated with a circulating water bath [19]. Thermal biosensors have been developed in all three groups, but most thermal biosensors are thermistor-based biosensors, which are based on the measurement of heat that generated during biochemical reactions involving specific enzymes. Two main reasons make thermistor biosensors more successful: the very sensitive miniaturized thermistor and the very easy flow-injection analysis (FIA). Thermistor biosensors require a very sensitive thermistor that can detect changes in temperature of 0.001°C.

Thermal biosensors were reviewed for their use in continuous monitoring of enzymatic processes by Lammers and Scheper [118]. Ramanathan and Danielsson [119] presented a good review on the principles of thermometric measurements with various instruments, materials and methods. They also described applications of thermistor-based calorimetric biosensors to enzyme activity measurements, clinical monitoring, process monitoring and control, multianalyte determination, hybrid sensing, environmental monitoring, and non-aqueous measurements. For example, a sol-gel based enzyme biosensor based on thermometric measurement was reported for detection of glucose in samples of fruit juice, cola, and human blood serum [120], and a flow injection calorimetric biosensor based on immobilized chicken liver esterase was developed for detection of dichlorvos residues in environment and foods [121].

Group	Principle	Sensors
Heat conduction	Measure the voltage of a thermoelectric	Calvet microcalorimeter
	transducer between the reaction vessel	Thermopile sensor
	and a surrounding isothermal heat sink	Integrated silicon thermopile sensor
Isoperibol calorimetry	Measure the temperature change in react-	Thermistor
	ing solution	Miniaturized thermistor
Isothermal calorimetry	Keep the temperature of a reaction vessel constant by heat compensation	Calorimetric microsensor

Table 1. Classification and principles of thermal sensors and instruments.

#### Magnetic Biosensors

Recently, magnetic biosensors have drawn more attention, and they are discussed in a review by Megens and Prins [122]. Miniaturized biosensors, based on sensitive detection of magnetic micro- and nanoparticles in microfluidics channels using the magneto-resistance effect, are promising in terms of sensitivity and size, but the challenge ahead is to integrate the sample pretreatment and detection into a cartridge for totally automatic operation. Magnetic particles are also used as solid supports for bioreaction in microchannels [123], and single bacterial cells can be tapped and transported by the magnetohydrodynamic flow [124].

#### 2.3.5 Immobilization of Biosensing Materials

In order to obtain the response signals of a biosensor, biosensing materials must be associated with a transducer. Physical and chemical methods typically used to immobilize biosensing materials onto transducers are adsorption, micro-encapsulation, entrapment, cross-linking, and covalent bonding [20,24]. Specific procedures of immobilization are dependent upon the surface nature of a transducer, the properties of biosensing materials, and the structure of the biosensor.

Consider enzymes and antibodies. These molecules adsorb on the surface of the metal, metal oxide, carbon, and glass commonly used for transducers due to hydrophobic, ionic, and Van der Waals interactions [23]. Adsorption is the simplest immobilization method with less disruption to the biosensing materials, but the bonding is weak and lifetime is short (several days). Absorbed biosensing material is susceptible to changes in temperature, pH, ionic strength, flow rate, and substrates. Also, adsorption does not promote molecular orientation of the biosensing materials for optimal efficiency in chemical transduction.

In the entrapment method, biosensing material is trapped in the matrix of a gel, a paste, or a polymer near the transducer surface. The gels include polyacrylamide, starch gel, nylon, and silastic gels. Polymetric materials can provide many functions such as a structural framework, selective ion permeability, enhanced conductivity, and mediation of electron transfer processes. The entrapment method does not lead to the orientation of macromolecules on the transducer's surface, and the lifetime is relatively short, in general several weeks.

In the cross-linking method, bifunctional agents, commonly including glutaraldehyde, hexamethylene diisocyanate and 1,5-dinitro-2,4-difluorobenzene, are used to establish intermolecular links between biomolecules for stabilizing the layers and preventing leaching from the reaction layers in the immobilization of biosensing materials onto a transducer [125]. This method does not provide high mechanical strength, and may limit the diffusion of the substrate.

When covalent attachment is used, covalent chemical bonds are formed between the biosensing material and the surface of a transducer. Enzymes, antibodies, carbohydrates, and oligonucleotides have specific molecular characteristics, and proteins are normally bound through the animo, carboxyl, sulfhydryl, or aromatic side chains of the amino acids in the macromolecules. Compared to adsorption, entrapment, and cross-linking methods, covalent binding provides better surface structure and durable surface modification lasting several months.

Mayes [125] presented detailed information on immobilization chemistry of biosensing materials and summarized available techniques. Various immobilization methods are discussed for different biosensing materials (enzymes, antibodies, and other biomolecules) to different transducer materials (gold, glass, silica, metal oxides, carbon, and polymers). The orientation of immobilized antibodies can be controlled using covalent coupling through attached glycosides, specifically located thiol groups, antibody-binding proteins [126], avidin/streptavidin-biotin complexation [127], and tags with engineered antibody fragments. Special control of surface immobilization can be done by patterning the substrate to introduce appropriate surface functionality, control of deposition by physical placement, light-directed immobilization and patterning, and electro-chemical control of deposition. Duschl [128] described the theories and practical approaches to determine binding constants and kinetic rate constants in immobilized biological systems. As mentioned by Cunningham [23], more and more surface modification and immobilization techniques have been developed to meet the needs of optimal loading of selective components, properly structured thin films, proper orientation for efficient biomolecular binding, and patterned surfaces for well-designed miniaturization.

## 2.3.6 Design of Biosensors

In design of a biosensor, main considerations are samples, performance criteria (or specification requirements), operation conditions, and available new technologies (Figure 9).

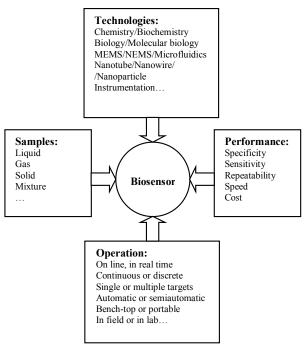


Figure 9. Major considerations in the design of a biosensor.

## Samples and Their Preparation

Sampling technique is critical in successful applications of biosensors. Specifically, there are many known, unknown, organic and inorganic materials in a sample from agricultural or food products, food processing, or the natural environment. At the same time, the concentration of target analytes, such as pesticide residues and foodborne pathogens, existing in an agricultural, food, or environmental sample, is usually very low (several ppb for pesticides and several cfu/mL for bacteria). It is very clear that in the sample pretreatment, the separation of target analytes enhances the specificity and the concentration of target analytes improves the low detection limit. Therefore, it is desired to have sample pretreatment as a part of the instrument of biosensors.

Usually three methods, filtration, centrifuge, and magnetic immunoseparation, are used in a laboratory to separate a target analyte from a sample and then concentrate it for the measurement of biosensors. Magnetic immunoseparation is more suitable to biosensors because of their rapid, easy, automatic operation. As reviewed by Gijs [129], magnetic microparticles and nanoparticles can be magnetically manipulated using permanent magnets or electromagnets. This is the basis of an improved exposure of the functionalized bead surface to the surrounding liquid and of higher sample preconcentration efficiencies, due to the increased relative motion of the bead with respect to the fluid. Magnetically labeled cells can be selected, separated, and aligned under a designed magnetic field [130]. A study indicated that magnetic nanoparticles coated with specific antibodies have better capture efficacy than magnetic microbeads in separating *E. coli* O157:H7 from ground beef and milk samples and they are more suitable to automatic operation and miniaturization [131].

## Performance Criteria

Specificity or selectivity or is the most important factor to evaluate the performance of a biosensor. It is determined by the nature of the biosensing material used in the biosensor, for example, enzymes, antibodies, DNA probes, or microbes. In most cases, it can be measured by false positive rate to a target analyte. Similarly, the false negative rate is also used to help define the specificity of some biosensors. The range of false positive/negative rate is from 0.1% to 5%. It is not practical, at present, to require a biosensor to have a rate less than 0.1%, whereas it is not acceptable if a biosensor shows a rate more than 5%.

A lower detection limit is the second important factor that shows the quality of a biosensor. The lower detection limit can be determined in different ways depending on the nature of analytes, samples, and biosensor structures. One way is based on minimum detectable signals. The detection limit can be also determined based on a defined the ratio of signal to noise, conventionally  $\geq 3$ . In case that both the noise and signal vary a lot, a statistical test may be used to determine the lower detection limit based on significant difference between the signal and noise.

Sensitivity is defined as the ratio of the change in the biosensor output signal over the change in the concentration of target analyte. In indirect biosensors, usually changes in concentrations of co-reactants or co-products of a chemical reaction are measured, not the concentration of target analyte. Thus, a calibration curve should be 78 Chapter 2 Hardware

plotted using the data of the magnitude of output signals and the concentration of target analyte. This can be obtained by exposing the biosensor to a standard solution containing different concentrations of target analyte. Then, the slope of a linear part of the calibration curve is the sensitivity. There are numerous factors affecting the sensitivity of a biosensor, due to different biosensing materials and transducing devices. It is expected that the sensitivity of a biosensor will remain constant all the time and high enough to ensure reliable quantitative results.

Detection time is one of the most important factors to differentiate biosensors from other analytical instruments. The detection time of a biosensor should be counted as that from sampling to the final reading, and its value varies from several seconds to several minutes. For some biosensors, response time and recovery time need to be counted, too. The response time is for the measurement signal of a biosensor to come to equilibrium. Compared to physical sensors or even chemical sensors, biological sensing materials usually increase the response time of a biosensor. The recovery time is the time before a biosensor is ready to be used for the next sample, because the biosensor may need to resume a baseline equilibrium after washing, regeneration of biosensing materials, or changing sensing chips or cartridges.

Repeatability or reproducibility is specifically important in the evaluation of a biosensor since large variation of both biological samples and biosensing materials is expected. The repeatability is usually expressed as relative standard deviation which can be determined by pooling enough replicates to plot a calibration curve and calculating standard deviations. The expected reproducibility of a biosensor is  $\pm 1\%$  to  $\pm 10\%$ .

Lifetime is a quite unique factor in biosensors since the biosensing materials are organic materials that deteriorate with time at a rate affected by temperature, humidity, pH, and other factors. In fact, the response signal of a biosensor to a standard sample changes in months, days, or even hours, depending on the biosensing materials used. Three types of lifetime are considered: the lifetime of a biosensor in use, the lifetime of a biosensor in storage, and the lifetime of the biosensing materials stored separately [132]. It is expected that most biosensing materials should have at least 6 months lifetime when they are stored in a controlled environment, and most biosensors are supposed to be used continuously for several days or weeks.

## **Operation Conditions**

In applications in agriculture, food, and the environment, biosensors may be used on-line or off-line, in real time or not, continuously or discretely, and in the field or in the laboratory. Also, biosensors can be bench-top or portable, stand-alone or with a PC, and totally automatic or semi-automatic. Compared with applications of biosensors to clinical diagnostics and drug screening, the operation environment is a very challenging factor since biosensors may be exposed to significant temperature change, vibration, electromagnetic field, and dust when they are used in a food processing plant, on farm, or in the field. At present, there is still a big gap between the laboratory and in-field use of biosensors. Four different operation modes for a biosensor used in a food processing or bioprocessing are distinguished: on-line and in real time, on-line with feedback, local off-line, and remote laboratory off-line.

#### Available New Technologies

Since the development of biosensors involves chemistry, biology, materials science, engineering, and other areas of study, it is obviously beneficial to update biosensors with new technologies available in these disciplines. Microfluidic chips and nanomaterials are two examples.

Microfluidic chips were introduced in the early 1990s for electrophoretic separations. In comparison with conventional capillary electrophoresis, much shorter analysis times were reported for these novel devices [133]. This advantage is attributed to low thermal and efficient mass transfer, as well as to the large channel surface area-to-channel volume ratios [134]. Furthermore, microfluidic devices are suitable for integration with pumps, valves, and detectors to lead to a *micro-total analysis system* (μTAS) or *lab-on-a-chip*, which is advantageous in terms of high analysis speed and low reagent and power consumption. Microfluidic chips have been extensively used in electrophoresis for detection of small molecules, DNA, and antibodies/antigens. Very recently, Bange et al. [135] presented a review on microfluidic immunosensor systems. They pointed out that the most-used transducing method is fluorescent, followed by electrochemistry, and in applications, the sandwich procedure and competitive assays are commonly used for detection of large molecules and small molecules, respectively.

As new nanoscale materials, including nanotubes/nanowires/nanofibers (e.g., carbon nanotubes and TiO<sub>2</sub> nanofibers) and nanoparticles (e.g., magnetic beads and quantum dots), are available, there is a great interest to integrate them into biosensors for target separation/concentration, miniaturization and performance enhancement. High aspect-ratio one dimensional (1D) nanomaterials such as nanotubes are of particular interest. Recently, for example, a carbon nanotubes-based dual amplification route was tested for ultrasensitive electrochemical biosensors in detection of proteins and DNA [136], OPH-gold nanoparticles were used in an optical biosensors for the direct detection of organophosphate pesticides [137], and silver nanoparticles were reported to enhance current response of a glucose biosensor [138]. More information on materials and techniques in biosensor design and construction can be found in the review by Davis et al. [139] and Zhang et al. [140].

## 2.3.7 Applications of Biosensors in Agriculture, Food, and the Environment

Biosensor technologies have been developed and applied to agriculture, food, and the environment. Table 2 lists the biosensors for analytical targets in agricultural production, food processing, food quality and safety, and environmental monitoring, based on published reports. Many review articles and books address the applications of biosensors in agriculture [141,142], food [4,143-146] and environment [147-149].

In agriculture, biosensors are mainly applied to the measurement of pesticides, fertilizers, odors, and animal diseases. Pesticides are a group of several thousand organic compounds of different structures being used in contemporary agriculture. Because of extensive use and inappropriate management, pesticides are one of the major chemical hazards in agriculture. There are 64 pesticides used in large amounts that are considered potential contaminants of food products and ground water [150]. Applications of

80

biosensors are reviewed for agricultural chemicals [141], field analysis of pollutants in crops and soils, rapid detection of diseases in crops and livestock, and monitoring of animal fertility [151]. Enzymes, antibodies and cell-based biosensors with amperometric, potentiometric, absorption, and fluorescence-labeled optic transducing methods have been developed for detection of insecticides, herbicides, and fungicides used in agriculture. The detection of pesticide and antibiotic residues in agricultural samples has been intensively studied and practiced using electrochemical enzymatic biosensors [152,153], SAW biosensors [154], SPR biosensors [101], and bioluminescence biosensors [89].

In the food area, biosensors have found many applications in determining food components for quality control and to detect microbial and chemical contaminants for

Table 2. Applications of biosensors in agriculture, food, and the environment.

Application	Analytical Targets	Types of Biosensors
Agriculture		•
Insecticides	Paraoxon, dichlorvos, aldicard,	Amperometric immuno- and enzyme sensors
	chlorpyrifos ethyl oxon, methami-	Potentiometric immuno- and enzyme sensors
	dophos, malaoxon, etofenprox,	Absorbance immuno- and enzyme sensors
	etc.	Fluorescent immuno- and enzyme sensors
Herbicides	Atrazine, cyanazine, propazine,	Cell-based electrochemical, and optical sensors
	smazine, prometon, terbutryn, etc.	LAPS immuno- and enzyme sensors
		Fiber optic immunosensors
Fungicides	Dithiocarbamate, carbofuran,	QCM immunosensors
	metalaxyl, triadimefon, etc.	
Food		
Meat quality	Putrescine, cadaverine	Amperometric enzyme sensors
Fish	hypoxanthine	Amperometric immunosensors
freshness		DNA/RNA sensors
Food	L-lactate, D-lactate	Amperometric cell-based and tissue sensors
components		
	Glucose, sucrose, lactose, fructose	Potentiometric enzyme sensors
		Potentiometric enzyme sensors
	Glutamate, aspartame	LAPS immunosensors
	Ethanol, sulfite	Optical FIA biosensor
Food safety	Listeria monocytogenes	Fluorescent immunosensors, PCR-DNA sensors
	E. coli O157:H7	SPR immunosensors, DNA sensors
	Salmonella typhimurium	Chemiluminescence immunosensors
	Staphylococcus aureus	QCM immunosensors, QCM-DNA sensors
		Impedance immunosensors
	Other bacteria and viruses	Thermal enzyme sensors
	Antibiotics, pesticides, fungicides	
Environment		
BOD	Oxygen	Amperometric immuno- and microbe sensors
Pesticides	Paraoxon, atrazine, dithiocar-	Potentiometric cell-based sensors
	bamate, etc.	Impedance immunosensors
Detergents	Alkylate sulfonates	Fiber optic immunosensors, fluorescence
•	•	sensors
Antibiotics	Cephalosporins, penicillins,	Optical cell-based sensors
	nystatin	LAPS immunosensors

food safety and security. Some commercial biosensors have been used to determine the freshness of fish and meat products. In recent years, various biosensors have been studied for rapid, sensitive, specific, low-cost detection of foodborne pathogens including *Listeria monocytogenes*, *E. coli* O157:H7, and *Salmonella typhimurium*. Several review articles present enzyme-based amperometric biosensors for food analysis [145], biosensors for measurement of biological and chemical analytes in food safety [155], biosensors for detection of pathogenic bacteria [156-158] and biosensor applications in food industries [54].

In environmental monitoring, biosensors have been used in detection of pesticide residues, antibiotic residues, toxins, and microbes, and for BOD measurement, in air, water, and soil samples. Biosensors were reviewed by Dennison and Turner [159] for environmental monitoring. Wang et al. [160] reviewed DNA electrochemical biosensors for environmental monitoring. Electrochemical enzymatic biosensors for detection of pesticides in ground and surface water were reviewed by several researchers [152,153,161,162]. In another review article, Rodriguez-Mozaz et al. [149] described various biosensors for environmental applications and their future development trends.

#### 2.3.8 Commercial Biosensor Products

Since the first commercial biosensor for glucose detection was developed in the 1970s, a variety of biosensor products have been developed and commercialized by more than 150 companies worldwide. Biosensors had a \$7.3 billion market in 2003 [27]. Even as glucose biosensors still hold more than 80% of the biosensor market, more and more biosensors have been applied to agricultural production, food processing, and environmental monitoring. Many automatic biosensors, either bench-top or portable types, are on the market. Enzyme-based electrochemical sensors by Yellow Springs Instruments [163], for example, YSI 2700 SELECT, are widely used in food analysis for measurement of glucose, sucrose, lactose, lactate, glutamate, galactose, choline, ethanol, etc. SPR biosensors developed by Biacore [164], with integrated microfluidics cartridges (IFC) for precisely controlling sample and buffer solutions, are able to detect both chemical and biological agents including pathogenic bacteria. Biacore Q is specifically designed for food safety and quality. Microfluidics-based portable RAPTOR fiber optic biosensors of Research International [165] are able to detect toxins, chemicals, bacteria, and viruses. They can quantitatively detect toxins and bacteria at levels as low as 1 ng/mL and 100 cfu/mL, depending upon the species of bacteria and types of samples. Several microbial BOD biosensor systems (BOD-2000 by Nisshin Electric Co., Tokyo, Japan; BODypoint by Aucoteam GmbH, Berlin, Germany; BSBmodul by Prüfgeräte-Medingen GmbH, Dresden, Germany; and ARAS by Dr. Lange GmbH, Berlin, Germany) are used in process control of wastewater and environmental control. Most of these BOD biosensors are based on amperometric oxygen transducers with a flow-through setup. More commercial optical biosensors were reviewed by Baird and Myszka [166] and Rich and Myszka [167].

Selected biosensor manufacturers, instruments, and contact information are summarized in Table 3. Many research prototypes of biosensors are not listed here since they are

Table 3. A selection of biosensors manufacturers, instruments and contact information.

Manufacturer	Instrument	Contact
Electrochemical biosensors		
Analox Instruments	Analyzer LM5, Am2, Gm10, GM7	analox.com
Aucoteam GmbH	BODypoint	Berlin, Germany
BioFutura S.r.1	PerBaco 2000, PeBaco 2002	biofutura.com
Biosensori SpA	Midas Pro	Genoa, Italy
BioSensor Technology GmbH	Thick Film Biosensor	bst-biosensor.de
Biomerieux	Bactometer 64/128, M128	biomerieux-usa.com
Biotech Products	Micro Dialysis Biosensors by Sycopel	biotechproducts.com
Dr. Lange GmbH	ARAS	Berlin, Germany
EKF Diagnostic GmbH	Biosen 5020, 5040, 6020	ekf-diagnostic.de
Flownamics Analytical Instruments	FAIZA 110-P	flownamics.com
Gwent Sensors Ltd	The Answer 8000	g-s-l.co.uk
IBA GmbH	OLGA, on-Line General Analyzer	iba-go.de
Ismatec S.A.	ASIA FIA	ismatec.com
Malthus	Malthus 2000	Crawley, UK
Nisshin Electric Co.	BOD-2000	Japan
Nova Biomedical	Bio Profile Chemistry Analyzer	novabiomedical.com
Oriental Electric Co.	Fresh Meter KV-101	Japan
Prüfgeräte Medingen GmbH	BSBmodule	Dresden, Germany
SensAlyse Ltd	Alcohol Sensor	sensanalyse.com
Toyo Jozo	PM-1000, PM-1000DC, M-100, AZ- 200 Analyzers	Shizuoka, Japan
TRACE Biotech AG	Process TRACE 1.2	trace-ag.de
Universal Sensors	ABD 3000	intel.ucc.ie/sensors/universal/
Yellow Springs Instruments	YSI 2300, YSI 2700 SELECT	ysi.com
Optical biosensors		
Analytical μ-Systems	BIO-Suplar 2	micro-systems.de
Affinity Sensors	IAsys, IAsys Plus, IAsys Auto+	affinity-sensors.com
AVIV Instruments	PWR Model 400	avivins.com
Biacore AB	Biacore 1000, 2000, 3000, X, J, Quant	biacore.com
Farfield Sensors Ltd	AnaLight Bio250	farfield-sensors.co.ik
Graffinity	Plasmon Imager	graffinity.com
HTS Biosystems	SPR array	htsbiosystems.com
IBIS Technologies	IBIS I, IBIS II	ibis-spr.nl
Nippon Laser Electronics	SPR670, SPR Cellia	nle-lab.co.jp
Prolinx	OCTAVE	prolinxinc.com
Quantech Ltd	FarTraQ SPR Aarray	quantechltd.com
Research International	RAPTOR	resrchintl.com
SRU Biosystems	BIND	srubiosystems.com
Texas Instruments	Spreeta	ti.com
ThreeFold Sensors	HH01 BioSensor Fluorometer	threefoldsensors.com
Piezoelectric instruments		
CH Instruments	EQCM 400 Electrochemical QCM	chinstruments.com
Elchema	EQCN-700 and EQCN-900 QCM	elchema.com
Maxtek	PM-700 Series	maxtek.com
Maxtek Princeton Applied Research	PM-700 Series QCA-917, QCA-922 QC Analyzers	maxtek.com prinstonappliedreseacrh.com

	i able 3 (continued).	
QCM Research	Mark Series Cryogenic QCM, and Thermoelectric Pieoelectric Detector	qcmresearch.com
Universal Sensors	PZ-105 Gas Phase Piezoelectric Detec- intel.ucc.ie/sensors/universal/ tor, PZ-1000 Immunobioensor	
Thermal sensors	,	
Minco	Thermistors and thermometers	minco.com
Thermal Metric	Thermal activity monitors	UK
Thermometric	Thermistors, thermistor sensors, thermoprobes	thermometric.com
Yellow Springs Instruments	Thermistors, thermoprobes, thermome- ysi.com ters	

Table 3 (continued).

not on the market. Some reports [27,167], books [147] and review articles [149, 166,168] provided the extensive information on the commercial products of biosensors and their markets.

In the development of new biosensors, there are two great needs which are also great challenges. One is ultra small size and the other one is ultra high sensitivity [14]. MEMS, bio-MEMS, NEMS, and microfluidics technologies have made it possible to have a biosensor with pumps, valves, reactors, separators, detectors, controllers, etc. on chips or arrays [124,135]. Nanomaterials such as nanotubes, nanowires, nanofibers, and nanoparticles applied to biosensors, as reviewed by Kohli et al. [169], Chen et al. [170], Li et al. [171] and Seydack [172], as well as engineered biosensing materials, will play important roles in the development of biosensors that will be able to detect single molecules. As these advanced technologies are integrated into biosensors, the applications of biosensors in agriculture, food, and the environment for rapid, specific, sensitive, inexpensive, infield, on-line and/or in real time detection of biological and chemical agents will be realized.

## Acknowledgements

The author thanks all research associates, research assistants, and graduate students working in the Biosensors and Bioinstrumentation Laboratory at the University of Arkansas for their inputs in preparing this manuscript.

#### References

- 1. Turner, A. P. F., I. Karube, and G. S. Wilson. 1987. *Biosensors Fundamentals and Applications*. Oxford, UK: Oxford University Press.
- 2. Datta, A. K. 1990. Novel chemical and biological sensors for monitoring and control of food processing operations. *J. Food Eng.* 12: 223-238.
- 3. Coulet, P. R. 1991. What is a biosensor? *Biosensor Principles and Applications*. L. J. C. Blum, ed. New York, NY: Marcel Dekker.
- 4. Wangner, G., and G. G. Guibault, eds. 1994. *Food Biosensor Analysis*. New York, NY: Marcel Dekker.
- 5. Clark, Jr. L. C., and C. Lyons. 1962. Electrode systems for continuous monitoring in cardiovascular surgery. *Ann. N.Y. Acad. Sci.* 105: 20-45.

- 6. Guilbault, G. G., and J. Montalvo. 1969. Urea-specific enzyme electrode. *JACS* 91: 2164-2569.
- Scheller, W., R. Hinstche, D. Pfeiffer, F. Schubert, and R. Kindservater. 1991. Biosensors: Fundamentals, applications and trends. *Sensors Actuators* B4: 197-206.
- 8. Turner, A. P. F. 1989. Current trends in biosensor research and development. *Sensors Actuators* 17: 433-450.
- 9. Griffiths, D., and G. Hall. 1993. Biosensors—What real progress is being made? *Trends Biotechnol.* 11: 122-130.
- 10. Gizeli, E., and C. R. Lowe. 1996. Immunosensors. *Current Opinion Biotechnol.* 7: 66-71.
- 11. Powner, E. T., and F. Yalcinkaya. 1997. Intelligent Biosensors. *Sensor Review* 17(2): 107-116.
- 12. Rogers, K. R., and M. Mascini. 1998. Biosensors for field anlytical monitoring. *Field Anal. Chem. Technol.* 2: 317-331.
- 13. Janshoff, A., H.-J. Galla, and G. Steinem. 2000. Piezoelectric mass-sensing devices as biosensors—An alternative to optical biosensors? *Angew. Chem. Int. Ed.* 39: 4004-4032.
- 14. Nakamura, H., and I. Karube. 2003. Current research activity in biosensors. *Anal. Bioanal. Chem.* 377: 446-468.
- 15. Kissinger, P. T. 2005. Biosensors—A perspective. *Biosens. Bioelectron.* 20(12): 2521-2516.
- 16. Marquette, C. A., and L. J. Blum. 2006. State of the art and recent advances in immunoanalytical systems. *Biosens. Bioelectron.* 21: 1424-1433.
- 17. Cass, E. A. G., ed. 1990. *Biosensors: A Practical Approach*. Oxford, UK: Oxford University Press.
- 18. Hall, E. A. H. 1991. Biosensors. Englewood Cliffs, NJ: Prentice Hall.
- 19. Buerk, D. G. 1993. *Biosensors: Theory and Applications*. Lancaster, PA: Technomic Publishing.
- Eggins, B. 1996. Biosensors: An Introduction. New York, NY: John Wiley & Sons.
- 21. Taylor, R., and J. Schultz. 1996. *Handbook of Chemical and Biological Sensors*. Philadelphia, PA: IOP Publishing.
- 22. Kress-Rogers, E. 1997. *Handbook of Biosensors and Electronic Noses*. Boca Raton, FL: CRC Press.
- 23. Cunningham, A. 1998. *Introduction to Bioanalytical Sensors*. New York, NY: John Wiley & Sons.
- 24. Yang, V. C., and T. T. Ngo, eds. 2000. *Biosensors and Their Applications*. New York, NY: Kluwer Academic/Plenum Publishers.
- 25. Sadana, A. 2001. Engineering Biosensors: Kinetics and Design Applications. London, UK: Academic Press.
- 26. Gizeli, E., and C. R. Lowe, eds. 2002. *Biomolecular Sensors*. London, UK: Taylor & Francis.

27. Newman, J. D., L. J. Tigwell, A. P. F. Turner, and P. J. Warner. 2004. *Biosensors: A Clearer View*. Cranfield, UK: Cranfield University.

- 28. Stryer, L. 1988. Biochemistry. New York, NY: W. H. Freeman.
- 29. Bilitewski, U, W. Drewes, J. Neermann, J. Schrader, R. Surkow, R. D. Schmid, and J. Bradley. 1993. Comparison of different biosensor systems suitable for bioprocess monitoring. *J. Biotechnol.* 31: 257-266.
- 30. Jenkins, D. M., and M. J. Delwiche. 2002. Manometric biosensor for on-line measurement of milk urea. *Biosens. Bioelectron.* 17: 557-563.
- 31. Bilitewski, U. M. Genrich, S. Kadow, and G. Mersal. 2003. Biochemical analysis with microfluidic systems. *Anal. Bioanal. Chem.* 377: 556-569.
- 32. Soldatkin, A. P., V. N. Arkhypova, S. V. Dzyadevch, A. V. El'skaya, J.-M. Gravoueille, N. Jaffrezic-Renault, and C. Martelet. 2005. Analysis of potato glycoalkaloids by using of enzyme biosensor based on pH-ISFETs. *Talanta* 66: 28-33.
- 33. Stanier, R. Y., J. L. Ingraham, M. L. Wheelis, and R. P. Painter. 1986. *The Microbial World*. 5<sup>th</sup> ed. Upper Saddle River, NJ: Prentice Hall.
- 34. Roitt, I., J. Brostoff, and D. Male. 1989. *Immunology*. London, UK: Elsevier Science Health Science Div.
- 35. Delwiche, M., X. Tang, R. BonDurant, and C. Munro. 2001. Improved biosensor for measurement of progesterone in bovine milk. *Trans. ASAE* 44: 1997-2002.
- 36. Sanaikone, K., M. J. Delwiche, R. H. BonDurant, and C. J. Munro. 2004. Quantitative lateral flow immunoassay for measuring progesterone in bovine milk. *Trans. ASAE* 47: 1357-1365.
- 37. Su, X. L., and Y. Li. 2004. A self-assembled monolayer-based piezoelectric immunosensor for rapid detection of *Escherichia coli* O157:H7. *Biosens. Bioelectron.* 19(6): 563-574.
- 38. Tschmelak, J., G. Proll, and G. Gauglitz. 2005. Optical biosensor for pharmaceuticals, antibiotics, hormones, endocrine disrupting chemicals and pesticides in water: Assay optimization process for estrone as example. *Talanta* 65: 313-323.
- 39. Rajendran, M., and A. D. Ellington. 2002. Chapter 12: Nucleic acids for reagentless biosensors. *Optical Biosensors—Present & Future*, 369-396. F. S. Ligler, and C. A. R. Taitt, eds. Amsterdam, Netherlands: Elsevier.
- 40. Tombelli, S., M. Minunni, and M. Mascini. 2005. Analytical applications of aptamers. *Biosens. Bioelectron.* 20: 2424-2434.
- 41. Chan, S. D. H., K. Dill, J. Blomdahl, and H. G. Wada. 1996. Non-isotopic quantitation of mRNA using a novel RNAse protection assay. Measurements of erb-2 mRNA in tumor cell lines. *Anal. Biochem.* 242: 214-220.
- 42. Kai, E., S. Sawata, K. Ikebukuro, T. Lida, T. Honda, and I. Karube. 1999. Detection of PCR products in solution using surface plasmon resonance. *Anal. Chem.* 71: 796-800.
- 43. McCauley, T. G., N. Hamaguchi, and M. Stanton. 2003. Aptamer-based biosensor arrays for detection and quantification of biological macromolecules. *Anal. Biochem.* 319: 244-250.

- 44. Kalogianni, D. P., T. Koraki, T. K. Christopoulos, and P. C. Ioannou. 2006. Nanoparticle-based DNA biosensor for visual detection of genetically modified organisms. *Biosens. Bioelectron.* 21: 1069-1076.
- 45. Zhao, W., S. Yao, and I.-M. Hsing. 2006. A microsystem compatible strategy for viable *Escherichia coli* detection. *Biosens. Bioelectron.* 21: 1163-1170.
- 46. Davis, C. 1975. Annals Microbiol. 126A: 175-186.
- 47. Karube, and Suzuki. 1990. Microbial biosensors. *Biosensors: A Practical Approach*, 155-170. A. E. G. Cass, ed. Oxford, UK: Oxford University Press.
- 48. Rack, J. 1995. Cell-Based Biosensor. Lancaster, PA: Technomic Publishing Co.
- 49. Chang, I. S., J. K. Jang, G. C. Gil, M. Kim, H. J. Kim, B. W. Cho, and B. H. Kim. 2004. Continuous determination of biochemical oxygen demand using microbial fuel cell based biosensor. *Biosens. Bioelectron.* 19: 607-613.
- 50. Goto, M., K. Sato, A. Murakami, M. Tokeshi, and T. Kitamori. 2005. Development of a microchip-based bioassay system using cultured cells. *Anal. Chem.* 77: 2125-2131.
- 51. Kumar, J., S. K. Jha, and S. F. D'Souza. 2006. Optical microbial biosensor for detection of methyl parathion pesticide using *Flavobacterium sp.* whole cells adsorbed on glass fiber filters as disposable biocomponent. *Biosens. Bioelectron*. 21: 2100-5. Available at: www.sicencedirect.com.
- 52. Bousse, L. 1996. Whole cell biosensors. Sensors Actuators B34: 270-275.
- 53. D'Souza, S. F. 2001. Microbial biosensors. Biosens. Bioelectron. 16: 337-353.
- 54. Mello, L. D., and L. T. Kubota. 2002. Review of the use of biosensors as analytical tools in the food and drink industries. *Food Chem.* 77: 237-256.
- 55. Rechnitz, G. A. 1978. Biochemical electrodes uses tissues slice. *Chem. Eng. News* 56: 16-21.
- 56. Wijesuriya, D. C., and G. A. Rechnitz. 1993. Biosensors based on plant and animal tissues. *Biosens. Bioelectron.* 8: 155-60.
- 57. Sezginturk, M. K., and E. Dinckaya. 2005. Direct determination of sulfite in food samples by a biosensor based plant tissue homogenate. *Talanta* 65: 998-1002.
- 58. Wu, F., Y. Huang, and C. Huang. 2005. Chemiluminescence biosensor system for lactic acid using natural animal tissue as recognition element. *Biosens. Bioelectron.* 21: 518-522.
- 59. Topcu, S., M. K. Sezginturk, and E. Dinckaya. 2004. Evaluation of a new biosensor-based mushroom (*Agaricus bisporus*) tissue homogenate: Investigation of certain phenolic compounds and some inhibitor effects. *Biosens. Bioelectron*. 20: 592-597.
- 60. Heinemann, W. R., C. W. Anderson, and H. B. Halsall. 1979. Immunoassay by differential pulse polarography. *Sci. Wash.* 204: 865-866.
- Niculescu, M., C. Nistor, I. Frebort, P. Pec, B. Mattiasson, and E. Csoregi. 2000. Redox hidrogel-based amperometric bienzyme electrode for fish freshness monitoring. *Anal. Chem.* 72: 1591-1597.
- 62. Ruan, C. M., H. Wang, and Y. Li. 2002. A bienzyme electrochemical biosensor coupled with immunomagnetic separation for rapid detection of *Escherichia coli* O157:H7 in food samples. *Trans. ASAE* 45(1): 249-255.

63. Liu, G., and Y. Lin. 2006. Biosensor based on self-assembling acetylcholinesterase on carbon nanotubes for flow injection/amperometric detection of organophosphate pesticides. *Anal. Chem.* 78: 835-843.

- 64. Bergveld, P. 1970. Development of an ion sensitive solid state device for neurophysiological measurements. *IEEE Trans. Biomed. Eng.* 17: 70-71.
- 65. Panfili, P. R., K. Dill, and J. D. Olson. 1994. Immunochemical detection using the light-addressable potentiometric sensor. *Curr. Opin. Biotechnol.* 5: 60-64.
- 66. Dill, K., J. H. Song, J. A. Blomdahl, and J. D. Olson. 1997. Rapid, sensitive and specific detection of whole cells and spores using the light-addressable potentiometric sensor. *J. Biochem. Biophys. Methods* 34: 161-166.
- 67. Tu, S., J. Uknalis, and P. L. Irwin. 2002. The capture of *Escherichia coli* O157:H7 for light addressable potentiometric sensor (LAPS) using two different types of magnetic beads. *J. Rapid Meths. Automat. Microbiol.* 10: 185-195.
- 68. Mourzina, G., E. Ermolenko, T. Yoshinobu, Y. Vlasov, H. Iwasaki, and M. J. Schöning. 2003. Anion-selective light-addressable potentiometric sensors (LAPS) for the determination of nitrate and suphate ions. *Sensors Actuators* B91: 32.
- Keusgen, M., J. P. Kloock, D. T. Knobbe, M. Jünger, I. Krest, M. Goldbach, W. Klein, and M. J. Schöning. 2004. Direct determination of cyanides by potentiometric biosensors. *Sensors Actuators* B103: 380-385.
- 70. Ivnitski, D., I. Abdel-Hamid, P. Atanasov, E. Wilkins, and S. Stricker. 2000. Review: Applications of electrochemical biosensors for detection of food pathogenic bacteria. *Electroanal*. 12(5): 317-325.
- 71. Shah, J., and E. Wilkins. 2003. Electrochemical biosensors for detection of biological warfare agents. *Electroanal*. 15: 157-167.
- 72. Bakker, E. 2004. Electrochemical sensors. Anal. Chem. 76: 3285-3298.
- 73. Watson, L. D., P. Maynard, D. C. Cullen, R. S. Sethi, J. Brettle, and C. R. Lowe. 1988. A microelectronic conductimetric biosensor. *Biosens. Bioelectron.* 3: 101-115.
- Gómez, R., R. Bashir, A. Sarikaya, M. R. Ladisch, J. Sturgis, J. P. Robinson, T. Geng, A. K. Bhunia, H. L. Apple, and S. Wereley. 2001. Microfluidic biochip for impedance spectroscopy of biological species. *Biomed. Microdevices* 3: 201-209.
- 75. Radke, S., and E. C. Alocilja. 2005. A high density microelectrode array biosensor for detection of *E. coli* O157:H7. *Biosens. Bioelectron*. 20: 1662-1667.
- 76. Yang, L., Y. Li, and G. Erf. 2004. Interdigitated array microelectrode-based electrochemical impedance immunosensor for detection of *Escherichia coli* O157:H7. *Anal. Chem.* 76(4): 1107-1113.
- 77. Yang, L., and Y. Li. 2005a. AFM and impedance spectroscopy characterization of the immobilization of antibodies on indium-tin oxide electrodes and their capture of *E. coli* O157:H7. *Biosens. Bioelectron.* 20(7): 1407-1416.
- 78. Louie, A. S., I. G. Marenchic, and R. H. Whelan. 1998. A fieldable modular biosensor for use in detection of foodborne pathogens. *Field Anal. Chem. Tech.* 2: 371-377.

- Schöning, M. J., M. Arzdorif, P. Mulchandani, W. Chen, and A. Mulchandani.
   2003. A capacitive field-effect sensor for the direct determination of organophosphorus pesticides. *Sensors Actuators* B91: 92-97.
- 80. Anh, T. M., S. V. Dzyadevych, M. C. Van, N. J. Renault, C. N. Duc, and J.-M. Chovelon. 2004. Conductometric tyrosinase biosensor for the detection of diuron, atrazine and its main metabolites. *Talanta* 63: 365-370.
- 81. Berggren, C., B. Bjarnason, and G. Johansson. 2001. Capacitive biosensors. *Electrognal.* 13: 173-180.
- 82. Katz, E., and I. Willner. 2003. Probing biomolecular interactions at conductive and semiconductive surfaces by impedance spectroscopy: Routes to impedimetric immunosensors, DNA-sensors, and enzyme biosensors. *Electroanal*. 15: 913-945.
- 83. Liley, M. 2002. Chapter 6: Optical transducers. *Biomolecular Sensors*, 121-175. E. Gizeli, and C. R. Lowe, eds. London, UK: Taylor & Francis.
- 84. Delwiche, M., E. Cox, B. Goddeeris, C. Van Dorpe, J. De Baerdemaeker, E. Decuypere, and W. Sansen. 2000. A biosensor to detect penicillin residues in food. *Trans. ASAE* 43: 153-159.
- 85. Liu, Y., and Y. Li. 2002. Detection of *Escherichia coli* O157:H7 using immunomagnetic separation and absorbance measurement. *J. Microbiol. Methods* 51(3): 369-377.
- 86. Capitan-Vallvey, L. F., E. Arroyo-Guerrero, M. D. Ferandez-Ramos, and F. Santoyo-Gonzalez. 2005. Disposable receptor-based optical sensor for nitrate. *Anal. Chem.* 77: 4459-4466.
- 87. Blum, L. J., S. M. Gautier, and P. R. Coulet. 1994. Chapter 5: Fiber-optic biosensors based on luminometric detection. *Food Biosensor Analysis*, 101-121. Wangner and Guibault, eds. New York, NY: Marcel Dekker.
- 88. Blum, L. J., and P. R. Coulet. 2000. Chapter 12: Luminescent biosensors. *Biosensors and Their Applications*, 213-222. V. C. Yang, and T. T. Ngo, eds. New York, NY: Kluwer Academic/Plenum Publishers.
- 89. Kim, B. C., and M. B. Gu. 2003. A bioluminescent sensor for high throughout toxicity classification. *Biosens. Bioelectron.* 18: 1015-1021.
- 90. Liu, Y., J. Ye, and Y. Li. 2003. Rapid detection of *E. coli* O157:H7 in ground beef, chicken carcass and lettuce samples with an immunomagnetic chemiluminescence fiber-optic biosensor. *J. Food Prot.* 66(3): 512-517.
- 91. Tu, S., A. G. Gehring, and P. L. Irwin. 2005. Bioenergetic confirmation of viable pathogens in foods by atp-bioluminescence. *Luminescence* 15: 445-448.
- 92. Taitt, C. R., G. P. Anderson, and F. S. Ligler. 2005. Evanescent wave fluorescence biosensors. *Biosens. Bioelectron.* 20(12): 2470-2487.
- 93. Barzen, C., A. Brecht, and G. Gauglitz. 2002. Optical multiple-analyte immunosensor for water pollution control. *Biosens. Bioelectron.* 17: 289-295.
- 94. Goh, Y. Y., V. Frecer, B. Ho, and J. L. Ding. 2002. Rational design of green fluorescent protein mutants as biosensor for bacterial endotoxin. *Protein Engr.* 15: 493-502.

95. Yang, L., and Y. Li. 2005b. Quantum dots as fluorescent labels for quantitative detection of *Salmonella typhimurium* in chicken carcass wash water. *J. Food Prot.* 68(6): 1241-1245.

- 96. Pickup, J. C., F. Hussain, N. D. Evans, O. J. Rolinski, and D. J. S. Birch. 2005. Fluorescence-based glucose sensors. *Biosens. Bioelectron*. 20: 2555-2565.
- 97. Hahn, K., and A. Toutchkine. 2002. Live-cell fluorescent biosensors for activated signaling proteins. *Current Opinion Cell Biol.* 14: 167-172.
- 98. Glaser, R. W. 2000. Surface plasmon resonance biosensors. *Biosensors and Their Applications*, 195-212. V. C. Yang, and T. T. Ngo, eds. New York, N.Y.: Kluwer Academic.
- 99. Rich, R. L., and D. G. Myszka. 2004. Why you should be using more SPR biosensors technology. *Drug Discovery Today: Tech.* 1: 301-308.
- 100. Meeusen, C. A., E. C. Alocilja, and W. N. Osburn. 2005. Detection of *E. coli* O157:H7 using a miniaturized surface plasmon resonance biosensor. *Trans. ASABE* 48(6): 2409-2416.
- 101. Dumont, V., A.-C. Huet, I. Traynor, C. Elliott, and P. Delahaut. 2006. An SPR biosensor assay for the simultaneous determination of thiamphenicol, florefenicol, florefenicol amine and chloramphenicol residues in shrimps. *Anal. Chim. Acta (in press)*.
- 102. Robinson, G. R., J. W. Attridge, J. K. Deacon, and S. C. Whiteley. 1993. The fluorescent capillary device. *Sensors Actuators* B11: 235-238.
- 103. Cush, R., J. M. Cronin, W. J. Stewart, C. H. Maul, J. Molloy, and N. J. Goddard. 1993. The resonance mirror: A novel optical biosensor for direct sensing of biomolecular interactions. Part I: Principle of operation and associated instrumentation. *Biosens. Bioelectron.* 8: 347-353.
- 104. Ligler, F. S., and C. A. R. Taitt. 2002. *Optical Biosensors—Present & Future*. Amsterdam, Netherlands: Elsevier.
- 105. Thompson, R. B., ed. 2006. *Fluorescence Sensors and Biosensors*. Boca Raton, FL: Taylor & Francis Group.
- 106. Shons, A., F. Dorman, and J. Najarian. 1972. The piezoelectric quartz immunosensor, *J. Biomed. Mater. Res.* 6: 565-570.
- 107. Su, X. D., S. F. Y. Li, J. Kwang, and S. Low. 2000. Piezoelectric quartz crystal based screening test for porcine reproductive and respiratory syndrome virus infection in pigs. *Anal. Chem.* 125: 725-730.
- 108. Berkenpas, E., P. Millard, and M. Cunha. 2005. Detection of *Escherichia coli* O157:H7 with langasite pure shear horizontal surface acoustic wave sensor. *Biosens. Bioelectron.* (in press).
- 109. Su, X. L., and Y. Li. 2005. A QCM immunosensor for *Salmonella* detection with simultaneous measurement of resonant frequency and motional resistance. *Biosens. Bioelectron.* 21(6): 840-848.
- 110. Ward, M. D., and D. A. Buttry. 1990. In situ interfacial mass detection with piezoelectric transducers. *Science* 249: 1000-1007.

- 111. Buttry, D. A., and M. D. Ward. 1992. Measurement of interfacial process at electrode surfaces with the electrochemical quartz crystal microbalance. *Chem. Rev.* 92(6): 1355-1379.
- 112. Suleiman, A. A., and G. G. Guilbault. 1994. Recent developments in piezo-electric immunosensors: A review. *Analyst* 119: 2279-2282.
- 113. Bunde, R. L., E. J. Jarvi, and J. J. Rosentreter. 1998. Piezoelectric quartz crystal biosensors. *Talanta* 46: 1223-1236.
- 114. Marx, K. A. 2003. Quartz crystal microbalance: A useful tool for studying thin polymer films and complex biomolecular systems at the solution-surface interface. *Biomacromolecules* 4(5): 1099-1120.
- 115. Buck, R. P., E. Lindner, W. Kutner, and A. G. Inzelt. 2004. Piezoelectric chemical sensors. *Pure Appl. Chem.* 76(6): 1139-1160.
- 116. Kroger, S., and B. Danielsson. 1997. Chapter 13: Calorimetric biosensors. Handbook of Biosensors and Electronic Noses—Medicine, Food, and the Environment, 279-298. E. Kress-Rogers, ed. Boca Raton, FL: CRC Press.
- 117. Danielsson, B., and K. Mosbach. 1987. Chapter 29: Theory and application of calorimetric sensors. *Biosensors: Fundamentals and Applications*, 575-597. A. P. F. Turner, I. Karube and G. S. Wilson, eds. Oxford, UK: Oxford University Press.
- 118. Lammers, F., and T. Scheper. 1999. Thermal biosensors in biotechnology. Adv. *Biochem. Eng. Biotechnol.* 64: 35-67.
- 119. Ramanathan, K., and B. Danielsson. 2001. Principles and applications of thermal biosensors. *Biosens. Bioelectron.* 16: 417-423.
- 120. Ramanathan, K., B. R. Jönsson, and B. Danielsson. 2001. Sol-gel based thermal biosensor for glucose. *Anal. Chimica Acta* 19: 1-10.
- 121. Zheng, Y.-H., T.-C. Hua, D.-W. Sun, J.-J. Xiao, F. Xu, and F.-F. Wang. 2006. Detection of dichlorvos residue by flow injection calorimetric biosensor based on immobilized chicken liver esterase. *J. Food Engr.* 74: 24-29.
- 122. Megens, M., and M. Prins. 2005. Magnetic biochips: A new option for sensitive diagnostics. *J. Magnet. Magnet. Materials* 293: 702-708.
- 123. Varshney, M., B. Srinivasan, S. Tung, and Y. Li. 2005a. Microfluidic filter chip based chemiluminescence filter optic biosensing method for the detection of *E. coli* O157:H7. ASABE Paper No. 057028. St. Joseph, MI: American Society of Agricultural and Biological Engineers.
- 124. Pamme, N. 2006. Magnetism and microfluidics. Lab Chip 6: 24-38.
- 125. Mayes, A. G. 2002. Chapter 4: Immobilization chemistry of biological recognition molecules. *Biomolecular Sensors*, 49-86. E. Gizeli, and C. R. Lowe, eds. London, UK: Taylor & Frances.
- 126. Dubrovsky, T., A. Tronin, S. V. Dubrovskaya, and C. Nicolini. 1995. Immunological activity of IgG Langmuir films oriented by protein A layer. *Sensors Actuators* B23: 1-7.
- 127. Anzai, J.-I., T. Hoshi, and T. Osa. 2000. Chapter 3: Avidin-biotin mediated biosensors. *Biosensors and Their Applications*, 35-46. V. G. Yang, and T. T. Ngo, eds. New York, NY: Kluwer Academic/Plenum Publishers.

128. Duschl, C. 2002. Chapter 5: Binding isotherms and kinetics of immobilized biological systems. *Biomolecular Sensors*, 87-120. E. Gizeli, and C. R. Lowe, eds. London, UK: Taylor & Frances.

- 129. Gijs, M. A. M. 2004. Magnetic bead handling on-chip: New opportunities for analytical applications (review). *Microfluidics Nanofluidics* 1: 22-40.
- 130. Tibbe, A. G. J., B. G. de Grooth, J. Greve, G. J. Dolan, C. Rao, and L. W. M. M. Terstappen. 2002. Magnetic field design for selecting and aligning immunomagnetic labeled cells. *Cytometry* 47: 163-172.
- 131. Varshney, M., L. Yang, X. Su, and Y. Li. 2005b. Magnetic nanoparticle-antibody conjugates for the separation of *Escherichia coli* O157:H7 in ground beef. *J. Food Prot.* 68: 1804-1811.
- 132. Deshpande, S. S. 1996. *Enzyme Immunoassays—From Concept to Product Development*. New York, NY: Chapman & Hall.
- 133. Schwarz, M. A., and P. C. Hauser. 2001. Recent development in detection methods for microfabricated analytical devices. *Lab on a Chip* 1: 1-6.
- 134. Meldrum, D. R., and M. R. Holl. 2002. Microscale bioanalytical systems. *Science* 297: 1197-1198.
- 135. Bange, A., B. H. Halsall, and W. R. Heinemann. 2005. Review: Microfluidic immunosensor systems. *Biosens. Bioelectron.* 20: 2488-2503.
- 136. Wang, J., G. Liu, and M. R. Jan. 2004. Ultrasensitive electrical biosensing of proteins and DNA: Carbon-nanotube derived amplification of the recognition and transduction events. *JACS Communicat*. 126: 3010-3011.
- 137. Simonian, A. L., T. A. Good, S.-S. Wang, and J. R. Wild. 2005. Nanoparticle-based optical biosensors for the direct detection of organophosphate chemical warfare agents and pesticides. *Anal. Chim. Acta* 534: 69-77.
- 138. Ren, X., X. Meng, D. Chen, F. Tang, and J. Jiao. 2005. Using silver nanoparticles to enhance current response of biosensor. *Biosens. Bioelectron*. 21: 433-437.
- 139. Davis, J., D. H. Vaughan, and M. F. Cardosi. 1995. Elements of biosensors construction. *Enzyme Microbial Tech.* 17: 1030-1035.
- 140. Zhang, S., G. Wright, and Y. Yang. 2000. Materials and techniques for electrochemical biosensors design and construction. *Biosens. Bioelectron.* 15: 273-282.
- 141. Wong, R. B. 2000. Chapter 16: Biosensors for agrochemicals. *Biosensors and Their Applications*, 283-298. V. G. Yand, and T. T. Ngo, eds. New York, NY: Kluwer Academic/Plenum Publishers.
- 142. Kindschy, L. M., and E. C. Alocilja. 2004. A review of molecularly imprinted polymers for biosensor development for food and agricultural applications. *Trans. ASAE* 47: 1375-1382.
- 143. Deshpande, S. S., and R. M. Rocco. 1994. Biosensors and their potential use in food quality control. *Food Technol*. 48(6): 146-150.
- 144. Scott, A. O., ed. 1998. *Biosesnors for Food Analysis*. Cambridge, UK: The Royal Society of Chemistry.
- 145. Prodromidis, M. I., and M. I. Karayannis. 2002. Enzyme based amperometric biosensors for food analysis. *Electroanal*. 14(4): 241-261.

- 146. Amine, A., H. Mohammadi, I. Bourais, and G. Palleschi. 2006. Enzyme inhibition-based biosensors for food safety and environmental monitoring. *Biosens. Bioelectron.* 21: 1405-1423.
- 147. Ramsay, G., ed. 1998. Commercial Biosensors—Applications to Clinical, Bioprocess, and Environmental Samples. New York, NY: John Wiley & Sons.
- 148. Bilitewski, U., and A. Turner. 2000. *Biosensors in Environmental Monitoring*. Abingdon, UK: Taylor & Francis Group.
- 149. Rodriguez-Mozaz, S., M.-P. Marco, M. J. L de Alda, and D. Barcelo. 2004. Biosensors for environmental applications: Future development trends. *Pure Appl. Chem.* 76(4): 723-752.
- 150. Fielding, M., D. Barcelo, S. Helweg, L. Galassi, L. Torstensson, P. van Zoonen, R. Wolter, and G. Angeletti. 1992. Pesticides in Ground and Drinking Water, Water Pollution. Research Report No. 27, Commission of the European Communities, Brussels.
- 151. Velasco-Garcia, M. N., and T. Mottram. 2003. Biosensors technology addressing agricultural problems. *Biosystem Engr.* 84: 1-12.
- 152. Trojanowicz, M., and M. Hitchman. 1996. Determination of pesticides using electrochemical biosensors. *TrAC Trends Anal. Chem.* 15: 38-45.
- 153. Trojanowicz, M. 2002. Determination of pesticides using electrochemical enzymatic biosensors. *Electroanal*. 14: 19-20.
- 154. Tom-Moy, M., R. L. Baer, D. Solomon, and T. P. Doherty. 1995. Atrazine measurements using surface transverse wave device. *Anal. Chem.* 67: 1510-1516.
- 155. Patel, P. D. 2002. (Bio)sensors for measurement of analytes implicated in food safety: A review. *TrAC Trends Anal. Chem.* 21: 96-115.
- 156. Ivnitski, D., I. Abdel-Hamid, P. Atanasov, and E. Wilkins. 1999. Review: Biosensors for detection of pathogenic bacteria. *Biosens. Bioelectron.* 14: 599-624.
- 157. Leonard, P. S. Hearly, J. Brennan, L. Dunne, J. Quinn, T. Chakrabory, and R. O'Kennedy. 2003. Advances in biosensors fro detection of pathogens in food and water. *Enzyme Microbial Tech.* 32: 3-13.
- 158. Deisingh, A. K., and M. Thompson. 2004. Strategies for the detection of Escherichia coli O157:H7 in foods. *J. Appl. Microbiol.* 96: 419-429.
- 159. Dennison, M. J., and A. P. F. Turner. 1995. Biosensors for environmental monitoring. *Biotech Adv.* 13: 1-12.
- 160. Wang, J., G. Rivas, and X. Cai. 1997. DNA electrochemical biosensors for environmental monitoring. *Anal. Chim. Acta* 347: 1-8.
- 161. Karube, I., and Y. Nomura. 2000. Enzyme sensors for environmental analysis. *J Molecul. Catalysis* B10: 177-181.
- 162. Suri, C. R., M. Raje, and G. C. Varshney. 2002. Immunosensors for pesticide analysis: Antibody production and sensors development. *Critical Reviews in Biotech.* 22: 15-32.
- 163. Yellow Springs Instruments. 2005. YSI 2700 SELECT. Available at: www.ysi .com. Accessed December 2005.

- 164. BIACORE. 2005. An introduction to Biacore's SPR technology. Available at: www.biacore.com. Accessed October 2005.
- 165. Research International. 2005. RAPTOR biosensor. Available at: www.resrchintl.com/raptor.html. Accessed October 2005.
- 166. Baird, C., and D. G. Myszka. 2001. Current and emerging commercial optical biosensors. *J. Mol. Recognit.* 14: 261-268.
- 167. Rich, R. L., and D. G. Myszka. 2002. Review: A survey of the year 2002 commercial optical biosensor literature. *J. Mol. Recognit.* 16: 351-382.
- 168. Robinson, G. A. 1991. Optical immunosensing systems—Meeting the market needs. *Biosens. Bioelectron.* 6: 183-191.
- 169. Kohli, P., M. Wirtz, and C. R. Martin. 2004. Nanotubes membrane based biosensors. *Electroanal*. 16: 9-18.
- 170. Chen, J., Y. Mao, N. He, X. Wu, and S. Li. 2004. Nanotechnology and biosensors. *Biotech Adv.* 22: 505-518.
- 171. Li, J., J. E. Koehne, A. M. Cassell, H. Chen, H. T. Ng, Q. Ye, W. Fan, J. Han, and M. Meyyappan. 2005. Inlaid multi-walled carbon nanotube nanoelectrode arrays for electroanalysis. *Electroanal*. 17: 15-27.
- 172. Seydack, M. 2005. Nanoparticle labels in immunosensing using optical detection methods. *Biosens. Bioelectron.* 20(12): 2425-2469.

# 2.4 Robotics: Fundamentals and Prospects

K. J. Kyriakopoulos and S. G. Loizou

**Abstract.** Robotics is the branch of engineering that involves the conception, design, manufacture, and operation of robots. Robots are machines that are intended to replace human beings in the execution of tasks that involve physical activities or decision-making. Here we present some of the fundamentals from the fields of stationary and moving base robots and prospects for underwater, aerial, and walking robots.

Keywords. Robotics, Stationary base robots, Moving base robots.

#### 2.4.1 Robot Definitions and a Brief History

The ambition to give life to artifacts lies deep in the history of humankind. In Greek mythology, the titan Prometheus who molded humankind from clay, and the giant Talus, the bronze slave forged by Hephaestus entrusted to protect the island of Crete from invaders, testify to that ambition.

The concept of robots in the form that is now established in people's mind was first introduced by the Czech playwright Karel Čapek in 1921 with his science fiction tale *Rossum's Universal Robots*. In his tale he coined the term *robot* from the Slav *robota* which means executive labor, to denote the automaton built by Rossum. It ends up rising against humankind.

	List of Symbols and Abbreviations			
0	Cartesian frame	CCD	charge-coupled device	
$O_{xyz}$	Cartesian frame with x, y, z axes	IC	integrated circuit	
q	joint variable	PID	proportional integrator differenti-	
p	position vector		ator	
η, s, a	unit vectors	DAC	digital-to-analog converter	
A	rotation and translation matrix	AGV	automatic guided vehicle	
R	rotation matrix	AMR	autonomous mobile robot	
P	point	UUV	unmanned underwater vehicle	
λ	generalized coordinate	AUV	autonomous underwater vehicle	
L	Lagrangian	ROV	remotely operated vehicle	
T	kinetic energy	GPS	global positioning system	
U	potential energy	MBE	multi-beam echo-sounder	
ξ	generalized force	SSS	side-scan sonar	
DC	direct current	SBP	sub-bottom profiler	
AC	alternating current	CTD	conductivity, temperature and	
			depth	

In the 1940s, the famous science fiction writer Isaac Asimov conceived the robot as an automaton of human appearance. Apart from the tasks that it was programmed to perform, its behavior was dictated by a "positronic" brain programmed by a human being to follow certain ethical rules. The term *robotics* was introduced by Isaac Asimov as the science devoted to the study of robots, which was based on the *three fundamental laws of robotics*:

- 1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- 2. A robot must obey the orders given by human beings, except when such orders would conflict with the first law.
- 3. A robot must protect its own existence, as long as such protection does not conflict with the first or second law.

According to the widely accepted definition by the Robot Institute of America, a robot is a reprogrammable multifunctional manipulator designed to move materials, parts, tools or specialized devices through variable programmed motions for the performance of a variety of tasks. The structure of such an industrial robot consists of:

- A mechanical structure (the manipulator), which consists of a series of rigid bodies (links) connected with articulations (joints). The manipulator is the robotic analogue to the human hand and we can characterize it by the *arm*, the *wrist* and the *end effector*.
- The actuators, which are responsible for setting the robot in motion. The motors that are typically used for actuation are electric, hydraulic, pneumatic, etc.
- The sensing subsystem, with which the robot gathers information about its state and the state of its surrounding environment. This system is also responsible for accepting the external commands.

• The control system (computer) that integrates sensing and acting in an intelligent and efficient way, so the whole system performs as expected.

The first patent for an industrial robot was granted to George Devol Jr., who named his construction *Unimation*. The first robot application in industry was realized in 1959 at General Motors' Turnstead plant. The first robots were very big with motion produced by hydraulic systems. They were hard to program, control, and maintain. In the 1970s the robotics industry vastly benefited by advancements in microprocessors. New robot models were produced which were easy to reprogram. Fast control architectures that needed more computation could then be realized and the new robots had a greater diversity of applications. Hydraulic systems were replaced by DC motors and later by AC servomotors.

By the 1990s robots were a well-proven technology established in industry. Now they are used in dangerous and unhygienic environments. They perform repeating tasks with constant quality and unbeatable productivity. They may have displaced many people from their jobs, but they provided for vast technological advancements.

#### 2.4.2 Classification of Robots

In the following we present the most common criteria used for robot classification.

- Classification according to the number of their degrees of freedom or the number of their degrees of mobility (joints)—The degrees of mobility (number of joints) are distributed along the mechanical structure to provide the necessary degrees of freedom required for a given task. For an arbitrary task in the three-dimensional space, six degrees of freedom are required: three for the position and three for the orientation of an object. If there exist more degrees of mobility than degrees of freedom for a given manipulator then it is said to be redundant.
- Classification according to the configuration of their joints—Joints are used for the realization of relative motion between the links. There are two types of joints: prismatic joints, which are used for translational relative motion, and revolute joints used for rotational relative motion. Any other type of joint can be considered as a composition of those two types. Revolute joints are usually preferred to prismatic due to their compactness and reliability
- Classification according to the mechanical/geometric configuration—The type and sequence of the degrees of mobility of the robotic arm allows the classification of manipulators according to their mechanical/geometric configurations.

Three prismatic joints, whose axes are perpendicular to one another, characterize the *Cartesian* manipulators. Those joints are the three first joints of the kinematic chain. The Cartesian geometry is simple and provides good mechanical stiffness and constant positioning accuracy. The workspace is a parallelepiped.

For the *cylindrical* manipulators, the first joint in the kinematic chain is revolute and the second and third joints are prismatic like the Cartesian. The workspace is a portion of a hollow cylinder.

The SCARA type of robots have the first two joints of the kinematic chain revolute and the third is prismatic. The acronym SCARA stands for selective compliance assembly robot arm and characterizes the mechanical features of a

96 Chapter 2 Hardware

structure providing high stiffness to vertical loads and compliance to horizontal loads. The workspace is a portion of a hollow cylinder.

Three revolute joints realize the *anthropomorphic* geometry. The axis of the first joint is orthogonal to the axes of the other two joints, which are parallel. The workspace is a portion of a sphere.

- Classification according to the type of drive system—According to the type of drive system robots can be classified as hydraulic, pneumatic, electric or mixed actuated.
- Classification according to the type of control—Based on the control applied to a robot, it can be classified as non-servo point-to-point, point-to-point servo controlled, or continuous path servo controlled.

## 2.4.3 Stationary Base Robots

## Description

Robots that have the base fixed to a specific location in the workspace constitute the traditional type of industrial robotic systems. They consist of a chain of joints and links, which comprise a mechanical arm, animated through proper actuators and electronic elements. The structure of such a industrial robot system, described in Figure 1, consists of a mechanical subsystem, the actuators, the sensing subsystem, and the control system.

#### Characterization

Stationary base robots, are characterized according to their mechanical/geometrical structure as *Cartesian*, *gantry* (Cartesian type), *cylindrical*, *spherical*, *anthropomorphic*, and *SCARA*. According to their type of motion, they are characterized as *non-servo point-to-point*, *point-to-point servo controlled*, or *continuous path servo controlled*.

The most important characteristics of an industrial robot are its payload, repeatability, and accuracy. *Payload* is the weight that can be carried at the edge of the manipulator. The application point is considered the flange of the wrist. The specified load depends on the speed with which the mechanism is moving, e.g. a greater load than the specified can be carried but with lower speed. *Repeatability* is the robot's capability to return to the same place after a number of repetitions. *Repeatability* is given as the range within which the robot will stop the motion. It is a very important factor when it comes to industrial applications, where motions are taught to robots. *Accuracy* is the robot's capability to go to the exact place it was ordered to go. Usually accuracy is based on the control system's discretization, the mechanical connections of the robot parts, and the minimum allowed error imposed by the operation of the servos. Accuracy is affected by the type and size of the load, contrary to *repeatability*. That is why several manufacturing companies specify only the *repeatability*.

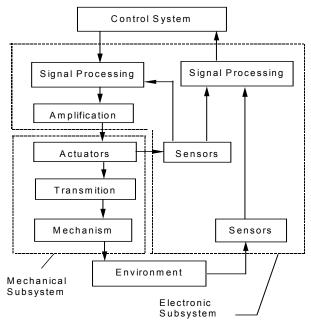


Figure 1. Basic architecture of a robotic system.

## Kinematics and Dynamics

Although robots with closed kinematic chain structure can be encountered, the fundamental structure of stationary base industrial robots is the open kinematic chain. One end of the chain is constrained to a base, while at the other end the end effector is connected to allow for object manipulation in space. The direct kinematics determine the position and orientation of the end effector based on the joint variables. The inverse kinematics determine the joint variables corresponding to a given end-effector position and orientation. For the inverse kinematics problem, a closed form solution is not always possible, a solution might not be unique, there might exist infinite solutions, or there might be no admissible solutions for the structure of the specific robot [1]. Here we will consider the case of direct kinematics. If we assume a reference frame  $O_0 - x_0 y_0 z_0$ , let  $q_i$  be the ith joint variable with i = 1,..., n. Let  $\eta$ , s, a be the unit vectors of a frame attached to the end effector and a the position vector of the origin of this frame. Then the end effector position is given as:

$$\begin{bmatrix} \eta^{0}(q) & s^{0}(q) & a^{0}(q) & p^{0}(q) \\ 0 & 0 & 0 & 1 \end{bmatrix} = A_{1}^{0}(q_{1})A_{2}^{1}(q_{2})...A_{n}^{n-1}(q_{n})$$
 (1)

with  $A_i^{i-1}(q_i) = \begin{bmatrix} R_i^{i-1} & r_i^{i-1} \\ 0^T & 1 \end{bmatrix}$ , where  $R_i^{i-1}$  the rotation matrix and  $r_i^{i-1}$  the translation

vector from frame i to frame i-1. For example the point **P** in Figure 2 has the coordi-

nates of the vector  $p^0$  with respect to the reference frame  $O_0$  and the coordinates of the vector  $p^1$  with respect to the reference frame  $O_1$ . Vectors  $p^1$  and  $p^0$  are connected with

each other through the transformation  $\begin{bmatrix} p^0 \\ 1 \end{bmatrix} = A_1^0 \begin{bmatrix} p^1 \\ 1 \end{bmatrix}$  that represents a rotation and

a translation.

The dynamic model or the robot provides the functional relationship between the joint actuator torques and the motion of the structure. With the *Lagrange* [5] formulation, the equations of motion can be derived in a systematic way. To this extend the set of variables termed *generalized coordinates*  $\lambda_i$ , i = 1,..., n is chosen. They describe the link positions of an *n*-degree-of-mobility manipulator. The *Lagrangian L* of the system can be defined as a function of the generalized coordinates:

$$L = T - U \tag{2}$$

where T is the total kinetic energy and U the potential energy of the system. The Lagrange equations are expressed by:

$$\frac{d}{dt}\frac{\partial L}{\partial \lambda_i} - \frac{\partial L}{\partial \lambda_i} = \xi_i \tag{3}$$

where  $\xi_i$  is the generalized force associated with the generalized coordinate  $\lambda_i$ . For an open kinematic chain robot, a usual choice of generalized coordinates is the vector of joint variables:

$$\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix} \tag{4}$$

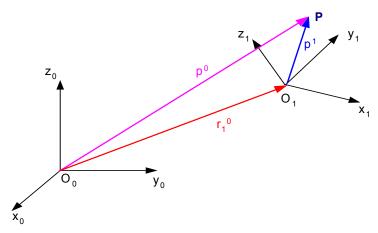


Figure 2. Representation of a point P in different coordinate frames.

#### Sensors

The sensors are a part of the electronic subsystem of the robot, which includes a variety of electronics and other low-power components. Those elements consist of measuring devices integrated in the robot, like optical encoders, potentiometers, analyzers, cameras, etc [2]. Another group of elements consists of analogue power electronics, which provide the motors with the needed electrical power, by appropriately amplifying a reference signal. Another group of electronics transfers the signals between subsystems by transforming them to a compatible form (analogue-digital converters, comparators, etc.).

In addition to sensors used for specifying the configuration of the robot itself, sensing of the surrounding environment is usually implemented. Environmental perception is usually realized through visual and tactile sensors. Visual sensors are used for object recognition or for gathering information regarding the presence of an object, its distance from the robot, the object's shape, its orientation, and its volume.

Robot vision systems may consist of several cameras and a processing unit. Cameras can be CCD, which provide very good resolution; they respond to a high bandwidth of light frequencies, are lightweight, and have low power consumption. The processing unit is usually realized with an integrated circuit (IC), which digitally processes the image signal.

Tactile sensing is realized through tactile sensors like piezoelectric crystals, microswitches, and force sensors [12].

#### **Motion Control**

The control subsystem [6] is the "brain" of the robot. It is responsible for monitoring, coordinating, controlling, decision-making, and giving commands to the rest of the system. All those functions are realized through a microprocessor, which is actually a complete computer. A typical control architecture is shown in Figure 3.

#### **Programming**

For a robot to carry out a task, it must first be programmed to do so. Describing a sequence of robot motions and activities and teaching the robot these motions and activities accomplish robot programming [3]. Robot programming can be divided into two groups: on-line programming and off-line programming. In on-line programming, the robot is involved at several stages during the programming process; in off-line programming, the robot does not participate in the program development, but only executes a predetermined program.

On-line programming is referred to as the *show-and-teach technique*. This technique involves manually moving the robot to each desired position. The resulting "program" is a sequence of joint coordinates and activation signals from external equipment, which is used to command the robot to perform the same motion at a later stage. Off-line programming is a language-based programming technique, similar to computer programming. With off-line programming, the programs are written in highlevel or low-level languages. Currently there are more than 100 robot languages. Table 1 shows some of the most widely used robotic languages.

100 Chapter 2 Hardware

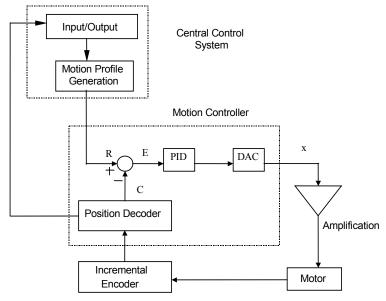


Figure 3. Control architecture.

Table 1. Typical robot programming languages.

Developer	Language	Robotic Platform
Cincinnati Milacron	Т3	Т3
Unimation	RPL, VALII	PUMA
Adept	VALII	Adept1
Sheinmann	AL, PAL	Stanford Arm
IBM	AML, Funky, Emily, Maple Autopass	IBM Arm
Bedix	RCL	PACS Arm
General Electric	Help	Allegro
Anorad	Anorad	Anomatic
Olivetti	Sigla	Sigma
Stanford	WAVE	Stanford Arm, Robovision
Automatix	RAIL	Autovision, Cyber Vision

## Typical Commercial Robots

The *PUMA* robot is one of the most known, especially in the scientific community. The word *PUMA* is an acronym of *programmable universal machine for assembly*. It has six axes of motion that allow for positioning of its end effector anywhere and with any orientation within its workspace. It is very dexterous due to its anthropomorphic structure and is easy to control due to its open architecture. It is used in industry as well as in research. It is manufactured by Unimation Inc.

The SCARA robot was designed by the Japanese professor Hiroshi Makino. It has great stiffness in vertical loading and elasticity in horizontal. That is why it is used for vertical assembly operations. Because the motors of the first two joints are in the first

link, the second link and subsequently the suspended part becomes lighter, allowing the robot to achieve greater speeds. Those robots are 10 times faster compared with other robot types. Adept (www.adept.com) is a manufacturer of such robots.

CRS ROBOTICS provides a variety of robots the most known of which are the anthropomorphic A465 and the Gantry G365. The anthropomorphic A465 is a six-degree-of-freedom robot, with a 3-kg payload and a repeatability of 0.05 mm. Usual applications of this robot are material handling, tool loading on machine tools, assembly, color spraying, and quality control. The G365 has a three-degree-of-freedom geometry, which allows it to reach any position within its three-dimensional workspace with a constant orientation. There is an option to add a two- or three-degree-of-freedom wrist for controlling the orientation. G365 can be used for material handling, quality control, packaging, assembly, etc.

Komatsu provides *LM15-1*, which is actually a handy robotic lift, able to be divided into two parts and carried on a pickup truck. Its weight is 520 kg and its payload ranges from 150 to 350 kg. Its maximum height is 4.2 m and it can manipulate objects on the ground level. Its price is about 30,000 US dollars.

## 2.4.4 Moving Base Robots—Wheeled Robots

## Description

Mobile robots are the ones capable to move any part of their mechanism. The motion capability is provided by specialized propulsion systems that can range from the very simple, such as wheels, to the more complicated, such as jets, propellers, or mechanical legs, depending on the operational environment.

#### Automatic Guided Vehicles

Mobile robots, according to their level of autonomy, can be divided to AGVs (automatic guided vehicles) and AMRs (autonomous mobile robots). AGVs have limited motion autonomy, and their motion is predetermined through ground cables or transmitters in the environment.

An AGV is a moving platform, usually powered by batteries but other mobile power sources are possible. Its motion is along predefined trajectories within an industrial environment. It can be used as a material transporter, a trailer, or a moving base for a manipulation device. Its trajectory can be a phosphorescent or magnetic stripe or a guiding cable integrated in the floor.

As shown in Figure 4, motion control is performed by a central computer that is not onboard the vehicle, and which can coordinate a number of vehicles. The computational subsystem that is implemented onboard the AGV has the less difficult tasks of controlling the motors of the vehicle to follow the predetermined trajectory and to establish two-way communication with the central computer.

#### Autonomous Mobile Robots

One of the goals in robotics is to create robots that will act with limited or no human intervention. The development of AGVs in this direction has resulted in AMRs

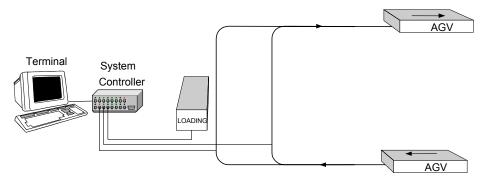


Figure 4. Autonomous guided system layout.

(autonomous mobile robots), which have no need of continuous external supervision and are capable of carrying out tasks in an autonomous manner, by accepting only high level commands.

Research institutes all over the world are studying mobile robots because their applications are limitless. Wheeled robots have been studied the most. They are used in a variety of applications, including object transportation, helping people with disabilities, security services, and in various agricultural applications such as planting, seed-bed preparation, spraying, cultivation, etc. Operating in environments hazardous to humans is one of their noteworthy applications. Those applications range from supervision and diagnosis for nuclear power stations to rescues to performing dangerous assignments like mapping hostile areas or disarming minefields.

The basic differences of AMRs with AGVs are at the level of autonomy, the capability of performing various tasks, and intelligence. Both utilize a motor-driven wheel subsystem, a sensing subsystem for position estimation, environment perception and navigation, a processing unit and of course the power source. They can both be categorized further based on the previous characteristics.

## Steering Systems of Mobile Robots

Both AGVs and AMRs can be characterized by their steering systems. From the variety of solutions used for the steering system [7], we mention several here.

Differential steering is a mechanical structure with two thrusting wheels, which can rotate with different angular velocities to allow steering and kinematic consistency. In addition to the thrusting wheels, there exist a number of non-thrusting wheels called castor wheels that help support the vehicle.

The Ackerman differential is a well-known geometry [8] usually met in outdoor robots. It is a very reliable structure. To establish kinematic consistency the turning angles of the two front wheels must obey a kinematic constrain, as shown in Figure 5, to have the same center of rotation

With the *synchronous steering system* all wheels turn while steering. The system exhibits stability and adequate traction. A proper torque distribution provides the ability to overcome obstacles.

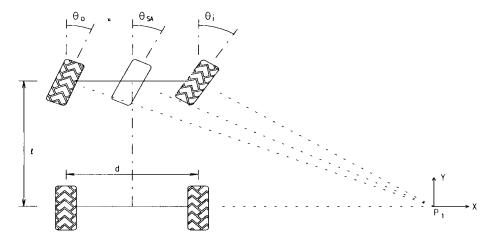


Figure 5. Kinematic constraints for the steering wheels.

All the above-mentioned systems are subject to kinematic constraints called *non-holonomic* (i.e., vehicles cannot move in a direction perpendicular to the direction of their wheels). Although there are steering systems that overcome those constraints (*omni-directional steering*), they have increased friction.

## Integration

The architecture of an autonomous wheeled robot is more complicated than that of AGVs. The robot must gather and then process, by itself, information from its environment and from itself, and most of all to make decisions [9,11]. Its architecture is summarized in Figure 6.

## 2.4.5 Future Directions and Prospects

#### **Underwater Robots**

Unmanned underwater vehicles (UUVs), either remotely operated ROVs (remote operated vehicles) or autonomous AUVs, constitute a technological challenge [4] with a very important economical and social impact due to the need to exploit the wealth of the oceans as well as various climatological and ecological problems directly related to the sea. UUV exploitation has been limited due to the operational costs.

They have many similarities, but the basic difference between ROVs and AUVs is that ROVs are connected to the mother ship with a cable used for powering and communications, while AUVs are not connected (Figure 7). ROVs are box-shaped and move at a low speed or hover. AUVs move with high speed, and so they are streamlined; conventional AUVs cannot hover. Due to their infinite power supply and the teleoperation capability provided by the cable, ROVs are used in underwater construction. AUVs have greater autonomy due to the absence of the cable, but their power resources are limited to the onboard batteries. Communication with the surface

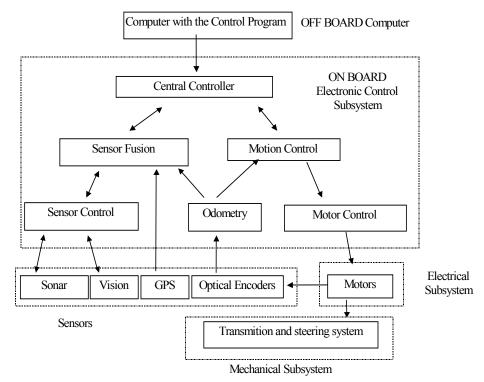


Figure 6. Architecture of an AMR. An off-board computer sends high level commands to the onboard central controller. The on-board electronic control subsystem interacts with the sensors and the electrical subsystem to produce the appropriate actuations of the mechanical subsystem.

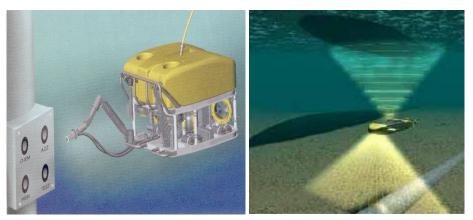


Figure 7. Conceptual artistic drawings of an ROV (left) and an AUV (right).

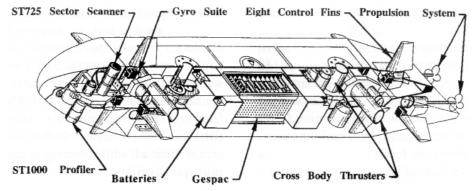


Figure 8. AUV showing its subsystems.

is performed with sonars. AUVs are usually used for inspection of simple structures like underwater pipes.

Figure 8 shows the subsystems of an AUV. The advanced sensors used on AUVs are multibeam sonars (MBE), sidescan sonars (SSS), sub-bottom profilers (SBPs) and CTD probes.

#### Aerial Robots

Aerial robots are a rapidly growing field of robotics. Although there are a large number of possible applications of aerial robots, their applications have been limited so far to military purposes, since they are a fairly new technology and are not yet considered to be stable enough to be used in sensitive areas like cities or close to other air traffic. However, as aerial robot technology gets cheaper and more stable, they will be used in other applications, such as agricultural spraying or monitoring and coordinating ground AMRs. Other applications include traffic surveillance, today done by helicopters in large cities, and environmental and meteorological inspection, where the number of weather measurement stations is key and the main problem is the need to get measurements from sea stations. Normal traffic planes are used, together with sea transporters, as mobile weather stations, but these only provide data from the standard flight routes and altitudes.

An example of a remotely controlled airplane used for environmental recognition is Helios by NASA, shown in Figure 9a. Since no pilot is needed, this solution is cost-effective. Figure 9b shows the AVATAR (autonomous vehicle for aerial tracking and retrieval), developed under the Autonomous Flying Vehicle project at the robotics lab of the University of Southern California. Other applications where the human pilot would not be needed are regular transportation, pesticide applications, etc. The main limitations to the exploitation of this technology come from the airplane stability and governmental laws and regulations. There is already an organization, Euro UVS, created by universities and companies working together with governmental air-safety bureaus.





Figure 9. (left) NASA's Helios and (right) robotic helicopter of the University of Southern California robotics lab.

## Walking Robots

The main motivations for using walking robots rather than more conventional wheeled ones for certain tasks are due to their versatility for obstacle avoidance, their adaptation capability on uneven ground, and their ability to climb irregular surfaces. Figure 10a shows the Dante-II robot of NASA's JPL lab during mountain climbing in Alaska. The most common walking robots are bipeds. Figure 10b shows the Asimo biped, a humanoid robot of the Honda Motor Company. Biped robots are a particular class of under-actuated systems and pose very interesting control problems. The characterization as under-actuated is due to the fact that they are not able to produce torque at their ankles.

Walking robots exhibit a non-holonomic behavior during the flight phase and they are not globally controllable during the single support phase during which they act as an under-actuated pendulum-like system. Biped motion under a walking cycle is





Figure 10. (a) Dante-II of NASA's JPL lab during mountain climbing in Alaska. (b) The Asimo Humanoid robot of Honda (courtesy of American Honda Motor Company, Inc.).

composed of the single support phase (one foot on the ground), the contact phase (transition phase), and the double support phase (both feet on the ground). Hence this is a system with multiple dynamics (one for each phase), subject to unilateral constraints (ground presence) and hybrid in nature (discrete changes due to the transition phases). The complexity of the walking mechanisms make their control and analysis difficult and adequate mathematical techniques accounting for the periodic nature of the biped's motion and the dynamic behavior at the shocks should be considered.

#### 2.4.6 Conclusions

Today robots are a fact. The increasing use is inevitable since their cost is decreasing and their functionalities increasing. In the future, due to competition, only the most productive enterprises will be able to survive, and robots are expected to serve a big role in this.

## References

- 1. Sciavicco L., and B. Siciliano. 1996. *Modeling and Control of Robot Manipulators*. Columbus, OH: The McGrawHill Companies, Inc.
- 2. Everett, H. R. 1995. Sensors for Mobile Robots. Wellesley, MA: A. K. Peters Ltd.
- 3. Flynn, A. M., and J. L. Jones. 1993. *Mobile Robots Inspiration to Implementation*. Wellesley, MA: A. K. Peters Ltd.
- 4. Fossen, T. I. 1994. *Guidance and Control of Ocean Vehicles*. New York, NY: John Wiley and Sons.
- 5. Baruh, H. 1999. *Analytical Dynamics*. WCB/McGraw-Hill.
- 6. Canudas De Wit C., B. Siciliano, and G. Bastin. eds. 1996. *Theory of Robot Control* (Communications and Control Engineering Series). Berlin, Germany: Springer Verlag.
- 7. Wong, J. Y. 2001. *Theory of Ground Vehicles*. Wiley-Interscience.
- 8. Colyer, R. E., and J. T. Economou. 1998. Comparison of steering geometries for multi-wheeled vehicles by modelling and simulation. *Proceedings of the 37<sup>th</sup> IEEE Conference on Decision and Control* 3: 3131-3133.
- 9. Chatila, R. 1994. Control architectures for autonomous mobile robots. *Proceedings of the IEEE Conference From Perception to Action*, 254-265.
- 10. Edan, Y., D. Rogozin, T. Flash, and G. E. Miles. 2000. Robotic melon harvesting. *IEEE Transactions on Robotics and Automation* 16: 831-835.
- 11. Edan, Y., and G. E. Miles. 1994. Systems engineering of agricultural robot design. *IEEE Transactions on Systems, Man and Cybernetics* 24: 1259-1265.
- 12. Allotta, B., G. Buttazzo, P. Dario, and F. Quaglia. 1990. A force/torque sensor-based technique for robot harvesting of fruits and vegetables. *Proceedings of the IEEE International Workshop on Intelligent Robots and Systems* (1): 231-235.
- 13. Oilfield Publications Limited (OPL). 2000-2001. Remotely Operated Vehicles of the World. England: OPL.
- 14. www.ece.eps.hw.ac.uk/~autotrack/. Accessed 2002.

# Methods, Algorithms, and Software

# 3.1 Modeling and Simulation

M.-J. Cros, F. Garcia, R. Martin-Clouaire, and J.-P. Rellier

Abstract. Modeling and simulation are major assets in the engineering of systems of all types, in particular for production management problems. This section provides an overview of the basic aspects and key issues in modeling and simulation of agricultural production systems. Particular attention is given to the kind of structures, processes, requirements, and difficulties that have to be dealt with in the modeling of the interactive processes underlying management activities and biophysical components. Simulation can be used in combination with optimization techniques to find settings that optimize expected performance criteria; various such techniques are surveyed. The issue of calibration and validation of an agricultural production system model is also addressed.

**Keywords.** Agricultural production systems, Production management, Decision, Modeling, Simulation, Optimization.

#### 3.1.1 Introduction

The field of *computer simulation* is usually combined with the term *modeling* to form *modeling and simulation* or *simulation modeling*. To model is to create a structure that captures interesting or noteworthy attributes and processes of an object of study. A model is a simplified description of a system through a computerized representation that enables us to conduct virtual experiments about its behavior. Basically, computer simulation is the imitation for some period of consideration of the functioning of a system that accepts inputs, produces outputs, and interacts with its environment. A model of a system in its environment is an abstraction of reality, and as such, certain details are excluded from it. It consists of a set of instructions, rules, equations or constraints that generate input/output behavior once processed by a simulation mechanism.

Simulation modeling is an amazingly diverse topic whose coverage can be found in just about every discipline. In particular, modeling and simulation are major assets in the engineering of systems of all types, in particular for production management problems. This section focuses on the modeling and simulation of agricultural production systems, which constitute one of the main application domains in agricultural engi-

neering. The modeling of an agricultural production system essentially aims at using virtual experimentation as a means to predict performances and improve or design the production process, in particular, the management commitments underlying this process. Simulation models of the production system are useful where agronomic experimentation results are lacking, incomplete, too expensive, or impossible to get. They can contribute to furthering our understanding of the processes involved and enable us to study the sensitivity to different factors, economical performance, resource use feasibility, and environmental impacts for different weather scenarios and management practices. These models can also be used to determine empirically or by use of mathematical optimization techniques the most appropriate management principles that enable us to adapt to climate variability in order to reach desired economic ends and meet the various constraints (e.g., resource limitation, environmental regulation). In addition, by virtue of their capabilities to demonstrate how the systems work, these models can also be used for educational purposes.

In this section our intent is to provide the reader with an overview of the aspects and issues in modeling and simulation of agricultural production systems. Section 3.1.2 surveys the types of modeling and simulation approaches that might be used depending of the model purpose. A short presentation of the simulation mechanisms underlying these approaches is provided in 3.1.3. The subsequent two sections discuss conceptual aspects involved in modeling the two interactive processes at work in any agricultural production system. They concern the biophysical components at the core of the production process (the crop and/or animal components in their physical settings) and the human activities that include the sequential making and implementation of decisions about the technical operations to perform on the biophysical system in order to reach the intended objectives. The focus there is on the kind of structures, processes, requirements, and difficulties that have to be dealt with in the modeling of management (3.1.4) and biophysical (3.1.5) aspects. Section 3.1.6 briefly discusses the use of weather generators that are required in order to create sets of artificial weather scenarios representative of likely cases in the given region of interest in the simulation study. Simulation can be used in combination with optimization techniques to find settings that optimize expected performance criteria; a survey of the various approaches is given in 3.1.7. The issue of calibration and validation of an agricultural production system model is addressed in 3.1.8. Some concluding remarks are made in 3.1.9 about the potential benefits that one can expect from simulation modeling of agricultural production systems and about the current developments undertaken to enhance and extend these capabilities.

### 3.1.2 Different Questions Imply Different Modeling and Simulation Needs

Many computer models of agricultural production systems have been developed. Depending on their specific purpose and intended use the models are quite different in terms of the modeling and simulation approaches they rely on. The models might be deterministic or stochastic if uncertainty plays a central role in the phenomenon we seek to demonstrate, understand, or cope with. A model might be static if intermediate steps of the production process are of no interest; it needs to be dynamic if we want to

follow interactions as time goes on. Statistical relations may suffice for some phenomena, whereas mechanistic models of the functioning of some processes may be required in order to be able to extrapolate outside the range of data available and to perform fine-grained simulations, where order and time of occurrence of events matter. A dynamic model may be continuous (e.g., a system of differential equations) or discrete-event in which properties change instantaneously at separate instants. The model of the external environment is also an important difference between simulation systems in terms of what is taken into account (e.g., climatic factors) and whether there are stochastic inputs. Finally, another important discriminant feature concerns the representation of the decision-making behavior and the various production constraints to take into account in decision making and implementation of decisions.

The software systems that, to some extent, simulate farm or field production systems can be looked at from the standpoint of their use during simulation: users may interact with the simulator to provide decision inputs as the execution progresses or may passively watch the results. Examples of the first type of use include on-line decision-making support (guidance based on real data) by exploiting the predictive power of biophysical models. Crop problems or requirements (water stress, fertilizer needs) can be foreseen ahead of time and testing of an appropriate anticipation can be done on computer or looked for with computer help [1]. Another application where on-line use of a simulator might be helpful is to study the on-line reaction of an operator in the course of realizing a particular task: the user can enter decisions as the simulation proceeds step by step (by day or week, for instance) and can learn to operate better, by their mistakes or poor performance results. In these cases, only a dynamic model, either mechanistic or empirical, of the biophysical system is required; no decision-making model is needed.

Simulation software working in the passive mode is more common. The simple examples concern models aimed at budgeting analysis and production system configuration. A static model constituted by simple algebraic relations (a spreadsheet model) involving decision variables may suffice. In case some parameters are uncertain, probability distributions can be assumed and the prediction can be done by Monte Carlo simulation. This computes the probability distributions of the variables connected to these parameters. In another class of applications, simulation is used as a device for helping capture and communicate intelligibly and convincingly how things work (i.e., interactions). By providing a well-founded, encompassing and shared body of knowledge about the behavior of a production system of interest, this capability can serve as a virtual experimentation platform for multiparticipant training sessions involving extension agents and farmers; see for instance the DSSAT family of systems [2]. Thanks to insightful visual presentation, simulation is helpful in giving an overall understanding about how the system functions and creating objective opinion and consensus about where some important issues lie. The underlying model, in particular as regards the decision-making aspects, may be more or less complex depending on the target population. Such systems may also be able to provide a rough estimate of the economical performance or environmental impact of the management options across a range of climatic scenarios. They might also be used to support diagnosis and for explanatory purposes. Indeed, they can help in understanding why certain physiological phenomena occur (Why did this growth problem happen on this crop?) by exhibiting the chronology of interactions that are responsible for the observed results.

Usually these systems are essentially crop models and only highly simplified management situations are dealt with under the form of a set of fixed management options (e.g., crop species, sowing date, rate of nitrogen fertilization). Systems belonging to this line of work have very limited capabilities to represent management practices and to take into account production constraints during a season. They do not allow simulated decisions to be planned and made in relation to the dynamic conditions on the farm both as regards biophysical states and operational constraints; decisions have to be made before each simulation regardless of what might happen and feasibility considerations. Consequently, they provide little help in evaluating realistic management strategies and operating procedures, identifying production resource bottlenecks (e.g., peak labor force demand), and designing by iterative local improvements in new management solutions adopted to changes in the technical, economical, and social context. Overcoming these limitations to address management issues with simulation approaches is still a challenge that requires modeling human activities as an integral part of the production system model. What is involved in the modeling of human activities is the subject of the next section. The subsequent section deals with the modeling of biophysical aspects.

#### 3.1.3 Simulation Mechanisms

As discussed in the preceding section the modeler's way of representing systems might be different depending on the model purpose. The underlying simulation algorithms are specific to the modeling approach chosen. The book by Zeigler et al. [3] is a good technical reference about the continuous and discrete systems approaches that are briefly presented here together with the popular spreadsheet approach.

### Spreadsheet Simulation

Spreadsheet simulation refers to the use of a spreadsheet as a platform for representing the simulation model and performing the simulation experiments. For agricultural production system application (e.g., [4]) this approach is often employed when the study can be done with a simple static model that represents mathematical and logical relationships between variables (e.g., feedlot ration analysis). Essentially a spreadsheet is an electronic grid consisting of rows and columns of cells. Each cell can be addressed individually and can contain data or a formula. The resolution mechanism propagates any change in the data cell to the formula cells referring directly or indirectly to the data cells. The data cell can be random numbers and propagation then employs Monte-Carlo computation techniques.

Spreadsheet simulation is popular thanks to the wide availability of its software, its intuitive interface, and its ease of use. At the same time, this approach is not recommended for complex models that require dedicated simulation tools implementing one of the next approaches presented.

### Continuous Systems

The continuous systems approach assumes continuous state variables and time through differential equation formulation: the rates of change of the state variables are defined by derivative functions. See [5] for an example of the use of this approach in modeling an agricultural production system. The simulation mechanism is based on a numerical integrator [6] such as the Euler or Runge-Kutta method that uses a discretized time base.

Systems dynamics [7] software implements this approach within a framework supporting graphical representation of the structure of the model. System dynamics models are centrally concerned with cyclic relationships that represent positive and negative feedback loops. This approach looks at systems at a very high level of abstraction, which makes it more appropriate for strategic analysis than for a detailed analysis as required, for instance, in process re-engineering. The next two approaches are more often used for this latter purpose.

### Discrete Time Systems

Discrete time systems (e.g., [8]) assume a stepwise mode of execution. The dynamics of the system are represented by difference equations or more generally by transition functions that each express how to update a state variable on the basis of the state at the previous time step and the inputs (influencing factors). The simulation mechanism relies on a fixed-step iterative algorithm that jumps from one simulation step to another and computes the next state given the state and input at the current time. All the model variables are scanned at each step.

The finite state automaton formalism is subsumed by discrete time systems. Cellular automata are particular examples widely used for studying physical spreading phenomena (e.g., infestation propagation) or group phenomena (e.g., population dynamics).

### Discrete Event Systems

In the discrete event system approach [9], the modeling of the dynamic behavior of the system is similar to that used in discrete time systems: transition functions specify local changes. In addition, this approach relies on the identification of events that cause the transitions to be fired at their occurrence time. The important difference with discrete time systems is in the mechanism of event processing, which jumps from one interesting point in time (when an event occurs) to another. It scans only time points and variables concerned with the current event. As long as no events occur, no state changes are made. The simulation clock is event-driven. Discrete event simulation works by maintaining a list of events (agenda) sorted by their scheduled event times. Events are treated by taking them from the event list in sequential order and executing the associated processes, which produce state transitions. Processing events results in new events being scheduled and inserted into the event list, as well as events being canceled or removed from the event list. Events can also be caused by the environment, which is not under the control of the system itself.

# 3.1.4 Human Activities in a Production System

As argued in Section 3.1.2, if studying production management is the main purpose of a simulation project, one has to look at the production system as a human-centered process that involves sequential sensing, decision making, and execution of actions. At the core of such an investigation is the problem of gaining a better understanding of the interactions and interrelations that occur throughout the production period among the human actors (decision makers and operating agents), the biophysical entities (e.g., fields, crops) and the events in the external environment (e.g., climatic circumstances). A prerequisite for such a modeling undertaking is to have a comprehensive view of what these human activities are about and to identify and define some basic concepts enabling the description and modeling of this part of agricultural production systems.

Little attention has been paid to the managerial practices of farmers. See [10] and [11] for analyses reflecting essentially the thought of the French farming systems community. What kinds of activities does a manager perform and what are the distinguishing characteristics of this managerial work? The managerial functions involved in dealing with the technical aspect of production (excluding budgeting and marketing aspects) include:

- *Organizing*, such as configuring the production system resources both material (land, machine) and human (hiring, role assignment);
- Planning, such as formulating an outline of activities to be done along the production season and methods for their execution with respect to production objectives and constraints;
- Information and knowledge handling. Decision making is based on the ability to handle information and knowledge (what has to be acquired and stored, when to acquire it or recall it, how to use it in decision making, either directly or in a derivation process of a decision-relevant synthetic indicator). The manager must therefore have monitoring and forecasting activities in which he seeks and acquires information to be aware of the present situation of the production system and envision its likely evolution.
- Determining actions and commanding, such as identifying possible courses of action in face of the current or anticipated situation (essentially in the biophysical domain) and constraints (availability and compatibility of resources, compatibility of activities); selecting actions in function of different criteria such as cost/benefit considerations, continuity and harmonization of activities, urgency (having some activities finished or a state achieved before a deadline, having particular resources released by a date); and issuing instructions to cause activities to happen in a specific manner.
- Event handling, such as looking at the environment for opportunities or disturbances and initiating or devising changes in planned activities, i.e. if the wind is strong then postpone spraying.

In his management task the farmer has to deal with several interactive dynamics concerning the biophysical components, the environment of the system of interest (external events) and the unfolding of the technical actions resulting from the decision he has made. Most actions are durative; their execution may be disturbed, interrupted, and even never finished. The farmer has to combine planned and reactive behaviors in order to organize the work as a function of known and exploitable regularities and adapt to contingencies as they occur. Therefore, the tight coupling between sensing and decision making in a timely manner is of primary importance. Managing an agricultural production process is a dynamic process in which plan revision and execution must be interwoven because the external environment changes dynamically beyond the control of the farmers and because relevant aspects are revealed incrementally.

The decision-making behavior of farmers in their management task has been the subject of different kinds of investigation, but the mental process that intervene between stimulus and response and by which that behavior is exhibited is still largely unknown. Nevertheless, the concept of management strategy is often used as a means to express beforehand the farmer's management behavior. These management strategies do not pretend to reproduce what is happening in the manager's head. More modestly they attempt to enable the derivation of what the farmer does depending on the current state of affairs. A management strategy can be seen informally as a manually elaborated construct that specifies a kind of flexible plan coming with its contextresponsive adaptations and the necessary implementation details to constrain the stepwise determination and execution of the actions to perform. Due to evolving and unpredictable circumstances the plans are flexible with respect to the temporal organization of the constituting activities. The commitments to particular activities are delayed until run-time conditions are known. In particular, what can be executed is strongly constrained by the availability of resources and state-dependent requirements on the operations suggested by the plan.

The modeling of human activities should not be of a black-box type when management aspects are central for the study carried out with simulation. How can we determine that a management strategy is likely to succeed or can be improved if there is no explicit representation of it? A model of human activities requires a computational and declarative representation of the activities, execution processes, information, resources, constraints, and behavior involved in the production process. The model should enable us to determine the impact of changes on all parts of the production process. For example, if some activities are inserted, deleted, modified, or coordinated differently with others, the model must be able to represent how this will affect the resource use, economic performance, and other aspects of interest. The model should make explicit what is planned, what might happen, and, more generally, any information and knowledge necessary in managing and operating the production process.

The agent modeling [12, 13] and enterprise modeling [14] fields of artificial intelligence provide a set of inspiring and useful formalisms for representing such knowledge structures and afferent processing mechanisms. See [15-21] for agricultural production system examples involving some modeling of human activities.

# 3.1.5 Modeling the Biophysical Aspects of the Production System

Many biophysical models [22] are described in the agronomic literature, but very few are designed to be used within agricultural production models that also include management models. For instance, most published grassland models address only the herbage growth or vegetation dynamics aspects, but these models cannot be used to address pasture-based production systems because dealing with the plant/animal interaction is required. The choice of the processes to be modeled, the management variables to be considered, and the level of detail must be determined by the intended use of the simulation model. The biophysical models embedded in agricultural production models [23] often require the user to integrate knowledge from different areas of expertise (crop science, animal science, farming systems research) from a farm management perspective at the seasonal scale. Such models often operate at farm scale.

The models must be able to react dynamically to external stimuli and to actions prescribed along the management process. They must account for the influence of climatic factors such as rainfall, temperature, and solar radiation. The other driving variables of such models should include the main factors that farmers can control (for instance, in the case of a pasture-based dairy production system, nitrogen fertilizer rate, defoliation frequency and intensity, composition of cows' diet). Since the management commitments to be evaluated refer to state variables, the models must provide accurate enough information about their current values. They have to incorporate all the variables playing a role in the biophysical processes and those that are looked at in evaluating the simulation results.

Making the biophysical component of the simulation model as simple as possible is an important design principle in order to facilitate its parameterization or make it tractable. It would be unwise to model some biophysical processes in great detail if the parameters involved could not be obtained in current conditions. Straightforward assembly of existing detailed research models is inadequate; one should start by identifying the requirements imposed by the purpose of the model and find out the appropriate level of comprehensiveness and the temporal and spatial scope. Clearly, the degree of abstraction and coarseness of inner processes are an impediment to the pertinence, usability, and acceptance of the model. For instance, choosing field and sward rather than plants or organs of plants as the basic units of the model has shown to be a crucial design step. Having more than one hierarchical level below the level of immediate interest would not necessarily serve the purpose any better because it leads to greater complexity.

# 3.1.6 Random Inputs: Stochastic Weather Generators

Long-term weather data are commonly needed for evaluating the impacts of different management strategies in agricultural production models. However, long-term weather records are generally not available and weather generators have to be used in order to produce weather data long enough to assess agricultural risk. A stochastic weather generator is a model that produces on a daily or sub-daily basis weather data (such as precipitation, temperature, solar radiation, and wind speed) that have the same statistical characteristics as observed weather data for a given location.

Stochastic weather generators are playing a more and more important role in decision support systems in conjunction with simulation models of agricultural production systems. They are generally used to estimate, for every management strategy, the probability distribution of different output variables of interest (such as yield, economical margin, and labor hours) in order to compare them and to select the best management strategy. Stochastic weather generators can also be used in real time. Given the observed data at the current time and a set of possible future weather data produced by the weather generator, the distribution of simulated crop yields, for instance, can be obtained.

Several stochastic weather generators, such as WGEN [24] or LARS-WG [25, 26], are now available and can be used in agricultural production models. They differ in structure and complexity. However, these generators generally work similarly by estimating the model parameters on the basis of some observed daily weather data for a given site, and then by producing daily weather data, using a pseudo-random number generator.

# 3.1.7 Simulation Optimization

The problem of optimizing an agricultural production system is in finding the best management strategy that yields the optimal expected performance. When a strategy is defined as a sequence of decision rules, this optimization problem can be seen as a control problem (see Section 3.2), for which dynamic programming methods exist [27]. In this approach simulation models can be used for estimating the rewards [28] or the probabilities of the state transitions [29] when a stochastic dynamic programming model is used. However, dynamic programming methods are often inappropriate in cases of large state and decision spaces, despite some promising improvements that have been obtained recently with the reinforcement learning approach. This approach, also called neuro-dynamic programming, directly approximates the solution of the dynamic programming procedure during simulation [30].

A more flexible and realistic formulation of the optimization problem consists in searching, with the use of simulation, for the best values for the parameters of a predefined strategy. In that case the simulation model is considered as a black box function where the vector of strategy parameters  $\theta = (\theta_1, ..., \theta_p)$  is taken as input variable, and where the output is the objective function J. In most of the simulation models of agricultural production systems, the objective function J also depends of the current climatic series, which has to be considered as an unknown and uncontrollable random variable  $\xi$ . Optimizing a strategy thus consists in searching for the set of parameters  $\theta^*$  that maximize the expected value of the objective function J (assuming J models a benefit function) according to the random climatic series. In mathematical terms:

$$\theta^* = \operatorname{argmax} E(J(\theta, \xi)), \text{ with } \theta \in \Theta$$
 (1)

When the value domain  $\Theta$  is in  $R^p$  (the p-dimensional Cartesian product of reals), different efficient methods have been developed for solving this stochastic simulation optimization problem, which belongs to the most difficult problems of mathematical programming [31-33]. Until the beginning of the 1990s, these methods involved

essentially stochastic approximation algorithms. Basically, they are gradient search procedures that approach to  $\theta^*$  using some type of gradient estimation technique (e.g., perturbation analysis). Stochastic approximation is a variant of the steepest descent gradient search from deterministic optimization, and convergence to locally optimal parameters can be guaranteed under appropriate conditions. Also designed for continuous problems, the Nelder-Mead (simplex) method is an alternative to stochastic approximation; it is not based on a gradient but instead uses a geometric figure to move from one point to the other on the search space.

For discrete optimization problems, ranking and selection methods and multiple comparison procedures have been designed for selecting the optimal parameter  $\theta^*$  over a small finite-set  $\Theta$  having no more than 20 elements. If the set  $\Theta$  is larger, the ordinal optimization approach first obtains a rough ranking of the different alternatives, and then discards all but the r top alternatives, where r is a small positive integer. Then ranking and selection or multiple comparison methods are used for locating the best solution among the r candidates.

General search methods can also be applied for solving very large discrete problems, and can even be adapted for continuous problems. Random search methods move iteratively from a current point to another point in its neighborhood, and can be guaranteed to converge to the set of global optimal solutions. Recently, deterministic optimization methods like genetic algorithms, simulated annealing or tabu search have also been adapted for the stochastic environment associated with simulation. In these approaches, the criterion  $E(J(\theta))$  is estimated by averaging the objective function  $J(\theta,\xi)$  over a large number of sampled variables  $\xi_i$ :

$$E(J(\theta,\xi)) \approx \frac{1}{N} \sum_{i=1,N} J(\theta,\xi_i)$$
 (2)

When the  $\xi_i$  are fixed, the objective function J thus becomes a deterministic function of the input variables  $\theta$ , and a range of classical optimization algorithms can be used. Note however that a good estimate of  $E(J(\theta, \xi))$  may require a large number N of samples and thus multiple simulation run replications. This technique, called sample path optimization, thus converts a stochastic problem into a deterministic one. It was originally developed for solving continuous parameter simulation optimization problems, and has recently been studied in a variety of contexts.

More recently, stochastic versions of branching methods for discrete global optimization problems have been developed including, in particular, a stochastic version of the branch-and-bound method and the nested partition method, which hierarchically partitions and evaluates by random sampling the search space.

Evolutionary algorithms and other standard deterministic optimization methods have all been applied to agricultural simulation models in the last years [34], where they have shown good performance. However, most of these optimization problems consist in optimizing postseason optimal management decisions, based on a complete knowledge of the weather (e.g., [35-39]). Conversely, very few optimization methods have been applied so far to agricultural systems for optimizing the expected value of

an objective function, as described above. See, for instance, [40] where the Kiefer-Wolfowitz stochastic gradient method is used to derive the best values of some parameters involved in a grazing management strategy, or [41] where irrigation strategies are optimized by using the Nelder-Mead simplex algorithm. The easy availability of stochastic weather generators should lead to a fast development of this stochastic optimization approach.

#### 3.1.8 Calibration and Validation

Calibration involves estimating the values of various parameters in the model structure on the basis of real world data. Such a task is commonly accomplished with specialized statistical computer programs designed for just such purposes. In some cases model calibration is also performed in-line by exploiting acquired data from real field measurements during implementation of the model. The errors computed from the model predictions and the subsequent measurements are back-propagated for tuning the model. This procedure is very useful when little data are available for initial model calibration or when the modeled process evolves in time by itself. Model calibration can also be accomplished by using values of parameters from models estimated for another context that is similar to the one being considered; this strategy is referred to as importing model parameters and should be employed only by experienced practitioners.

Classically, validation of a model involves experiments comparing model behavior against actual measurements. Obviously this approach is feasible only in cases of models involving a relatively small number of variables and parameters. Validation becomes increasingly difficult for an agricultural production model which is expected to provide realistic estimates of different dynamic aspects varying over a period of several months. The extent of variation of most input variables (weather and management) is large and precludes any systematic exploration. Consequently, it is impossible for this kind of simulation model to be validated over its entire domain of application [42]. Moreover, records from the observation of the natural system (the production system and weather) are scarce and incomplete with respect to the set of aspects listed above and the time frame of interest. Although these data might be used to perform standard statistical validations of some parts of the model there is no assurance that the assembled model will necessarily behave acceptably well. Some errors may be introduced through linking model components at a higher level. The various parts of the model may be unequally checked and some interactions may not be predictable.

Hence, the only possible approach is a subjective one in which scientists and experts in the agricultural production domain are provided with simulations of cases familiar to them and asked if the model behavior is consistent and reasonably accurate. The validation consists then in checking that the results of simulation are in agreement with those expected by the knowledgeable people involved in the validation process. Furthermore, validation has to form part of the development of the whole simulation system; the results of evaluation provide feedback to make corrective changes to the biophysical model. Ultimately validation must provide substantiation that a computer-

ized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model.

The validity of the model is defined in terms of its ability to reproduce faithfully and accurately biophysical system behavior induced by a particular management strategy and its consistency and sensitivity for different weather patterns. Strategies are evaluated with respect to user-defined criteria that are valued as a function of output results provided by the model. Typically, the criteria are concerned with the dynamics and timing of key events or the magnitude or trend of relevant quantities. More specifically the kinds of outputs that might be used in evaluation include:

- a time series of daily values of model variables;
- the date and duration of operations;
- a time series of daily values of user-defined variables, that is, variables defined as functions of model variables for inspection purposes or for synthesizing decision-relevant indicators:
- a calendar of key events, that is, the date of occurrence of decision-relevant situations:
- histograms of variable values simulated for different climatic scenarios.

The analysis of outputs obtained with particular farm configuration and management options needs to be made using the weather pattern assumed for the simulation.

### 3.1.9 Conclusion

Thanks to their ability to model complex things as they are, simulation models offer great potential for the study and design of agricultural production systems and for education and training. Simulation, in the spirit advocated in this section, enables us to evaluate management practices and understand why and in what cases they may perform acceptably well or fail. The approach can be used to provide more compelling evidence to demonstrate beforehand that the production processes are in compliance with norms of social acceptability. Simulators can make insightful and reliable projections in a range of external environment scenarios, which nobody would even attempt due to the complexity of the system. In particular, they can support uncertainty analysis (how robust is a management strategy), timing analysis (when is some particular phenomenon occurring in the absolute time or relatively to another phenomenon) and resource use analysis (what are the critical needs). Simulation can stimulate farmers' thoughts about the management problems and potentially augment their innate knowledge-handling abilities. From the point of view of the model developers, the activity of modeling may reveal new ways of thinking about the decision domain and help partially formalize aspects of decision making.

Simulation modeling is no longer limited by computational difficulties thanks to computing power advances. Comprehensive systems that integrate all the aspects discussed in this section are, however, just starting to emerge, in particular for the modeling of human activities. Developing integrative farm-level models is new and little methodological support is available to support the learning/discovery process: how to choose what to feed to the simulator (what to experiment) and how to analyze outputs when inputs involve management strategies. Some aspects of agricultural production,

such as precision agriculture and beyond farm-level interactions, are in need of further investigation with the simulation modeling approach. The modeling task could be eased by promoting reuse of knowledge and software modules and developing generic frameworks dedicated to agricultural production systems.

Simulation modeling of agricultural production system requires a multidisciplinary collaboration between specialists of various domains including crop and animal sciences, farming systems, ergonomics of agricultural production system, agent modeling, and software engineering.

This section has focused on the modeling and simulation of an isolated agricultural production system. There is, however, growing interest in using computer simulation to explore and clarify the relationship between small-scale features observable at the level of the individual production system and large-scale, societal (or *macro*) phenomena that emerge from the micro-level interactions. This is the subject of multi-agent simulation that has its roots in distributed artificial intelligence [43]. Multi-agent simulation approaches [44] have great potential to study agricultural problems involving resource sharing (e.g., irrigation water) or environmental consequences of the activities of a community.

# References

- Cabelguenne, M., C. A. Jones, and J. R. Williams. 1995. Strategies for limited irrigation of maize in south-western France: A modeling approach. *Trans. ASAE* 38: 507-511.
- Jones, J. W., G. Y. Tsuji, G. Hoogenboom, L. A. Hunt, P. K. Thornton, P. W. Wilkens, D. T. Imamura, W. T. Bowen, and U. Singh. 1998. Decision support system for agrotechnology transfer: DSSAT v3. *Understanding Options for Agricultural Production*, eds. G. Y. Tsuji, G. Hoogenboom, and P. K. Thornton, 157-177. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- 3. Zeigler, B. P., H. Praehofer, and T. G. Kim. 2000. *Theory of Modeling and Simulation*. San Diego, CA: Academic Press.
- 4. Coléno, F. C., and M. Duru. 1999. A model to find and test decision rules for turnout date and grazing area allocation for a dairy cow system in spring. *Agricultural Systems* 61: 151-164.
- 5. Guerrin F. 2001. MAGMA: A simulation model to help manage animal wastes at the farm level. *Computers and Electronics in Agriculture* 33: 35-54.
- 6. Burden, R. L., and J. D. Faires. 1989. *Numerical Analysis*. Boston, MA: PWS-KENT Publishing Company.
- 7. Roberts, N., D. F. Andersen, R. M. Deal, M. S. Grant, and W. A. Shaffer. 1983. *Introduction to Computer Simulation: A System Dynamics Modeling Approach*. Boston, MA: Addison-Wesley.
- 8. Cros, M.-J., M. Duru, F. Garcia, and R. Martin-Clouaire. 2001. Simulating rotational grazing management. *Environment International* 27: 139-145.

- 9. Rellier, J.-P. 1992. Prediction and design problems in crop management. *Proc.* 12<sup>th</sup> International Conference on Artificial Intelligence, Expert Systems and Natural Language 2: 739-748.
- 10. Aubry, C., F. Papy, and A. Capillon. 1998. Modeling of decision-making processes for annual crop management. *Agricultural Systems* 56: 45-65.
- 11. Papy, F. 2000. Farm models and decision support: A summary review. *Research on Agricultural Systems: Accomplishments, Perspective and Issues*, eds. J.-P. Colin, and E. W. Crawford, 89-107. Huntington, NY: Nova Science Publishers.
- 12. Wooldbridge, M. J., and N. R. Jennings. 1995. Intelligent agents: Theory and practice. *Knowledge Engineering Review* 10(2): 115-152.
- 13. De Giacomo G., Y. Lespérance, and H. Levesque. 2000 Congolog, a concurrent programming language based on the situation calculus. *Artificial Intelligence* 121: 109-169.
- 14. Fox, M. S., and M. Gruninger. 1998. Enterprise modeling. *AI Magazine* (fall 98): 109-121.
- 15. Papy, F., J.-M. Attonaty, C. Laporte, and L.-G. Soler. 1988. Work organization simulation as a basis for farm management advice. *Agricultural Systems* 27: 295-314.
- Sherlock, R. A., and K. Bright. 1999. An object framework for farm system simulation. *Proc. MODSIM '99*, eds. L. Oxley, F. Scrimgeour, and A. Jakeman, 3: 783-788. The Modeling and Simulation Society of Australia and New Zealand Inc.
- 17. Martin-Clouaire, R., and J.-P. Rellier. 2000. Modeling needs in agricultural decision support systems. *Electronic Proc. of the XIVth CIGR Memorial World Congress*.
- Bergez, J.-E., P. Debaeke, J.-M. Deumier, B. Lacroix, D. Leenhardt, P. Leroy, and D. Wallach. 2001. MODERATO: An object-oriented decision tool for designing maize irrigation schedules. *Ecological Modelling* 137: 43-60.
- Jeannequin, B., R. Martin-Clouaire, M. Navarrete, and J.-P. Rellier. 2003. Modeling management strategies for greenhouse tomato production. *Proc. of CIOSTA-CIGR V Congress*, 506-513.
- 20. Martin-Clouaire R., and J.-P. Rellier. 2003. A conceptualization of farm management strategies. *Proc. of EFITA-03 Conference*, 719-726.
- 21. Cros, M.-J., M. Duru, F. Garcia, and R. Martin-Clouaire. 2004. Simulating management strategies: the rotational grazing example. *Agricultural Systems* 80: 23-42.
- 22. Boote, K. J., J. W. Jones, and G. Hoogenboom. 1998. Simulation of crop growth: CROPGRO model. *Agricultural Systems Modeling and Simulation*, eds. R. Peart, and R. Curry, 651-691. New York, NY: Marcel Dekker Inc.
- 23. Cros, M.-J., M. Duru, F. Garcia, and R. Martin-Clouaire. 2003. A biophysical dairy farm model to evaluate rotational grazing management strategies. *Agronomie* 23(2): 105-122.

- Richarson, C. W., and D. A. Wright. 1984. WGEN: A model for generating daily weather variables. U.S. Department of Agriculture, Agricultural Research Service, ARS-8.
- 25. Racsko, P., L. Szeidl, and M. Semenov. 1991. A serial approach to local stochastic weather models. *Ecological Modelling* 57: 27-41.
- Semenov, M. A., R. J. Brooks, E. M. Barrow, and C. W. Richardson. 1998. Comparison of the WGEN and LARS-WG stochastic weather generators in diverse climates. *Climatic Research* 10: 95-107.
- 27. Kennedy, J. O. 1986. *Dynamic Programming: Application to Agricultural and Natural Resources*. London, UK: Elsevier Applied Science.
- Epperson, J. E., J. E. Hook, and Y. Mustafa. 1993. Dynamic programming for improving irrigation scheduling strategies of maize. *Agricultural Systems* 42: 85-101.
- 29. Bergez, J.-E., M. Eigenraam, and F. Garcia. 2001. Comparison between dynamic programming and reinforcement learning: A case study on maize irrigation management. *Proc.* 3<sup>rd</sup> European Conference for Information Technology in Agriculture, Food and the Environment, ed. J. Steffe, (2): 343-348. France: Agro Montpellier.
- 30. Sutton, R. S., and A. G. Barto. 1998. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
- 31. Andradóttir, S. 1998. Simulation optimization. *Handbook of Simulation*. *Principles, Methodology, Advances, Applications, and Practice*, ed. J. Banks, 307-334. New York, NY: John Wiley and Sons, Inc.
- 32. Swisher, J. R., P. D. Hyden, S. H. Jacobson, and L. W. Schruben. 2000. A survey of simulation optimization techniques and procedures. *Proc. of the 2000 Winter Simulation Conference*, 119-128.
- 33. Fu, M. C. 2001. Simulation optimization. *Proc. of the 2001 Winter Simulation Conference*, 53-61.
- 34. Mayer, D. G., J. A. Belward, and K. Burrage. 1998. Optimizing simulation models of agricultural systems. *Annals of Operations Research* 82: 219-231.
- 35. Li, M., and R. S. Yest. 2000. Management-oriented modeling: Optimizing nitrogen management with artificial intelligence. *Agricultural Systems* 65: 1-27.
- 36. Mayer, D. G., J. A. Belward, and K. Burrage. 1996. Use of advanced techniques to optimize a multi-dimensional dairy model. *Agricultural Systems* 50: 239-253.
- 37. Mayer, D. G., J. A. Belward, and K. Burrage. 1998. Tabu search not an optimal choice for models of agricultural systems. *Agricultural Systems* 58: 243-251.
- 38. Mayer, D. G., J. A. Belward, and K. Burrage. 2001. Robust parameter settings of evolutionary algorithms for the optimisation of agricultural systems models. *Agricultural Systems* 69: 199-213.
- 39. Parsons, D. J. 1998. Optimizing silage harvesting plans in a grass and grazing simulation using the revised simplex method and a genetic algorithm. *Agricultural Systems* 56: 29-44.
- 40. Cros, M.-J., M. Duru, F. Garcia, and R. Martin-Clouaire. 2001. Simulation optimization of grazing management strategies. *Proc.* 3<sup>nd</sup> European Conference

- for Information Technology in Agriculture, Food and the Environment, ed. J. Steffe, (2): 349-354. France: Agro Montpellier.
- 41. Botes, J. H. F., D. J. Bosch, and L. K Oosthuizen. 1996. A simulation and optimization approach for evaluating irrigation information. *Agricultural Systems* 51: 165-183.
- 42. Sargent, R. G. 1999. Validation and verification of simulation models. *Proc. of the 1999 Winter Simulation Conference*, 39-48.
- 43. Weiss, G. 1999. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA: The MIT Press.
- 44. Bousquet, F., and C. Le Page. 2004. Multiagent simulations and ecosystem management: A review. *Ecological Modelling* 176: 313-332.

# 3.2 Control and Optimization

G. van Straten and L. G. van Willigenburg

**Abstract.** A wide range of various classical and modern control methods is classified from the point of view of required information and put in the context of optimization.

Keywords. Automation, Control, Optimization, Control classification.

### 3.2.1 Introduction

Control is a subset of automation in the wider sense. Automation aims at performing operations without constant human effort. This section provides an overview of control methods and algorithms, including dynamic optimization, in agricultural applications with a focus on continuous processes. There is another main stream in automation that deals with discrete operations like harvesting, packing, and internal transport. This field is characterized by the use of programmed sequences and PLCs in combination with suitable on/off sensors and simple actuators. Though important and widespread, this special field is not covered in this section.

There are two major fields of applications of control in the agricultural area: *mechanics and mechatronics* (power, implements, field operations, robotics) and *processing* (drying, storage, greenhouse cultivation).

Despite differences in time scales and targets, there are many similarities in approach. Therefore, in this section we concentrate on generic concepts. Our aim is to present an overview of the various controller paradigms, in order to help the reader in making the proper choice in practical situations. Section 4.1 Automation and Control of this handbook provides an overview of components and techniques of practical control systems.

# 3.2.2 The Purpose of Control

Before designing or purchasing any control system, it is important to ask what the main goal or objective of the controlled system should be. The goal is a very central

issue. Control can be defined as "to make the system behave as we desire." The desire expresses the objective. Possible control objectives are:

- To provide precision, accuracy and quality. Keeping things constant and repeatable is a major derived objective of this overall goal.
- To provide comfort and relieve people of monotonous work. This includes tasks that without control would be hard to do.
- To ensure safety, and to avoid risks. The control system is used to provide alarms and to stabilize a system within safety bounds.
- To prevent waste and abuse of valuable resources. The control system allows operations to be performed near the margins, thus saving energy, preventing quality loss, and making the best use of scarce resources. Such goals often ask for more than just set-point control: constraint satisfaction and optimization methods are important here.
- To enhance the economic result of the process, in the wide sense. Often, this entails a trade-off between purely economic factors and constraints due to environment and product quality. This may include savings on labor. It is obvious that optimization methods play a major role here.
- To design systems that without control would not be possible, e.g., precision farming. The control in these applications is not just an additional tool, but is an essential element of the system as a whole.

# 3.2.3 Systems, Signals, and Models

The first step in designing or choosing a controller for a system is to define the system boundaries. Once the system boundaries have been defined, it is necessary to specify how the system interacts with its environment. This process can be viewed as *modeling*. We need a model in order to know how the system responds to actions that we impose upon it, so that we can make the system behave as desired. In practice, sometimes successful control can be achieved without explicit models, by trial and error. Yet, modeling is a very central issue in the science-based design and development of controllers.

From the point of view of modeling for control purposes, physical streams such as energy and mass exchange can all be viewed as information streams or *signals*, i.e., entities that vary in time. A dominant approach in control theory is to distinguish between *input* signals and *output* signals. Inputs, or forcing variables, are exerted upon the system from the outside. Outputs represent the response of the system. Inputs can be further subdivided into unobserved disturbance inputs (*v*), observed disturbance inputs (*d*), and control inputs (*u*) (see Figure 1). Outputs can be subdivided into observed outputs (*y*) and outputs of interest (*q*). These are not observed, but can be calculated from known variables of the system. From the point of view of the system dynamics, the subdivision is not important, but from the point of view of controller design it is, because each control input is associated with an *actuator*, and each observed variable requires a *sensor*. In practice, actuators and sensors determine the cost of the control system, and are therefore subject to trade-off between costs and achievable performance.

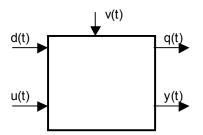


Figure 1. System with input and output signals.

### Linear and Non-Linear Systems

Systems that have the property that the output responds proportionally to the input are called *linear* systems. The output should be measured with respect to the system output at input zero. Most real life systems are non-linear, but many can be approximated by linear system models over a specific operating range. Control theory is much more developed for linear systems than for non-linear systems.

#### Models

There are two major approaches to obtaining a dynamic model for a system: *data-oriented modeling* and *mechanistic modeling*. The data-oriented approach starts with available system data, and tries to identify the dynamic behavior between input and output from the data. Such signal-driven models have the advantage that they can be derived from available data sets, and therefore do not rely on any pre-set judgment about how the system works. There is extensive software available to extract models from data. This process is known as *system identification*. The result is an *input-output model* (I/O model, Figure 2, left), which may take the form of a higher order differential equation, or, in the linear case, of a transfer function. Signal-driven models are very useful in classical controller design, but they are hard to transfer to other similar situations, and they cannot be used when the system does not yet exist.

In contrast, mechanistic models are derived from mass and energy conservation laws, or from known relationships between accessible variables within the system. In this type of modeling, there are auxiliary variables called *state variables* that represent the system's memory (Figure 2, right). Knowing the current state, the past behavior

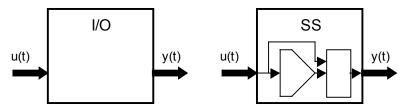


Figure 2. Input-output models (left) and state space models (right).

All signals and functions are vector valued.

Table 1. Standard form for continuous input-output (I/O) models and for state space models (SS). The non-linear and linear differential equation forms are given, as well as the transfer function form in the linear case.

	I/O	SS
Non-linear	$f\left\{y, \frac{dy}{dt}, \frac{d^2y}{dt^2}, \cdots\right\} = g\left\{u, \frac{du}{dt}, \frac{d^2u}{dt^2}, \cdots\right\}$	$\dot{x} = f\{x, u, t\}$ $y = g\{x, u, t\}$
Linear	$a_{n} \frac{d^{n} y}{dt^{n}} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \dots + a_{1} \frac{dy}{dt} + a_{o} y = $ $b_{q} \frac{d^{q} u}{dt^{q}} + b_{q-1} \frac{d^{q-1} u}{dt^{q-1}} + \dots + b_{1} \frac{du}{dt} + b_{o} u$	$\dot{x} = Ax + Bu$ $y = Cx + Du$
Linear transfer function	$G_{u \to y} = \frac{b_q s^q + b_{q-1} s^{q-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0}$	$\overline{y} = \left(C(sI - A)^{-I}B + D\right)\overline{u}$

becomes irrelevant. State variables can have a physical meaning, but they can also be virtual mathematical constructs. It should be noted that there may exist several state space representations that yield the same input-output behavior, showing that the state is not unique. The state space approach is very powerful, in particular in the frame of design and optimization.

Table 1 presents the common form of continuous models in I/O form and in state space form, both for the general, non-linear case and for the linear case. In this table u(t) is the input and y(t) is the output. The I/O equations are given for a single input and single output only, but the equations can easily be expanded to multiple inputs and multiple outputs. In the state space representation u(t) is an m-dimensional input and y(t) is a p-dimensional output, whereas x(t) is an n-dimensional state vector. For linear models, the transfer function form (TF) is also frequently used. The transfer function shows how the Laplace transformed output  $\overline{y}$  depends upon the Laplace transformed input  $\overline{u}$ . There are as many transfer functions as there are input-output combinations. The expression between brackets for the state space case is an m · p matrix. The linear I/O differential equation form, the transfer function form, and the linear state-space form can be transformed in one another. This means that controller designs based on one of these forms can also be made applicable to any of the other forms.

#### 3.2.4 PID Control

Before introducing a generic framework to control problems, it is appropriate to present briefly the well-known classical Proportional-Integral-Derivative (PID) controller. This controller uses a single input and a single output (SISO). The goal is to reject disturbances, and to track a reference trajectory r(t), which in many applications is just a constant set-point r. The controller is given by:

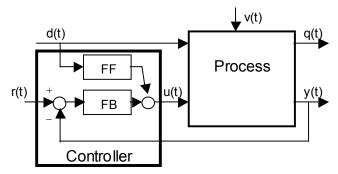


Figure 3. Signal flows for a system with feed-forward and feedback controller.

$$u(t) = c + K \left( \varepsilon(t) + \frac{1}{\tau_I} \int \varepsilon(t) dt + \tau_d \frac{d\varepsilon(t)}{dt} \right)$$
 (1)

The main idea is to check whether the output deviates from its reference value, and to correct the input u in proportion to the error  $\varepsilon(t) = r(t) - y(t)$ , using controller gain K. This idea of feedback (FB) is central to control. If the error is zero, the input is at its nominal value c. Because in practice due to load variations it can happen that the steady input needs to be adjusted to maintain the output, an integral action is added to achieve this automatically. The derivative action is introduced to speed up the controller response to fast disturbances.

Figure 3 shows the signal flow for this type of controller. The scheme also shows the option of feed-forward (FF) compensation of measured disturbances.

#### 3.2.5 A Generic Framework

### General System Model Description

A generic model description in state space form is:

$$\dot{x}(t) = f\{x(t), u(t), d(t)\}, \qquad x(t_o) = x_o$$
 (2a)

$$\begin{bmatrix} y(t) \\ q(t) \end{bmatrix} = \begin{bmatrix} g_y \{ x(t), u(t), d(t) \} \\ g_q \{ x(t), u(t), d(t) \} \end{bmatrix}$$
 (2b)

Here, x is the state vector, with an initial value  $x_o$  at the initial time  $t_o$ . The time evolution of the state is given by differential equations (Equation 2). The vector valued function f represents the rate of change, as a function of the state itself and the external forcing variables. These input signals are separated into manipulated variables u(t), and disturbance variables d(t). From the state output signals can be computed. The outputs are separated into observed outputs y(t) and variables of interest ("quality" variables) q(t) that either cannot be measured on-line or can only be computed.

The functions  $g_y$  and  $g_q$  are vector valued algebraic read-out functions. In practice, sometimes some of the variables of interest are computed by integration of states or outputs, for instance total energy consumption in a greenhouse model. In the frame-

work above, such integrated values are handled as states. By proper spatial discretization, the state space description can also accommodate spatially distributed systems.

Prior to setting up this scheme, the designer has to make decisions about the choice of suitable control signals and output signals. This choice depends upon the availability and cost of actuators and sensors in relation to their expected effectiveness. Making this choice is a very important step in control system design, and often requires an iterative procedure.

## Ideal Control over a System

Provided we have a model without error, including full knowledge of future disturbances, and a suitable goal function J(u, d) to be optimized, then it is possible—in theory—to compute a control trajectory  $u^*(t)$ , such that J is minimal. Also, the optimal paths of the states  $x^*(t)$  and the signals  $y^*(t)$  and  $q^*(t)$  will become available. Note that the answer depends upon the future disturbances, so that a prediction is needed for d. The goal function can be formulated as a *cost function* in which case its minimum must be found, or as a *benefit function* that needs to be maximized. Since one can be converted into the other by simply changing the sign, we will use the terms *cost minimization* and *goal optimization* as equivalents.

# A Typical Goal Function

A generic form of a goal function is:

$$J = \Phi\left\{x(t_f)\right\} + \int_{t_f}^{t_f} L\left\{x(\tau), u(\tau), d(\tau), q(\tau), \tau\right\} d\tau \tag{3}$$

Here  $\Phi$  represents the costs associated to the final state  $x(t_f)$ . Usually these are benefits, such as crop at harvest time, or biomass value at completion of the cultivation in batch reactors. L represents the running costs. An example is the heating in a greenhouse, where the costs relate to the heating power u(t) and the temperature x(t). The final time  $t_f$  can be fixed, free, or infinite.

## Dynamic Optimization

The task is to find the optimal control trajectory  $u^*(t)$ , such that the cost J is minimized, while satisfying the model equations. In mathematical terms:

$$u^*(t) = \arg\min J \tag{4}$$

subject to Equations 1 and 2. In addition, there may be constraints on x, u, and y.

There is software available to solve these kinds of problems, both for continuous systems as presented here, as well as for discrete time and sampled data systems. The solutions differ depending upon the nature of the final time and the presence of state or control constraints [1].

## **Open Loop Control**

Under the assumptions made above the calculated optimal control  $u^*(t)$  can be supplied to the system. If the model is correct, the system will behave in an optimal way, such that  $x(t) = x^*(t)$ ,  $y(t) = y^*(t)$ ,  $q(t) = q^*(t)$ , and the goal function  $J = J^*$ . This is

called *open loop optimal control*. An example is the steering of a robot to perform a time-optimal movement under fixed load conditions. Another example is the optimal feed rate control to a fed-batch bioreactor, so as to optimize biomass yield.

# The Need for Feedback

In real life, the picture outlined above is spoiled for several reasons:

- modeling errors,
- initial state deviations,
- deviations from the predicted disturbances, and
- unpredictable and unknown disturbances.

Deviations from predicted disturbances can also be treated as modelling errors, because disturbance prediction can be viewed as a special case of modeling. Modeling errors arise, for instance, from neglect of sub-processes, aggregation of variables, and erroneous parameters.

Because of uncertainties, the system behavior will deviate from the ideal behavior  $x^*(t)$ . The answer to uncertainty is *feedback*. The system is observed, via the observable outputs y(t), and if deviations occur from the expected trajectory, a correction is made. So, the scheme is enhanced with a controller, which can be viewed as a device—possibly in the form of an algorithm implemented on the computer—that maps the observed output y(t), and/or the systems state x(t) derived from it, into a control input u(t), such that the goal function J is optimized. The controller may also use the observed disturbance inputs d(t). Finding such maps is called *feedback controller design*. In order to be effective, feedback needs to be computed on-line.

#### 3.2.6 Feedback Control Families

# Optimal Control with Linear Feedback Compensator

The addition of a feedback compensator in addition to optimization leads to the scheme of Figure 4a [2]. The dynamic optimization is performed off-line. The compensator corrects the pre-calculated optimal controls with a correction signal, on the basis of deviations between predicted output and observed output. If the disturbance deviations are small, the compensator can be designed as a linear quadratic (LQ) controller. The underlying model is a locally linearized transfer function or state space model of the system. The goal function is the following quadratic expression:

$$J = \int_{t_0}^{t_f} \left( \delta u(\tau)^T Q(\tau) \delta u(\tau) + \delta x(\tau)^T R(\tau) \delta x \right) d\tau$$
 (5)

where  $\delta u$  and  $\delta x$  are deviations from the nominal optimal trajectory, and Q and R are weighting matrices, that allow the designer to balance between tracking properties and control effort. It can be shown that for normally distributed independent stochastic modeling and measurement errors the closed loop optimal control law is a state feedback law of the form:

$$\delta u(t) = -F(t)\delta x(t) \tag{6}$$

where F follows from the matrices Q and R [2][3].

Application of control law (Equation 6) requires that the state can be observed. The system state is often not accessible directly. In that case, *state reconstruction* is needed. This can be done with a Kalman filter (LQG, linear quadratic Gaussian state reconstruction) or by an observer [4]. The principle idea is that the state  $\delta x$  in Equation 6 is replaced by an estimate, obtained as:

$$\frac{d\delta\hat{x}(t)}{dt} = A(t)\delta\hat{x}(t) + B(t)\delta u(t) + L(t)\delta y(t) \tag{7}$$

where  $\delta y$  is the difference between the optimal output  $y^*(t)$  and the observed output y(t). In case of the Kalman filter the gain L is computed from the assumed system and measurement noise variances. If such information is not available, the observer gain can be designed such that the state error vanishes in a prescribed way.

Note that the compensator scheme fails when the optimal trajectory hits control constraints.

# Feedback Control with a Supervisor (Decision Maker)

A very common scheme appears if the off-line dynamic optimization is replaced by a supervisory decision maker. This is shown in Figure 4b. The observed output is compared to some set-point or set-point trajectory, which is dictated by a decision

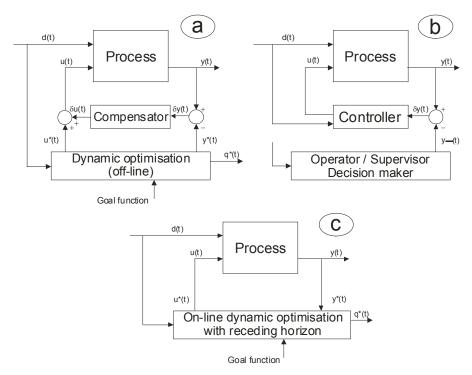


Figure 4. (a) Optimal control with linear feedback compensator, (b) standard feedback/feed-forward control, (c) receding horizon optimal control.

maker as supervisor. This can be a human operator, a scheduler (e.g., the program sequence in common greenhouse climate control computers), or the result of static or dynamic optimization.

Note that the overall benefits or costs of this scheme depend entirely upon the supervisor. This can be sub-optimal, no matter how well the controller has been designed. In many practical applications, sub-optimal control is quite satisfactory. The reason why the main stream of control applications belongs to this scheme is that responsibilities are split: the (economic) optimality is put into the hands of the supervisor, while the controller "only" needs to make sure that set-points are tracked and that disturbances are rejected. Any measured disturbance can be accommodated by feed-forward compensation in the controller.

A large range of controller designs is available, as outlined below, depending upon the number of controls and observations and the available knowledge about the system dynamics. The common classical single input-single output designs based on a linear transfer function description belong to this family. But, also, the multiple input-multiple output LQ-controller of Equation 6 based on the off-line optimization of Equation 5 can be used in this frame. In that case,  $\delta x$  represents the deviation from a predefined set-point, and  $\delta u$  is taken with respect to a pre-set nominal control value.

# Model Predictive Receding Horizon Control with On-line Dynamic Optimization

This scheme is presented in Figure 4c. Instead of off-line optimization in conjunction with a compensator, on-line optimization is performed, which directly calculates the controls needed in the forthcoming control interval, from an optimization of the type outlined before over a specific horizon. The length of the horizon must be sufficiently long as compared to the slowest time constant of the system. At the next interval, the optimization is repeated, and the final time is "receded" [5, 6]. Receding horizon schemes are particularly applicable when the disturbances are unpredictable but can be measured, and when there are control and state constraints. A typical example is greenhouse climate control. Because on-line optimizations must be carried out, this method can only be applied to systems that are slow enough, compared to the required computation time. With the ever-increasing speed of computers the class of potential applications is growing.

On-line optimization can be avoided under the assumption of linearity in conjunction with a quadratic goal function. In those situations closed loop solutions can be found, as for instance in *generalized predictive control (GPC)* [7]. It is clear that the original optimality might be lost in this way. However, such designs can still be used as the controller in the supervisory scheme of Figure 4, as an alternative to a classical design or LQ design. The advantage of these model predictive feedback controllers over classical feedback designs is that they can easily handle constraints.

# 3.2.7 Controller Paradigms and Design Methods

It is now possible to place the various control methods available in the literature in perspective. The choice of a method is dictated essentially by the available information about a problem. Important issues are:

- 1. Is the system already available, or should it still be designed? In the first case, data-based modeling methods can be used. System identification tools are available to estimate linear dynamic models from the data [8, 9]. If the system is supposed to be non-linear, neural-net modeling yields good results [10].
- 2. Is the model structure known, partially known, qualitatively known, or not known at all? If the model structure is known, state space methods are superior. Many controllers can be designed on the basis of transfer function information, which can be derived from the state space model (possibly after linearization) or from data [11, 12]. If only qualitative knowledge is available, fuzzy models may be used. If nothing is known at all, nothing remains but trial and error, probably best by starting with off-the-shelf PID control.
- 3. Are reliable model parameters available? If not, experiments for parameter estimation are needed, or adaptive methods may be selected that obtain the parameters "on the fly." Experimental design and identifiability are important issues here. The field of parameter estimation is related to regression, where methods are available to assess the confidence in the estimates [13].
- 4. If disturbances are measurable, they can be incorporated by feed-forward control. A model is needed. Feed-forward control is particularly useful for measurable load variations, in order to compute the necessary pre-compensation in advance [12].
- 5. The states of a system may not be fully accessible. In that case they need to be reconstructed by state estimation methods (e.g., observers, Kalman filter). For linear systems, observability conditions can be checked. Additional sensors may be needed.
- 6. For linear systems, a proposed scheme can be checked for controllability. A rearrangement of actuators and sensors may be required.

Table 2 summarizes the most important situations and associates to it the most suitable control methods. It is important to notice that the most important step in improving the performance of a controlled system is in improving the understanding of the system, i.e., in improving the model.

Tables 3 to 6 give a non-exhaustive summary of a number of well-known control methods. They are intended to help the reader in finding an appropriate method. Some features are explained below.

System Knowledge		<u></u>
Model Information	Additional Features	Preferred Method
Accurate		Optimal control $+ LQ(G)$
Improvable	While running	Adaptive control
Improvable	By repetition	Iterative learning control
Uncertain	With accurate state observations	Receding horizon optimal control (rhoc)
Uncertain	With some experience	PID / fuzzy control / rule-based control
Uncertain	With some idea on uncertainty bounds	Robust control
Uncertain		Improve model off line!

Table 2. The preferred control family as function of the available information.

			Frequency 1	Domain Methods	
Туре	PID	Cascade Control	Smith Predictor	FF Compensation	Robust
Model	TF form or response curve			TF /Bode with uncertainty bounds	
Typical problem area	SISO 2 <sup>nd</sup> order like	Slow/fast	Dead time	Measurable disturbances	Unknown variable dynamics
Synthesis	Tuning/pole-placement			Loop shaping	
Analysis	Phase & gain margin/root-locus			μ-synthesis	
Result	Tuned PID		Tuned PID + FF		Tuned FB law in TF format
Advantage	Off-the-shelf; pretty robust in practice				Works under parameter uncertainty
Drawback	Fixed structure limits performance		Fixed dead time	Good FF model needed	Tends to be conservative (worst case)
Remarks	MIMO: via decoupling	2 sensors/ actuators			More flexible controllers than PID
References		[1	1, 12]		[14]

Table 3. Frequency domain design methods.

Table 4. State space design methods.

		Linear State Space M	ath o da
Туре	LQ	LQG	Robust H <sub>inf</sub>
Model		Linear State Space	ce
Typical problem area	Accurate model + state observation	Accurate model + output observation	Uncertain model
Synthesis	Solve Riccati equation	Kalman filter design + Riccati equation	Minimize min-max goal (H-infinite norm)
Analysis	Tune Q,R matrices Tune Q,R + u		uncertainty parameters
Result	u = -Fx		u = -Fy
Advantage	MIMO / optimal for quadratic goals		MIMO
Drawback	Not robust		Conservative
Remarks	Not optimal when goal is not quadratic		Requires uncertainty information
References	[1-3]		[14]

### The Model

In many methods, the model plays a key role, yet the final controller may or may not explicitly contain the model itself. Sometimes the model is used only to find the controller parameters, such as in PID control. Sometimes the model is used to derive the controller form, such as in LQ and robust control, in conjunction with the controller parameters, but in all these cases the model itself is not part of the controller. In contrast, in optimizing control, including receding horizon model predictive control, and some adaptive control schemes, the model itself is embedded in the controller algorithm.

Table 5. Model predictive methods and adaptive control.

	Model Predic			re Methods			
Туре	GPC	IMC, DMC	RHOC	Adaptive Control			
	GFC		KIIOC	Self tuning	MRAC		
Model	CARIMA(X)	Step response	State space	(C)AR(I)MA(X)	State space		
Typical problem area	Data oriented problems with state and input constraints		Uncertain model + accurate state observations	Unknown or time varying parameters	Servo problems		
Synthesis	Optimizing quadratic goal: Diophantine equations		Optimization of goal	As GPC + recursive parameter estimation	Lyapunov theory + reference model		
Analysis	Simu	Simulation; stability analysis			Simulation; stability analysis		
Result	Explicit FB law: solution to Diophantine equations; horizon tuning		No closed form; on-line optimization	Autoregressive controller	FB + adjustment mechanism		
Advantage	Intuitively appealing		Optimal for any goal	Self-tuning and adapting to changing dynamics			
Drawback			Computationally intensive	Stability not easy to prove			
Remarks	Not optimal when goal is not quadratic		Can exploit meas- urable disturbances	Suitable for iterative learning			
References	[6, 7]		[2, 5, 22]	[15-17]			

Table 6. Non-linear state space design methods and artificial intelligence design methods.

	Non-Linear Star	te Space Methods	AI Methods		
Type	Optimal Control	Non-Linear Control	Neural Control	Fuzzy Control	
Model	Non-linear state space		Neural net (NARMAX)	None	
Typical problem area	Accurate general model; any goal function	Non-linear systems with wide operating range	Many data, no mechanistic model	Prior qualitative knowledge	
Design method	Hamiltonian & co-states or NL programming	Back stepping; feedback linearization	MPC but with neural model or direct mimicking desired behavior	Membership function tuning on available behavior	
Analysis	Sim	ılation	Simulation		
Result	State & control trajectories	Non-linear FB	Non-linear FB	Fuzzy FB rules	
Advantage	Truly optimal		No need for mechanistic model	Exploit human knowledge	
Drawback	Computationally demanding		Requires many data	Subject to prejudice	
Remarks	Usually in combination with compensator	Theory under development	Many variants; suitable for iterative learning	Many variants	
References	[1, 22, 23]	[17]	[10]	[18]	

# Frequency Domain Methods: Classical Control

These controllers are designed on the basis of linear *transfer functions*. The name *frequency domain* stems from the fact that the closed loop behavior of system plus controller can be studied in a Bode plot, i.e., a plot of amplitude amplification of the closed loop system as function of the frequency of the disturbance or set point. The task is to choose the controller parameters to obtain good tracking performance and good disturbance rejection while guaranteeing stability. Although valid for linear systems only, approximate transfer functions can be obtained for non-linear systems by linearization around a suitable working point.

#### Robust Control

Stability of a feedback loop is the major concern in classical design. The results of the design procedure depend upon the quality of the model. Since there are usually uncertainties, in classical design phase and gain margins are used to accommodate model mismatch. Another approach is *robust control* where the uncertainties about the model and parameters are expressed explicitly. In the frequency domain, the closed loop response is shaped such that the closed loop system will always be within the uncertainty bounds. Robust-designed controllers tend to be conservative, since they need to allow for the worst case [14]. Robust design is related to  $H_{\infty}$ -design in state space.

# Adaptive Control

The idea of *adaptive control* is that controller performance can be improved by updating either the controller or the underlying model parameters on-line. As such, it is an alternative to robust control. Since the system is adjusted steadily to new situations, rather than handling the worst-case scenario, it is less conservative than robust control. However, the identification step needed to obtain information about the system requires perturbing (sometimes called "probing") the system, which is incompatible in most cases with the desire to control the system. So, stability of adaptive schemes is a major issue. Adaptive control can also be used to design self-tuning or self-learning controllers [15]. Moreover, when the system is repetitive, such as batch processes or repetitive mechanical operations, each run can be used to improve the model. Here, iterative learning techniques are very useful.

# Model Predictive and Adaptive Control

The idea is very simple: the model is used to predict future behavior. Since this behavior is a function of the control action, the action that maximizes the performance can be searched. It has already been discussed. Adaptive control is achieved when parameters are obtained on-line [16]. Sufficient excitation is needed to be able to adjust the parameters. All adaptive schemes result in nonlinear controllers, and can therefore be seen as a subset of nonlinear controller design in general [17].

## Artificial Intelligence Methods

The term *artificial intelligence* or AI is used to indicate modeling methods such as neural net modeling and fuzzy modeling. Neural nets and fuzzy models [18] can be

seen as non-linear mappings. It should be noted that neural nets require lots of measurement data to train the network, whereas the formulation of fuzzy rules requires good qualitative knowledge about the system. In the control context, dynamic models are needed, which can be achieved in two ways. In one approach the state equations are as usual, but the right-hand side is approximated by a static neural net or fuzzy model. In the other approach the dynamics are approximated by discretization in time, and a mapping is sought between the output and past values of output and input. Apart from these model-based approaches, it is also possible to have neural or fuzzy controllers. In these cases the controllers are tuned to obtain the desired behavior or to mimic the control strategies of a skilled operator.

### 3.2.8 Methods for Dynamic Optimization

In the previous paragraphs optimality has been chosen as the fundamental idea. The question remains how optimal patterns can be computed.

In what are called *direct methods*, the (discretized) control sequence is obtained by common optimization techniques. Within this class, one group of methods is formed by gradient algorithms using Newton's method and variants. They are relatively fast, but they require the gradients and may get stuck in a local minimum. To overcome this, other algorithms have been developed, such as *controlled random search* [19] or *evolutionary algorithms*, and in particular the relatively efficient *differential evolution method* [20]. They are usually slower, but are better in locating a global optimum. The second group of direct methods tries to find the optimal path by variants of dynamic programming, which is basically a planning method. A fairly efficient method is *iterative dynamic programming* [21].

Indirect methods exploit the dynamic nature of the problem. They work with adjoint variables and a Hamiltonian function. The conditions for a minimum are derived from what are called the *Euler-Lagrange equations* [22]. An additional advantage is that these methods provide insight in the sensitivity of the solutions. Depending upon the various control, state and end constraints, there are many variants of this algorithm (see, e.g., [23]).

## References

- 1. Bryson, A. E., Jr. 1999. *Dynamic Optimization*. Menlo Park, CA: Addison-Wesley.
- 2. Athans, M. 1971. The role and use of the stochastic linear-quadratic-gaussian problem in control system design. *IEEE Transactions on Automatic Control*, AC-16(6): 529-552.
- 3. Kwakernaak H., and R. Sivan. 1972. *Linear Optimal Control Systems*. New York, NY: Wiley.
- 4. Lewis, F. L. 1986. Optimal Estimation. New York, NY: Wiley Interscience.
- 5. Mayne, D. Q., and H. Michalska. 1990. Receding horizon control of nonlinear systems. *IEEE Transactions on Automatic Control* 35(7): 814-824.

- 6. García, C. E., D. M. Prett, and M. Morari. 1989. Model predictive control: Theory and practice—A survey. *Automatica* 25(3): 335-348.
- 7. Clarke, D. W., C. Mohtadi, and P. S. Tuffs. 1987. Generalized predictive control —Parts I and II. *Automatica* 23(2): 137-160.
- 8. Norton, J. P. 1986. An Introduction to Identification. London, UK: Academic Press.
- 9. Ljung, L. 1987. System Identification—Theory for the User. Englewood Cliffs, NJ: Prentice-Hall.
- 10. Sjöberg, J., Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, H. Hjalmarsson, and A. Juditsky. 1995. Nonlinear black-box modeling in system identification: A unified overview. *Automatica* 31(12): 1691-1724.
- 11. Nise, N. S. 1995. *Control Systems Engineering*. Redwood City, CA: Benjamin/Cummings.
- 12. Stephanopoulos, G. 1984. *Chemical Process Control*. London, UK: Prentice-Hall International.
- 13. Montgomery, D. C, E. A. Peck, and G. G. Vining. 2001. *Introduction to Linear Regression Analysis*. New York, NY: Wiley Interscience.
- 14. Morari, M., and E. Zafiriou. 1989. *Robust Process Control*. London, UK: Prentice Hall.
- 15. Åström, K. J., and B. Wittenmark. 1995. *Adaptive Control*. Reading, MA: Addison-Wesley.
- 16. Martín Sánchez, J. M., and J. Rodellar. 1996. *Adaptive Predictive Control*. London, UK: Prentice Hall.
- 17. Krstíc, M, I. Kanellakopoulos, and P. Kokotovíc. 1995. *Nonlinear and Adaptive Control Design*. New York, NY: Wiley-Interscience.
- 18. Tagaki, T., and M. Sugeno. 1985. Fuzzy identification of systems and its application to modeling and control. *IEEE Trans. Fuzzy Systems* 1(1): 7-31.
- 19. Price, W. L. 1983. Global optimization by controlled random search. *J. Optimization Theory and Applications* 40(3): 333-348.
- 20. Storn, R., and K. Price. 1997. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optimization* 11: 341-359.
- 21. Luus, R. 2000. *Iterative Dynamic Programming*. Boca Raton, FL: Chapman and Hall/CRC.
- 22. Stengel, R. F. 1994. *Optimal Control and Estimation*. New York, NY: Dover Publications.
- 23. Bryson, A. E., Jr., and Y. C. Ho. 1975. *Applied Optimal Control: Optimization, Estimation and Control.* New York, NY: Hemisphere.

# 3.3 Topics on Software Evolution

A. M. Saraiva, A. R. Hirakawa, and C. E. Cugnasca

Abstract. Computer systems evolved beyond what could be anticipated decades ago. Although hardware evolution was basic to that process, maybe the software evolution is the most visible part to the user. What is not visible to the user are, in many cases, the changes that occurred in software development processes, software paradigms, and technologies that support the industry now present in our daily life. To mention only the most visible part, at the basic level the operating systems have changed from command-oriented to graphical systems. Computers or microprocessorbased systems have also invaded different objects from microwave ovens to agricultural tractors. Specific operating systems had to be developed for them, as well as for other novel handheld devices as cellular phones, palmtops, and games. Software cost became more critical compared to hardware. The systems that are now being developed are more complex. Consequently, new approaches, methodologies, paradigms, languages, technologies, and tools for software development have been required. The software industry is now a multi-billion-dollar market shared by big companies and free software communities. Entire books cover those subjects so only a brief overview of some selected aspects is presented here.

**Keywords.** Object-orientation, Open software, Operating system, Software modeling, Visual programming tools, UML, XML.

### 3.3.1 Introduction

The huge evolution of computer systems in the last decades has two equally important components: hardware and software. The software development process has received considerable attention because, compared to hardware, software costs have increased significantly. Besides cost, software quality and its user-friendliness, to mention a few aspects, have demanded a more professional approach. A few points related to these issues are presented in this section, which covers operating systems, software languages, object-oriented software modeling and its Unified Modeling Language, Extensible Markup Language for data interchange, and open software concepts.

# 3.3.2 Operating Systems

Each computer needs some type of system software developed to make it more convenient to use. This system, called its *operating system* (OS), is responsible for controlling the hardware and software resources of a computer, allowing the application programs to handle them properly to perform their specific tasks. The user's program is not allowed to directly control the systems resources (processor, memory, and peripherals); instead, multiple programs must share standard software drivers incorporated in the operating system. Due to that, OSs can be developed with different features, according to the application requirements, e.g., some are designed to simplify

computer use while other OSs are complex to make it more efficient, e.g., support multitasking, multi-user, or use in real-time applications.

The evolution of the OSs has been very important for microcomputer success. The following items show the most important OSs used in microcomputers nowadays [1].

### **Desktop Computer OS**

The DOS (Disc Operating System) has been used for many years, since personal computers—PCs—were launched. Due to its command-oriented interface it is not intuitive and quite difficult for most users. In the 1970s the first graphical user interface (GUI) was invented at Xerox Corporation's Palo Alto Research Center. The WIMP or Windows, Icons, Menus, and Pointing device was one of the GUIs available at that time and it was called a WIMP GUI; nowadays they are all known simply by the acronym GUI. In a GUI, pictures and graphic symbols represent the commands, choices, or actions. Another common aspect of those interfaces is the desktop metaphor: the area on the display screen where icons are grouped is referred to as the desktop because the icons are intended to represent real objects on a real desktop. The desktop is divided into windows, and in each of them different programs can be executed or different files can be displayed.

Well-designed graphical user interfaces can free the user from learning complex command languages, although some expert users feel that they work more effectively with a command-driven interface. They started to become commercially important since 1983, firstly with Mac OS from Apple, and then with Microsoft Windows, which has become the most popular for IBM-compatible microcomputers.

The next items show some steps of the evolution and the main characteristics of the most popular OSs for desktop computers.

#### Mac OS

The Apple Mac OS was developed to be simple and intuitive and has fewer compatibility and configuration problems. Many consumers consider Mac OS much better than Windows because they find it easier to use, more attractive, and more productive, especially when it comes to multimedia editing; it is also less subject to viruses and other malicious code attacks. Nevertheless it is only used in less than ten percent of computers. In spite of the fact that Mac OS has been continually improved its core has not changed substantially since its first version. However, it has limited multitask capabilities and presents difficulties in running multiple large applications at the same time.

An important update has been created, called Mac OS X, with many new features and enhancements. Its kernel, called Darwin, is based on BSD UNIX, the version of UNIX developed at the University of California, Berkeley. Unlike Microsoft, Apple has decided to adopt the idea of open source code so that it can encourage the community to submit modifications and enhancements, and it will also be easier to customize Mac OS X to meet specific needs. The main features of the Mac OS X are a virtual memory manager, no interference between application programs, a robust multitasking, availability of most of the services the users need, and significant Internet resources. Its appearance is quite different providing a more photo-realistic look to the

desktop. A lot of other interesting features are provided, although an important dilemma must be solved: few products based on Mac are available, usually with high performance but high cost and so only a small community uses Mac, compared to other OSs. The attempts to increase its market share have not been successful; maybe the release of the Mac Mini, a fairly powerful computer with a low price, can help change that picture.

#### Windows

The first versions of Windows were indeed an operating environment, requiring MS-DOS to be loaded on first. Later MS-DOS functions were incorporated to Windows.

Although more than 90% of PCs run on a Windows version, it has frequently received criticism, mainly due to its performance and faults. Compatibility with MS-DOS and the desire to gain market quickly are the main reasons of the problems. The Windows 95 version starts to solve these problems: its *Plug and Play* feature, although not always perfect, is an important evolution to usability of Windows PCs. But one of Windows' main advantages is the huge number of programs developed for it.

Other specialized versions of Windows that are available now include:

- Windows ME (Millennium Edition)—This is a consumer-oriented OS, designed to perform the wide variety of functions that people expect from PCs, such as home and office applications, simple networking and connectivity to the Internet, playing games, and multimedia applications (listening to MP3s, editing movies, etc). Based on the 9x kernel, an essential OS part that must be efficient, ME probably will be the last version to support MS-DOS programs. ME's more interesting new features are the System Restore tool, to return to a previous state after an important problem, the Auto Update, to check for critical OS updates and fixes (and download them) whenever the PC is connected to the Internet, and facilities for home networking and multimedia.
- Windows NT and Windows 2000—These products compete in the corporate server and workstation market that requires a reliable and secure environment. New Technology (NT) was designed as a 32-bit OS, unlike Windows 9x. Windows 2000 is an evolution of NT with the same reliability, but with the flexibility of a consumer OS. It is faster and more reliable than Windows 9x, reducing the chance that one software application interferes in another. Also, connecting to network resources is easier. Unlike NT, Windows 2000 offers full support for Plug and Play, supporting a large number of products (although a smaller number than ME).
- Windows XP—The intention in creating Windows XP was to provide a more stable and simplified computing environment by merging Windows 2000's stability with ME's versatility, and also simplifying software and hardware development. It is upgradeable from Windows 98/ME/2000. Two versions are available: Windows XP Professional for business and power users and Windows XP Home Edition for consumers. Besides ME features, many others were added: PC quick resume operation with low power consumption in standby mode, many

entertainment features and multimedia support, individual user accounts, protected memory space for each application, firewall protection, and remote access.

### Linux

Linux is a free OS, based on the old UNIX whose small kernel was written in assembly language, and the rest was written in C. Thus, it can be customized with relative simplicity. Since its introduction in the 1990s, Linux has gained more followers, mainly in the web server market, where Linux-based Apache has become a very popular Internet platform. It is a cost-effective OS: it is open-source and requires less system resources to operate, with quite good results in old microcomputers, an interesting and cheap solution for small group servers, or as workstations for simple uses.

Linux nevertheless has a command line interface, not suitable to most of the users. So, several GUI interfaces have been designed to simplify the Linux environment, such as Gnome and K Desktop Environment, which are similar to the Windows environment. Another point is that most of the software developed for Windows cannot be found in Linux versions. However, some companies have been investing in Linux products to the server market and other popular desktop applications. Many users consider Linux more difficult to install and use than Windows. Some companies such as Red Hat and VA Linux offer Linux versions that are easier to install. Although Microsoft products still dominate the market, Linux has grown consistently in the OS desktop market and also in the server market. Many important companies, IBM being an example, have invested strongly in Linux. Around the world some countries, states, and municipalities have stimulated the use of free software at the government level with a strong tendency to adopt Linux as their OS. This action may change the software scenario significantly in the coming years.

## Portable Computer OS

The portable computer market is recent but there are a lot of choices for consumers, including hardware and OS. Generically, they are known as *personal digital assistants* (*PDAs*) [2], and can be subdivided into categories based on the OS:

- Palm OS, from Palm Inc. [3], is a basic OS used in all Palm models, the Handspring Visor, and the Sony Clie. Its success is attributed mainly to its simplicity of use. No keyboard is required; it is quite fast and simple and can be easily customized. The Palm API is based on C language, simplifying development of new applications.
- Windows CE, from Microsoft [4], is based on Windows and was designed for handheld computing devices, mobile phones, and other applications. It allows Windows applications to be adapted to CE after some changes and recompilation. It can be used in a handheld PC, with a touch screen and a small keyboard, optional sound, and a CompactFlash and/or a PCMCIA slot; or in a pocket PC, without keyboard or mouse, optional sound, and a CompactFlash slot. Distinct from Palm OS, it depends on the user's familiarity with Windows and is a 32-bit, multithreaded, multi-tasking OS. This similarity is very important to new software development. This powerful solution allows processor-intensive appli-

cations, such as MP3 and video playback. The availability of popular applications such as Word, Excel, and Outlook is also very attractive to users. However, Windows CE requires more memory and the battery consumption is high. The new versions are similar to Windows XP, and have Internet Explorer, network access, MSN Messenger, a new look based on the Windows XP desktop.

Other OSs are available, such as Epoc in Psion PDAs, and recently the development of Linux-based PDAs has been encouraged: Agenda Computing, Samsung and Sharp are examples.

# Real-Time Operational Systems and Embedded Software

Many applications interact with environments that have time-varying properties and require predictable time-dependent responses. So the systems designed for these applications and their software must not only produce correctly calculated responses, but also exhibit predictable time-dependent behavior regardless of the system load and other conditions. They are called *Real-Time Systems* (RTS) [5, 6].

There are some important features that are desirable in RTS, e.g., time management, compliance with the tasks time requirement, time response predictability, and support peak load. RTS are frequently used in critical applications, such as industrial process, aircraft, and automotive, so fault tolerance is also a very important feature. They are not necessarily faster than a standard OS, but they must have a predictable time response, independently of whether or not the external stimuli are predictable. The processor's interrupt mechanisms can be used to assure predictability in many cases. The tasks scheduling must be carefully designed for RTOS, assigning the more suitable tasks to processors at each time. The pre-emptive priority scheduling should be supported, allowing a task to be interrupted at any point so that a more important task is able to gain the processor immediately. Table 1 has some examples of RTOS.

Nowadays there are many electronic products based on microprocessors (the *embedded system* or *ES*) such as cellular telephones, MP3 players, automobile devices, computer peripherals, and many others. *Embedded software* is used to control them, and it must be compact, efficient, reliable, and precise in treating inputs and outputs. With the advent of *systems-on-silicon*, RTOSs will be a component of ESs, expanding their applications even more.

RTOS	Supplier
pSOS Systems	Integrated Systems
VxWorks	Wind River Systems
VRTX	Microtec Research
OS-9	Microware Systems
SPOX	Spectron Microsystems
QNX software	QNX
HP-RT	Hewlett-Packard
LynxOS	Lynx Real-Time Systems
Inferno	Lucent
Virtouso RTOS	Eonic Systems

Table 1. Some commercial real-time operational systems.

## 3.3.3 Development Tools, Languages, and Platforms

Developing microprocessor and computer-based systems usually means developing software for which a programming language is selected and the system behavior is structured and implemented. Programming has changed a lot since the first computers were created. The original programs were very simple and straightforward compared to today's elaborate databases, word processors, schedulers, and action games [7].

Different computer languages have been created with which to write these increasingly complex computer programs. They can be categorized based on how close to normal speech they are, and thus how far from the computer internal language.

- *Machine language* or *1GL* (first-generation language) is the language of the CPU or central processing unit of the computer. The lowest-level language is composed of 0s and 1s.
- Assembly languages, or 2GL (second-generation languages), use mnemonics to represent the machine language, simplifying program understanding.
- *High-level languages*, or *3GL* (third-generation languages), use program statements—words and algebra-type expressions. After a program is written in one of the high-level languages, it must be either compiled (rewritten into machine language) or interpreted (translated). COBOL, FORTRAN, BASIC, PASCAL, PL/I, C, and JAVA are samples of 3GL.
- Very high-level languages, or 4GL (fourth-generation languages), are objectoriented and include database query languages. There are fewer options for programmers, but the programs are much easier to write than in lower-level languages. These, too, must be compiled or interpreted. The main example of this language is the SQL.
- *Natural languages*, or *5GL* (fifth-generation languages), are where the statements could be written as normal sentences. At present, there are no programming languages that use natural language.

There is another way to classify programming languages. A *visual programming language (VPL)* allows the user to specify a program in a two-(or more) dimensional way. Conventional textual languages are not considered two-dimensional since the compiler or interpreter processes them as one-dimensional streams of characters. Visual Basic, Visual C++, Delphy, Kylix and the other Visual family are not, despite their names, visual programming languages. A VPL allows programming with visual expressions: spatial arrangements of textual and graphical symbols. VPLs may be further classified, according to the type and extent of visual expression used, into icon-based languages, form-based languages, and diagram languages. The environments provide graphical or iconic elements, which can be manipulated by the user in an interactive way according to some specific spatial grammar for program construction. Some examples of visual programming languages are Prograph, Pict, Tinkertoy, Fabrik, CODE 2.0, ARK, VIPR, Cube and Hyperpascal.

The goal of all of these different approaches is the same: to make programming easier for programmers and accessible to non-specialized programmers. Some are used for rapid prototyping and rapid applications development, others are used for systems

or applications design, and still others can produce stand-alone applications for distribution.

Rapid application development (RAD) tools such as Visual Basic, Visual C, Delphi, Kylix and others are powerful languages that help develop software in a higher level of abstraction. They are based on textual languages, which use a graphical GUI builder to make programming interfaces easier. They are especially interesting for developing Windows interfaces on event-driven programs as they invoke fragments of code when the user performs certain operations on graphical objects on-screen. They have been widely used for in-house application program development and for prototyping. Prototypes can be powerful tools for eliciting requirements with the user, especially concerning man-machine interfaces and basic functionalities. When used at early phases of the development process, they can improve software quality and reduce the development costs; however, it is claimed that code maintainability is poor. In any case, they let users put more effort into solving their particular problems rather than learning about a programming language. Their success and the increasing complexity of applications, including web and distributed systems, promoted the development of new and more complete platforms such as .NET and J2.

#### .NET

.NET is a new programming model from Microsoft that incorporates almost everything related to the Windows environment and the facilities of the Internet [8]. It is claimed to have been developed from the ground up, based on a completely new framework, within which most programming tasks can be easily accomplished.

What could be done on Windows, including data access, windowing, connecting to the Internet, and much of the functionality of the Win32 API, is now mostly accessible through a very simple object model. The VB language has been widely upgraded, so it now includes classes and most of the features previously accessible in C++. A new language, C Sharp or C#, has been introduced, which combines the efficiency of C++ with some of the ease of development of VB. Memory management for .NET applications is much more sophisticated, meaning that a badly behaved .NET component is extremely unlikely to crash other components running in the same process. ASP.NET has replaced ASP (Active Server Pages). ASP is a specification for a dynamically created web page with the extension .ASP, which utilizes ActiveX scripting—usually VB Script or Jscript code. When a browser requests an ASP page, the web server generates a page with HTML code and sends it back to the browser. ASPs are thus similar to CGI scripts, but they allow Visual Basic programmers to work with familiar tools.

The new ASP.NET offers compiled web pages (making processing of web requests much more efficient) and includes a large number of pre-written components that can generate commonly used HTML form and user-interface items. The main programming languages have been moved far closer together, so code written in VB, C++, and C# can be intermixed. Components are wrapped up in a new unit called an assembly, which is highly self-describing, making installation and use of components very easy.

The most significant aspect of the .NET architecture is that code in VB and C# is compiled not to native executable code, but to an Intermediate Language (IL), with the

final step of converting to native executable normally happening at runtime. Such code is termed Managed Code C++. This makes your C++ code interoperable with VB and C# and allows you to take advantage of all the .NET features, but does restrict you to not using some features of C++ (such as multiple inheritance) that are not supported on .NET.

#### Java

Another popular programming language is Java and its new Platform Editions. Java is designed to solve a number of problems in modern programming processes. Java is a simple, object-oriented, network-ready, interpreted, robust, secure, architectureindependent, portable, high-performance, multithreaded and dynamic language. Java omits many rarely used, poorly understood, confusing structure features of languages such as C++. Another aspect of being simple is being small to enable the construction of software that can run stand-alone in small machines. Also, the Java interpreter and standard libraries have a small footprint, i.e., they require a small amount of disk or memory space. Java is object-oriented, which is very powerful because it facilitates the clean definition of interfaces and makes it possible to provide reusable software to dynamically link to C++ facilities. Java has an extensive library of routines for coping easily with TCP/IP protocols such as HTTP and FTP. This makes creating network connections easier and applications that open and access objects across the web via URLs in the same way as when accessing a local file system. With Java, the same version of one application runs on all platforms. The use of general bytecode instructions, which are not related to one specific computer architecture, makes the application portable. Surprisingly, Java has high performance. This is achieved by translating bytecode on the fly (at runtime) into machine code for the particular CPU the application is running on.

Recently, Sun Microsystems redefined the architecture for the Java platform, now named Java 2 [9]. Three products are part of the Java 2 Platform: Standard Edition (J2SE), Enterprise Edition (J2EE), and Micro Edition (J2ME). Each of these editions is composed of a Java virtual machine (JVM), Java programming language, technologies and features that are core to each product.

J2SE is based on JavaOne, optimized to run on individual desktops and workstations, which includes the Java Foundation Classes (JFC) API, Java plug-in software, international support, support for operability across heterogeneous environments, a 2D API, a new security model, and the Java HotSpot performance engine.

Building on the J2SE base, J2EE adds full support for Enterprise JavaBeans components, Java Servlets API, JavaServer Pages, and XML technology. The J2EE standard includes complete specifications to ensure portability of applications across the wide range of existing enterprise systems capable of supporting J2EE.

J2ME is a runtime environment optimized for very small and limited-memory devices, such as cellular phones, pagers, personal digital assistants, screenphones, digital kiosks, and automobile systems. J2ME's key component is the tiny-footprint K virtual machine (KVM). The most important thing about it is the connectivity of small devices with desktop and large enterprise systems.

#### 3.3.4 Object Orientation and UML: Unified Modeling Language

## **Object Orientation**

The traditional approach to software development is based on an algorithmic perspective, in which the building block is the procedure or function. Software specification is based on the functions it is supposed to execute, and the development process tends to focus on breaking each function into smaller ones to the point where they are easy to implement. The great problem of that approach is that functions are very prone to changes as system requirements change and as the system grows. As a result maintaining and evolving the system turns out to be very difficult.

The object-oriented approach proposes that the building blocks for software development be objects or classes of objects. An *object* is a thing that belongs to the problem domain, for instance a tractor, or to the solution domain, for instance a dialog box. The advantage of this approach is that, when modeling the problem, the objects are more stable than the functions or uses we have for them; in other words, the things that are important to the problem usually change less than the functions, that is, the way we want things to be manipulated. Object orientation tends to focus more on "what" than on "how," and that is more stable. Also, it allows an easier understanding of the problem and of its models even by non-software specialists, because it models the problem domain the way humans think, i.e. in terms of objects. As a result, it leads to less misunderstanding between software clients and developers.

A *class* is a description of a common set of objects that share some characteristics. For instance, a class "tractors" encompasses objects "tractors" (called instances) that share types of characteristics, though having different values for them (engine power, number of wheels, etc.). Objects have an *identity* (they can be distinguished from others, for instance because of their names), a state (data associated with it), and behavior (what can be done to the object, or what it can do to other objects). All of it is associated with the object and there is no separation between data and behavior, in the sense that they are encapsulated inside the object. Any object can only access the data and/or functions (behavior) of another object via its interface, and so it does not know about the other object's internal implementation. That is called *encapsulation* and is one of the key aspects of object-oriented software development. Two other aspects are inheritance and polymorphism. Inheritance allows a class to inherit characteristics of its parent class; for instance a class "wheel tractors" can inherit the properties of a parent class "tractors" (and can be more specific on other characteristics). Polymorphism is the characteristic that allows the same operation to generate different actions in different object classes; for example, an operation such as "print" can result in printing a map or a report depending on the object class. Object orientation also promotes maintainability of the systems, both because of the alleged stability of the objects and because the changes tend to be more restricted. They can even occur only inside the objects and therefore will not affect the rest of the system because of the property of encapsulation. Finally, object orientation provides the conceptual foundation for component-based software development, which is a key aspect for software reuse.

#### UML: A Little History

Object-oriented modeling languages date from the 1970s, in response to a need for different analysis and design methods to cope with new object-oriented programming languages and the increase in the complexity of software applications. By the mid-1990s, the number of object-oriented methods had reached more than 50, which confused the users, since none of them seemed to completely fill their needs. Three of the more prominent methods were Booch (from Grady Booch), Ivar Jacobson's OOSE (Object-Oriented Software Engineering), and James Rumbaugh's OMT (Object Modeling Technique). It became clear that, as they evolved, they were starting to adopt ideas from each other. These authors then began to develop the idea of a unified method and modeling language, which could mature more quickly thanks to the joint effort and accumulated experience, thus allowing the development of more powerful and useful software tools. Other partners, from big companies to individual collaborators, soon also supported the process and a few versions have been released since then. In the meantime, UML was adopted by the Object Management Group (OMG) as a standard in 1997 and a Revision Task Force is now responsible for its evolution [10]. Many books and software tools now available use UML, which is becoming a de facto standard for object-oriented modeling.

#### Key Aspects of UML

UML is a standard language for defining software structure. It can be used for visualization, specification, construction, and documentation of software-based systems. It is not tied to a specific development process and is only a part of a development method [10].

As a modeling language, its vocabulary and rules are intended for the representation of a system, focusing on different views of its architecture. It is a graphical language that helps facilitate understanding and communication between the parties. It allows specification in the sense that its models are precise, complete, and unambiguous, since the notation has a well-defined semantics. These models apply to different phases of the software development process. Although it is not a programming language, not even a visual programming language, it allows mapping from UML models into programming language constructs. The reverse is also possible. As a documentation language it provides a means for expressing system architecture and its details, its requirements, and tests. UML also supports project planning and release management.

There are three major elements that form a conceptual model of the language: the UML building blocks, the rules for combining the blocks, and the common mechanisms that are valid throughout UML.

There are three types of building blocks: things, relationships, and diagrams. *Things* are the main components in a model and are the basic object-oriented components of UML. They can represent static parts of a model (structural things), dynamic parts (behavioral things), organizational parts (grouping things), and explanatory parts (annotational things). *Relationships* describe the basic relations between things. There are four types of relationships in UML: dependency, association, generalization, and realization. *Diagrams* are graphic representations of sets of elements, usually things

and relationships that depict different views of a system. The diversity of diagrams (there are nine different types) is useful to stress different aspects of a system that otherwise would not be clear. The nine types of diagrams are: class, object, use case, sequence, collaboration, state chart, activity, component, and deployment diagram. UML rules define semantics for building the models in order to obtain well-formed models, which are self-consistent and harmonized with all related models. An example is the semantic rules for names (for things and relationships).

UML is a recent language but is increasingly used as it provides powerful, coherent and comprehensive concepts and notation for software modeling. In addition, many commercial tools support it and facilitate its use and integration with other software development tools. However, it is only a part (though an important part) of the software engineering process that must be used with a well-defined method. It is particularly adequate in a process that is use case-driven, architecture-based, iterative and incremental. Current information about UML and its formal specifications can be found at www.omg.org.

## 3.3.5 XML: The Universal Language for Data

XML or Extensible Markup Language can be summarized as a new tag-based language for describing data. It is a subset of the Standard Generalized Markup Language (SGML), a complex standard for describing document structure and content. It is a non-proprietary specification, supervised by the XML Working Group of the World Wide Web Consortium (W3C) [11], which was issued in 1998 as a recommendation, but that is still evolving with new functionalities and characteristics.

It is a metalanguage (a language for describing other languages), which means that it can be used to define customized markup languages for specific application domains (precision agriculture, for example) and their document classes. It is not a language for presenting data (such as HTML, Hypertext Markup Language), but for organizing data

#### SGML, HTML, and XML

SGML, an ISO standard (ISO 8879), was published in 1986 and introduced a format for embedding descriptive markup in documents and a method for the description of document structure. It allows the creation of hardware- and software-independent documents and supports a range of document structures. However, as it is very general, it is also complex, and it is difficult to implement processing programs with it.

HTML is an application of SGML that uses a very limited subset of tags that conform to a single SGML specification aimed at the presentation of data. Its simplicity led it to be the web-publishing language, but its fixed formats restrict its usefulness.

XML takes only the most important and simple features of SGML so it is more understandable and easier to use for application development, and it is more adequate for delivery and interoperability over the web. Compared to HTML, XML focuses on the content of the document, adding context and meaning to data. It allows the user to define the tags to be used so that data can be represented logically and in a structured way.

Its simplicity comes from its simple rules for creating a markup language to encapsulate data. The tags, which usually come in pairs, describe the information contained in the document, so it becomes almost self-describing. XML data is stored between tags as plain text, so it can be edited with any standard text editor. It supports the Unicode Standard, a character-encoding standard that supports all major languages, and that lets XML accept virtually all the world characters. This is especially interesting for developing applications that can be accessed by diverse cultures and nations. Its documents have a rooted tree structure, which is powerful to represent complex data for many applications, yet is easy to manipulate and to implement processing programs.

A document type definition (DTD) defines a grammar, or an XML document legal structure, specifying the tags that are available, where they may occur, and how to use them together. Although it is not compulsory to use a DTD it enables checking tags for validity and standardization on document naming and construction. An improvement to DTDs is the XML Schema language, which specifies the valid structure, constraints, and data types for the various elements and attributes of an XML document. Data typing is an important new feature, as it allows enforcement of proper syntax and semantics in XML documents, instead of treating all data as plain text.

In order to use XML documents an application program needs a parser, a software module that acts as an interface between them, reading the documents and giving the application program access to the document content and structure. Usually, there are parsers available in different languages and they are free.

#### Applications and Limitations

XML separates content definition from display instructions, so the same content can easily be used for different platforms or devices such as personal computers, personal digital assistants (PDAs), and cell phones, by using different style sheets (in Extensible Stylesheet Language, or XSL) to the same document.

Because XML documents also store meta-information (information about information) online information search and retrieval are easier.

One of the main points of XML is its potential for information exchange between different platforms and organizations. Since it is text-based all platforms can understand it very easily. It is claimed to be the universal language for data description. It also allows the definition of specific XML syntax between partners (for instance, agricultural organizations or companies) automating information transfer across the Internet.

However, XML is probably not a good choice for stand-alone systems. It does not provide any security features by itself, which can be a problem in a public network like the Internet unless some extra mechanism is used (cryptography, digital signatures, etc.).

Finally, standard vocabularies or tag sets are missing and must be developed, at least for specific industries (this is already occurring) to avoid misinterpretation of data. An example is the development of GML, or Geography Markup Language, which is an XML encoding for the transport and storage of geographic information,

including both the geometry and properties of geographic features (see http://www.opengeospatial.org/specs/).

XML is becoming an important language for many applications and will probably change the way information is used and delivered, especially on the Internet. It is still in progress and new characteristics are being incorporated. More information on XML can be found at http://www.w3.org/XML.

## 3.3.6 Open-Source Software

In a world where a few companies threaten to dominate software and the Internet, the strongest rival to their dominance is the collection of free software tools and operating systems collectively called *open-source software* (OSS). Unlike most commercial software, the core code of such software can be easily customized, modified, and improved.

Open source does not just mean access to the source code. The distribution terms of open source software must comply with the following criteria [12].

- 1. *Free redistribution*—The license shall not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programs from several different sources.
- 2. *Source code*—The program must include source code, and must allow distribution in source code as well as compiled form.
- 3. *Derived works*—The license must allow modifications and derived works, distributed under the same terms as the license of the original software.
- 4. *Integrity of the author's source code*—The license may restrict source code from being distributed only in original form. Path files, derived, or modified codes must be distributed with their own license.
- 5. *No discrimination against persons or groups*—The license must not discriminate against any person or group of persons.
- 6. *No discrimination against fields of endeavor*—The license must not restrict anyone from making use of the program in a specific field of endeavor.
- 7. *Distribution of license*—The rights attached to the program must apply to all of those to whom the program is redistributed without the need for execution of an additional license by those parties.
- 8. *The license must not be specific to a product*—The rights attached to the program must not depend on the program being part of a particular software distribution.
- 9. *The license must not restrict other software*—The license must not place restrictions on other software that is distributed along with the licensed software.

Nowadays, there are many products that can be considered OSS and many groups dedicated to its development [13]. The most popular one is the Linux Operating System and its related application software. Most UNIX-based operating systems observe the OSS features, which make UNIX sometimes be considered the father of OSS. The Internet is full of open-source software in heavy commercial use: Apache, which runs over half of the world's web servers; Perl, which is the engine of the World Wide

Web; and others such as FreeBSD, MySQL, Kylix, Java, and more recently, Open Office which is similar to MS Office.

Open-source software might seem to be a valorous fighter against world software monopoly. In 2002, the market share of Linux for web server OSs went from 29% to 34% while the market share for all other major OSs declined, including Windows, Sun Solaris, and other UNIX variants [14].

Some market researchers forecast that in the next few years, open source and free software will become the standard in operating systems, as well as in much of the commodity software in widespread use. It also means that one of two things may happen: either most computer users will switch to a new operating system and commodity software, or Microsoft will release Windows under an Open Source license. If people switch operating systems, it seems as though the most probable target is Linux, and other possibilities are FreeBSD or an Open Source release of BeOS. Regarding the second possibility, Microsoft might recognize the new trend and jump on it or it may have to do so because of the strong switch of users to OSSs. Most people see both scenarios as unlikely but there is no doubt that the OSSs are gaining popularity and market share.

#### 3.3.7 Conclusion

There is no doubt that software has evolved very much during the last decades, bringing more comprehensive and friendlier programs to a much wider public and applied to a wider range of applications, some very complex and demanding. However, there are many issues that must be further addressed, related to methodology, technology, compatibility, intellectual properties, and cost, among others.

Software became a very important industry and this is one of the main changes that occurred from the early days of programming. As a worldwide business involving huge resources it attracts many people and companies interested in that market. On the other side, a growing free software community interested on sharing knowledge and information proposes an alternative scenario to what may be the future of software.

#### References

- 1. Davis, S. J., J. MacCrisken, and K. M. Murphy. 1999. The evolution of the PC operating system: An economic analysis of software design. Accessed on March 03, 2005, from http://gsbwww.uchicago.edu/fac/steven.davis/research/.
- 2. Hansmann, U., L. Merk, M. S. Niklous, and T. Stober. 2003. *Pervasive Computing*, 2<sup>nd</sup> ed. New York, NY: Springer Professional Computing.
- 3. The Palm computing platform: An overview. Accessed on March 03, 2005, from http://www.wirelessdevnet.com/channels/pda/training/palmoverview.html.
- 4. Microsoft Windows CE: An overview. Accessed on March 03, 2005, from http://www.wirelessdevnet.com/channels/pda/training/winceoverview.html.
- 5. Li, Y., M. Potkonjak, and W. Wolf. 1997. Real-time operating systems for embedded computing. *International Conference on Computer Design (ICCD '97)*.

- 6. Rammig, F. J. 2001. Real time operating systems: Tutorial. *XIV Symposium on Integrated Circuits and System Design (SBCCI 2001)*.
- 7. Fenton, N., and S. L. Pfleeger. 1994. Science and substance: A challenge to software engineers. *IEEE Transactions on Software Engineering* (July): 86-95.
- 8. Robbins, T. 2000. Introduction to Microsoft.NET. Microsoft Partner Technical Workshop.
- 9. J2SETM Platform, J2EETM Platform, J2METM Platform documentation. 2002. Sun Microsystems, Inc. Accessed on March 03, 2005, from http://www.sun.com/java/about/.
- 10. Booch, G., J. Rumbaugh, and I. Jacobson. 1999. The Unified Modeling Language User Guide. Reading, MA: Addison Wesley.
- 11. W3C—World Wide Web Consortium. Extensible Markup Language (XML). Accessed on March 03, 2005, from http://www.w3.org/XML.
- 12. Perens, B. 1997. The Debian free software guidelines. Debian developers manual.
- 13. The Free Software Foundation. Accessed on March 03, 2005. http://www.fsf. org/.
- 14. Wheeler, D. A. 2002. Why open source software/free software (OSS/FS)? Look at the numbers! Accessed on March 03, 2005, from http://www.dwheeler.com/oss fs why.html.

# 3.4 Artificial Intelligence Methodologies

C. Borgelt and R. Kruse

Abstract. Artificial Intelligence is concerned with imitating the intelligent behavior of natural systems, such as animals and human beings, with artifacts such as computers and robots. It involves understanding how knowledge—especially uncertain and vague knowledge—can be represented so that it can be stored in computer memory and inferences can be drawn from it automatically, how decisions can be made and plans of action can be devised based on the stored knowledge, and how computer-processable knowledge can be acquired by interrogating human experts or by learning from example data.

Clearly, this section can only highlight some core methodologies and touch upon a few basic approaches. An interested reader is referred to the more-detailed expositions that can be found, for instance, in [1,2]. Textbooks for specific areas of artificial intelligence (formal logic, fuzzy systems, artificial neural networks, etc.) are pointed out in the corresponding sections.

**Keywords.** Artificial intelligence, Knowledge representation, Reasoning, Planning, Learning.

#### 3.4.1 Knowledge Representation and Reasoning

Although it is fairly difficult if not impossible to define *intelligence*, it is generally accepted that the ability to argue and to draw inferences is two of its key features. Therefore, if computers and robots are to show intelligent behavior, we have to provide them with the means needed to do reasoning, i.e., with appropriately encoded

knowledge and procedures to draw conclusions from this knowledge. In doing so, we have to place special emphasis on uncertainty and vagueness and methods to handle these phenomena, because most of our knowledge is neither certain nor precise.

#### Formal Logic

If we want to convey information to someone else or if we want to carefully analyze the arguments for and against a decision we have to make, we need to express the information or the arguments in a language. For this purpose, the language used must exhibit a certain structure. Of course, all natural languages, like English, French, German, etc., possess the necessary structure. However, natural language is most often not precise enough for computer representations, because in human communication the context in which some statement is made as well as common knowledge is implicitly drawn upon to fix the interpretation of certain words. In addition, natural language is very flexible, and allows for expressing the same thing in many different ways. Although this is surely an advantage in human communication, it turns out to be a hindrance for representing knowledge in a computer, because it can make it very difficult to check automatically whether two statements say the same thing.

Therefore, in artificial intelligence, formalized languages are used, in which the meaning of each term is precisely defined and which only exhibit the core structure needed for reasoning. These special languages are studied in the area of (formal) logic: Logic describes the core structure of argumentative languages, i.e., of languages in which one can argue. Formal logic reduces the complexity and ambiguity of natural language to a level on which it becomes possible to manage knowledge in a computer.

There are different logical calculi, depending on how far the complexity reduction goes. The most basic and simple system is *propositional logic*. It describes how the truth value of combined statements depends on the truth values of the basic statements they consist of. The basic statements considered are simple propositions like "Hannibal is a dog." Although such propositions have a structure, this structure is neglected. They are represented by simple *truth variables*, which can take one of the values *true* or *false*. Truth variables may be combined with *logical connectives* such as *and*, *or*, *not*, *if* ...then ..., etc. Propositional logic states, for instance, that the statement "A or B" is false if and only if both A and B are false. This *truth functionality* allows us to draw simple inferences. If we know, for instance, that the combined statement "A or B" is true and then find out that A is false, we can infer that B must be true.

However, propositional logic is not powerful enough for most applications, mainly because it neglects the structure of the basic propositions. Therefore an extension of propositional logic, namely (first order) *predicate logic* is frequently drawn upon. Like propositional logic it models the truth functionality of logical connectives. In addition it captures (part of) the structure of the basic statements using *constants*, *variables*, *functions*, and *predicates*. Constants denote specific objects, basically like names in natural language. The constant "John Steinbeck," for instance, refers to a specific American writer. Variables also reference objects, but in an unspecific way in order to make universal or existence statements in connection with the so-called *quantifiers* "For all ..." and "There exists ...." Functions model indirect characterizations, i.e., the

reference to an object through other objects. For instance, we may refer to John Steinbeck also as "author('The Grapes of Wrath')," where "author" is a function that is applied to the constant "The Grapes of Wrath." Finally, predicates are used to assign properties to objects or to describe relations between objects. For instance, the fact that John Steinbeck wrote the novel "The Grapes of Wrath" in 1939 can be expressed as "wrote('John Steinbeck', 'The Grapes of Wrath', 1939)," where "wrote" is a predicate and its arguments are constants. Note that predicate expressions can be true or false and thus correspond to the truth variables of propositional logic.

With predicate calculus more general inferences become possible. For example, using the fact that all humans are mortal, expressed as "For all x: if x is a human, then x is mortal" we can infer from the fact that Socrates is a human, that Socrates must be mortal. That is, with predicate calculus we can reason conveniently about properties of and relations between objects.

In artificial intelligence predicate logic is used, for instance, in *automatic theorem provers*, which can reason about situations described by logical formulae and can answer questions about these situations. There are also *special programming languages* like PROLOG that are based on specific subsets of predicate logic and are very powerful tools for artificial intelligence applications.

Further extensions of logic capture even more of the features of natural languages: *temporal logic*, for example, makes it possible to reason about time and takes care of changes of truth values in time ("The sun is shining" may be true today, but false tomorrow). *Modal logic* allows us to label statements as *possible* or *necessary* and thus provides the means for a–though limited–treatment of uncertainty. However, discussing these extensions of standard (first order) predicate calculus is beyond the scope of this section. More details about different logical calculi and their application can be found, for instance, in [3-5].

#### Rule-Based Systems

Rules, i.e. if-then statements, are a very convenient representation of many forms of knowledge, mainly because they are generally regarded as easily understandable. Rules can be used to specify inferences that can be drawn or to associate actions with conditions under which they have to be carried out. Therefore, rules are the most popular building blocks of expert or decision support systems as well as knowledge-based controllers. They seem to provide one of the best ways to reconcile human thinking and the abstraction and formal precision needed for computer representations.

A rule-based system usually consists of a *knowledge base*, which contains the rules representing the knowledge about the application domain; an *inference engine*, which checks which rules can be applied and carries out inferences or triggers actions; and an *application interface*, through which a user can communicate with the system or through which sensor information is fed into the controller. A decision support system may also comprise an explanation component, which generates justifications for the suggestions made by the system. This is necessary, because the responsibility rests, of course, with the human decision maker, who would like to know the reasons for a decision in order to be able to make it conscientiously. In a knowledge-based controller

there may also be state variables as an additional component, which can be used to make the behavior of the controller dependent on the preceding control state, i.e., to program control strategies that can take care of, for instance, time lags.

#### Uncertain Knowledge

Human expert knowledge is rarely certain. Most rules are only generally true and have exceptions. The best-known example is, of course, "All birds can fly." Although most birds can fly, there are some that cannot, such as penguins and ostriches. So if we infer from "Tweety is a bird" that "Tweety can fly" we may be mistaken.

However, the fact that most knowledge is *uncertain* does not mean that it is *useless*. In medicine, for example, certain symptoms point often—but not always—towards certain diseases. Therefore a physician is frequently in a situation in which he cannot infer the disease with certainty, but nevertheless his diagnoses are most often correct.

Unfortunately, rules that are only generally true, but have exceptions, cannot be handled easily with pure logic-based approaches. The problem is that in classical logic a statement must be either true or false. Since a "generally true" statement is, in a strict sense, neither true nor false, it cannot be used without introducing consistency problems. With modal logic, which we mentioned above, it is possible to label "generally true" rules as usable as long as there is no explicit information to the contrary. However, it lacks sufficiently powerful means to express and infer preferences between conclusions drawn from only "generally true" knowledge.

The most natural approach to handle uncertain knowledge is to soften the notion of truth by introducing additional truth values between true and false. For instance, one may introduce "maybe" levels or one may use real numbers between 0 and 1, identifying 0 with false and 1 with true. Enhancing rule-based systems with certainty factors, which specify the reliability of rules, and combination rules for these factors, have been successful in specific applications like the diagnosis of bacteriogenic infections. Unfortunately, however, they have been shown to lead to inconsistent results in the general case. The main problems are implicit independence assumptions, which are hidden in this approach and which are not satisfied in many applications.

A more promising approach to handle uncertainty, which has gained considerable popularity in recent years, is Bayesian networks. The basic idea underlying Bayesian networks is to make the dependence and independence relations (which hold between the attributes used to describe the domain of interest) explicit, thus avoiding the problems that result from implicit independence assumptions. The dependence and independence relations are encoded with a help of a graph, in which each node represents an attribute and each edge a direct dependence between attributes. From this graph the valid independences can be obtained with simple graph theory criteria. The graph also specifies how the (probabilistic) knowledge about the domain can be decomposed. As a consequence it prescribes the paths of probabilistic inference, so that mathematically sound and efficient evidence propagation methods can be derived. However, the mathematics of this decomposition and inference process are much too involved to be discussed here. An interested reader can find a detailed treatment of Bayesian networks and related approaches in, for instance, [6, 7].

#### Vague Knowledge

Human expert knowledge is not only *uncertain*, but also often *vague*. While uncertainty means that we cannot decide which of a set of (precisely defined) possible alternative statements correctly describes the obtaining situation, usually because we are lacking information, vagueness refers to situations in which the meaning or the applicability of a statement is in doubt. Vagueness results from the fact that the words of natural language usually have no precisely bounded domain of application. Although there are some situations in which they are surely applicable and others in which they are definitely not, there is a *penumbra* in between, where their applicability is not clear. For instance, 35°C is surely hot weather, while 10°C is surely not. But what about 25°C? Obviously, there is no precise boundary (a specific temperature) that separates the temperatures to which the term "hot" is applicable from those to which it is not, so that a decision is somewhat arbitrary.

The idea underlying *fuzzy set theory* and *fuzzy logic* is to describe the domain of applicability of a linguistic term like "hot" not by a crisp set (which would presuppose a sharp boundary), but by a *fuzzy set*, which has a soft or "fuzzy" boundary. This soft boundary is brought about by allowing for gradual membership in a set, which is described by a real number between 0 and 1. Zero means that an element is not contained in the set, 1 that it is contained without restriction. With such a gradual membership, linguistic terms like "cool," "warm," or "hot" may be interpreted as shown in Figure 1. Temperatures between 0°C and 10°C are definitely cool and neither warm nor hot. The temperatures between 10°C and 20°C, however, may be described as cool or warm, but with different degrees of membership. The lower the temperature, the higher the degree of membership for "cool" and the higher the temperature, the higher the degree of membership for "warm." Of course, the exact shape of the membership functions depend on the application.

The main advantage of fuzzy logic and fuzzy set theory is that they allow us to express our knowledge in rules that use linguistic terms, which are interpreted very intuitively with membership functions. They also provide very convenient means to interpolate between such rules if more than one rule is applicable. This is the reason why fuzzy logic is very popular in knowledge-based control: it enables a human expert to construct a controller by specifying some key input/output relations through linguistic rules, between which the fuzzy logic inference engine then interpolates to complete the control function. Fuzzy logic controllers have had enormous commercial success and can be found today in many household appliances. An extensive treatment of fuzzy theory and its applications is in, for example, [8-10].

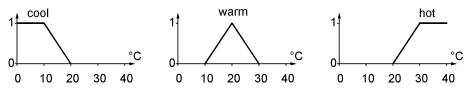


Figure 1. Fuzzy sets for the linguistic terms "cool," "warm," and "hot."

#### Knowledge Acquisition

Unfortunately, most knowledge needed in artificial intelligence systems is "buried" in human minds, and it has turned out to be very difficult to elucidate it in such a way that it can be fed into a computer. For instance, after an agricultural expert has looked at the landscape, has examined the soil, and has checked the weather in the region, he may be able to tell what crop the area is best suited for. But he may not be able to fully explain his reasoning, let alone state it as clear and simple rules. In other words, knowing something is not the same as teaching it—and we have to teach computers how to do certain things if we want them to help us.

As a consequence, knowledge acquisition is an important area of artificial intelligence. It is mainly concerned with the problem of how to get from a human expert the knowledge that is needed to automate processes and to provide computer-based decision support. As such, it is sometimes more a part of psychology than of computer science, because it involves interviewing human experts, setting up questionnaires, mapping the domain-specific terms and notions to computer-processable functions and predicates, etc. Since knowledge acquisition can be a very tedious and time-consuming process, recent research, supported by considerable advances in computing power, storage media, and sensor technology, has focused on learning from data, or, as we may say, by observing what the human expert does and trying to find automatically the rules that are needed to imitate him (see Section 3.4.3).

#### Case-Based Reasoning

A popular way to bypass the *knowledge acquisition bottleneck* is *case-based reasoning*. It draws on the idea that human beings often assess a new situation by comparing it to situations they (or their fellow humans) experienced in the past, and then act in a similar way as they acted in the most similar situations they have encountered or heard about. As a consequence, in a case-based reasoning system, the knowledge is embodied in a library of past cases, rather than being encoded in rules. Each entry in the library describes a case or a problem together with its outcome or its solution, respectively. Thus the knowledge or reasoning involved in deriving the outcome or solution is not made explicit, but is left implicit in the case descriptions. Therefore it is sometimes also called *lazy learning*, as all knowledge acquisition consists in simply storing example cases in a library.

Reasoning in case-based systems consists basically in retrieving those cases from the library that are most similar to the currently investigated case. It is then conjectured that the outcome of these cases or their solutions are also applicable to the new case. By trial applications of the solutions it is then determined whether they have to be revised or adapted in order to fit the current problem. Of course, after the outcome for the current case or the solution to the current problem has been found, it is added in a new case description to the library. A simple special form of cased-based reasoning is k-nearest neighbor classification. Based on a distance measure, which encodes the similarity of cases, the k closest examples in the library are retrieved and the class of the new case is predicted by a simple majority vote of these cases.

The advantage of case-based reasoning is that it is very intuitive as it works in a similar way as human beings reason about situations. Hence its suggestions are usually readily accepted by human decision makers. In addition, knowledge acquisition becomes fairly simple. Its disadvantages are that the library of example cases can become very large, which makes it difficult to devise methods for efficient look-up and retrieval of the most similar cases. In addition, the success of such methods depends heavily on the quality of the chosen similarity measure. More information about case-based reasoning and its applications can be found in [11, 12].

#### 3.4.2 Planning

An apt definition of *thinking*, on the simplest level, is that it consists in trial acting in an imaginary environment. Such trial acting serves the purpose of devising a plan for acting in the real world by exploring the likely consequences of possible courses of action. Its advantages are, of course, that actions that fail do not have harmful consequences. Or, as Konrad Lorenz remarked so pointedly about the evolutionary advantage of thinking: The hypothesis can die instead of the individual.

#### State Graphs

A planning problem is most conveniently represented as a *state graph*, in which each node represents a state of the domain of interest and each edge a possible direct transition between states that can be brought about by a specific action. Some states play a special role: the initial state, in which the domain is at the beginning, and the goal states (often there is only one), one of which is to be reached. The objective is to find a sequence of actions, called a *plan*, so that the state transitions brought about by this sequence of actions lead from the initial state to one of the goal states. In the state graph, this plan corresponds to *path*, i.e., a sequence of edges, from the node representing the initial state to a node representing a goal state.

As an illustrative example we consider the well-known problem of the man who carries a cabbage, a goat, and a wolf. He has to cross a river, but unfortunately the only boat that is available can hold only himself and one of the items he is carrying. He cannot leave the cabbage together with goat, nor the goat with the wolf, because in each case the latter will devour the former. The objective is to find a plan that will get him and his items safely to the other side of the river. The state graph for this problem is shown in Figure 2 with the initial state on the left and the goal state on the right. The letters near the states indicate which items are on this side of the river: m = man, c = cabbage, g = goat, w = wolf. Obviously, with this representation a solution can easily be found by searching a path from the initial state to the goal state.

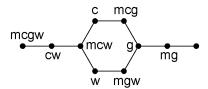


Figure 2. A state graph for the cabbage-goat-wolf problem. The initial state is on the left, the goal state on the right.

This is, of course, a toy example. State graphs for real world problems tend to be much larger, containing several hundred or even several thousand states. As a consequence sometimes it becomes impossible to construct the full state graph, as shown in this example. In such a case implicit representations of a state graph are used, consisting of an initial state, operations to construct the states that can be reached directly from a given state, and a function to identify goal states.

#### Search in State Graphs

State graphs are only representations of planning problems. They do not provide solutions immediately, as we still have to search for a path from the initial state to the goal state. In doing so we may have to take into account additional constraints; for example, each action may incur a cost and we may desire a minimum-cost solution.

Straightforward solutions to the path search problem are the *breadth-first search* or its more general version the *uniform-cost search*, as well as the *depth-first search*. In breadth-first and uniform-cost searches the states are visited in the order of increasing distance (measured as the number of edges) or increasing cost, respectively, from the initial state. In depth-first searches one always proceeds from the state that has been visited most recently, until a depth limit, specified as a number of edges, is reached. At this limit, or if no new states can be reached, the search backtracks to a state visited earlier, from which new states can be reached. The advantage of breadth-first (uniform-cost) search is that it always finds the solution that needs a minimum number of actions (causes minimum costs). However, it has the disadvantage that a lot of states have to be processed in parallel, so that a program proceeding in this way can consume a lot of computer memory. The advantage of the depth-first search is that it is much more memory-efficient, but it may not find the optimal solution if the search is terminated as soon as a goal state is reached. In addition, it may fail to find a solution if all goal states lie beyond the "search horizon" that is defined by the depth limit.

All these approaches are called *uninformed search*, because none of them uses special knowledge about the application domain to guide the search. In practice, however, it is often indispensable to exploit such domain-specific knowledge to guide the search, because otherwise the search would take prohibitively long. For instance, if we searched for the shortest route from town A to city B with a breadth-first search on the road network, treating each town or even each crossing as a state, the search explodes, except when A and B are very close together. Instead we use the useful heuristics to start the search by following only roads that emanate from A in the direction of city B or lead to the nearest highway. That is, we restrict the search to the most promising alternatives, considering other alternatives only after a failure.

A general method that makes use of such heuristics is the  $A^*$  algorithm. It is based on two functions g and h, which assess states. Function g measures the costs it takes to reach a state from the initial state. These costs are usually estimated from the best path that has been found up to now. Function h measures the costs needed to reach a goal state from a state. It is estimated by a problem-specific heuristic function. In a navigation problem like the one discussed above, we may simply use the straight-line distance between a town or crossing and city B. The  $A^*$  algorithm prescribes continuing

the search from the state for which the sum of the estimates of the two functions is minimal. This has the advantage that a path to the goal state can be found in considerably fewer steps, provided the heuristic estimate is appropriate. A nice property of the  $A^*$  algorithm is that it is guaranteed to find an optimal (i.e. minimum cost) solution, provided the function h never overestimates the costs.

More information about search in general and especially about heuristic search techniques can be found, for instance, in [13, 14].

#### Reasoning about States and Actions

More sophisticated approaches to planning use logic-based representations of states and how actions modify states and their properties. In this case the results of an action can be described in a generic way and it is left to the logical inference engine underlying the planning system to determine the outcomes of an action for a specific situation. Of course, this approach has the advantage of an immensely increased flexibility. However, it also leads to heavy computational costs, which can make it infeasible for some applications.

#### 3.4.3 Learning

One of the most striking abilities of all higher natural organisms is their ability to *learn*, that is, to modify their behavior as a result of past experience. Learning enables a (natural or artificial) agent to perform the same tasks in a more efficient way and to handle new tasks he could not manage before. Therefore, learning ability is commonly regarded as one of the cornerstones of intelligent behavior. In this section we highlight a few approaches to automated or *machine learning*. A more detailed treatment can be found, for instance, in [15, 16].

#### Inductive Logic Programming

Inductive logic programming consists of finding logical formulae that aptly describe the given data by systematically searching the space of all logical formulae that are applicable to the data. Most often the search is restricted to rules, i.e. if-then statements, because such rules are most convenient in many applications (see Section 3.4.1).

The search through the space of logical formulae is usually carried out by proceeding from the general to the specific (although there are also a few approaches which work the other way round). That is, the search is started with a very general formula, often an empty rule, which is applicable to all example cases. Thus it is made more specific, but hopefully without lowering the support of the rule too severely, i.e., the number of example cases to which it is applicable. The quality of each rule is then assessed with heuristic measures, which provide a ranking for the presentation of the rules found, but may also be used to guide the search. If the rules with the highest scores are the first to be expanded, chances are better that the best and most useful rules are found early in the search process. This has the advantage that the search can be stopped early, before the whole search space has been visited, without losing too much of the overall result.

Inductive logic programming is very well suited for classification and concept description tasks. In these cases the search starts from one rule for each class or concept, each having an empty antecedent and the class or concept in the consequent. Conditions are only added to the antecedent of the rule to make it more specific. The best rules found in the search can be used to classify cases for which the class is yet unknown or to understand what properties are characteristic for the concept. Applications are, for instance, tasks like assessing what properties characterize customers of a bank that are credit-worthy (in order to distinguish them from those who will presumably not pay back their loan) or the fault-prone vehicles of a car manufacturer (in order to find the causes and to improve the product quality).

The advantage of inductive logic programming is that it is not bound to single table representations of the training data, but can also handle multiple relations, which are connected by references and keys. This can be very important in applications where the data is often stored in relational database systems, divided into several relations that refer to each other. In such a case it can be difficult, if possible at all, to transform the data into tabular form, so that each line of a single table describes one example case. However, decision trees and artificial neural networks, which are described below, require such a single table representation.

The main disadvantage of inductive logic programming is that implementations are usually fairly slow, partly due to the fact that most of them are programmed in PROLOG, a language that is hardly renowned for fast execution. However, the core of the problem is rather that the search space that has to be traversed by inductive logic programming approaches is usually vast, so that prohibitively long execution times can result. Therefore, it is very important to restrict the search space and to guide the search by a *declarative bias*, which specifies the type of rules to look for by formal grammars or rule templates, and by applying search heuristics such as beam search, i.e., by concentrating in each step on the k best rules found in the preceding step.

More information about inductive logic programming approaches can be found, for instance, in [17, 18].

#### **Decision Trees**

The *decision tree* is the most popular machine learning method, because it is very simple and fast and produces easily understandable results. A decision tree is a classifier, i.e., a method to assign a class from a predefined set to a case or object under consideration, based on the values of a set of descriptive attributes.

As its name indicates, a decision tree has tree structure. Each inner node (i.e., each node having descendents) specifies a test of a descriptive attribute; each leaf (i.e., each node not having descendants) assigns a class. A case or an object is classified with a decision tree by starting at the root and descending along those branches, which correspond to the outcome of the test specified by an inner node, until a leaf is reached from which the predicted class can be read. As an illustration consider Figure 3, which shows a very simple decision tree for a medical task: a decision about a drug to administer. The tree prescribes to check the patient's blood pressure first. If the blood

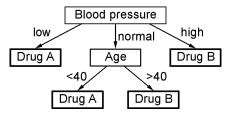


Figure 3. A simple decision tree which yields a suggestion for which drug to administer, depending on the blood pressure and the age of the patient.

pressure is low, drug A should be used; if it is high, drug B. If the blood pressure is normal, a second test is necessary. Depending on whether the patient is under or over forty years of age, it suggests administering drug A or drug B.

Decision trees can be constructed automatically from a set of preclassified example cases by a recursive procedure that is based on a "divide and conquer" approach together with a "greedy" selection of the test attribute. With this approach a decision tree is constructed from top to bottom (in the representation of Figure 3), which is why it is called *top-down induction of decision trees (TDIDT)*. It works as follows: The amount of information provided by each of the available descriptive attributes about the class is assessed with an evaluation measure. The attribute (or rather a specific test of this attribute) that receives the best score is selected and forms the root of the tree. The training cases are then split into subsets according to the outcome of the chosen test. The rest of the tree is constructed recursively by applying the same procedure to each of the subsets. The recursion stops if all cases in a subset are assigned to the same class, if there are too few cases for another split, or if there are no tests that could be performed left.

It should be noted that a decision tree can also be seen as a compact representation of a set of rules. Each path from the root to a leaf node describes one rule, with the conjunction of the test outcomes along the path providing the antecedent and the class assignment in the leaf providing the consequent of the rule. Therefore the induction of decision trees can also be used to find rules, although it should be kept in mind that the result may not contain the most expressive individual rules for a domain, due to the way in which they are constructed. The goal is rather to find a set of (related) rules that together yield a good classification of all training examples.

The advantage of decision trees is that learning is usually very fast and that the result is easily interpretable by human experts, so that it can be checked for plausibility. Their disadvantages are that the induction process usually selects only a small number of the available attributes so that information that is distributed on a large number of attributes, with each attribute carrying only limited information about the class, cannot be handled adequately, resulting in suboptimal prediction accuracy. In such situations (naive) Bayes classifiers and artificial neural networks are often superior.

A detailed treatment of the algorithms and mathematics underlying decision tree construction can be found, for instance, in [19, 20]. The latter also treats *regression trees*, which can predict numeric values instead of classes.

#### Artificial Neural Networks

Artificial neural networks are information processing systems whose structure and functionality are modeled after the nervous system, especially the brain, of humans and other animals. The basic idea is that intelligent behavior and especially learning ability may be brought about by imitating at least certain features of the "biological hardware" of natural intelligent systems. An artificial neural network consists of a (usually large) number of simple processing units, the so-called *neurons*, which work in parallel. These neurons process information by sending activation signals to each other via directed connections.

There are a large number of different types of artificial neural networks, which can be used for a large variety of tasks. Here we confine ourselves to the most common type, the *multilayer perceptron*, which can be used for classification and prediction purposes. The neurons of a multilayer perceptron are basically threshold units: If the weighted sum of its inputs exceeds a neuron-specific threshold, a neuron becomes active and sends out a signal to the neurons it is connected to. Otherwise it remains inactive. However, for training purposes the threshold behavior is not crisp, but described by a sigmoid (s-shaped) function, which is 0.5 at the threshold and approaches 0 for values less than the threshold and 1 for values greater than the threshold. The advantage of such a "softened" threshold function is that it is differentiable, which is important for training (see below).

As its name indicates, the neurons of a multilayer perceptron are organized into layers. Connections exist only between consecutive layers and carry weights with which the transmitted activation signals are multiplied. Furthermore, a multilayer perceptron is a *feed-forward network*. That is, the connections run in one direction only, or, in other words, there are no backward connections. Most commonly, three layers are used: an input layer, one hidden layer, and an output layer (see Figure 4). The input layer receives the input values and distributes them, usually unchanged, to the neurons of the hidden layer. The neurons of the hidden layer, which are called hidden because they do not interact with the environment (i.e., they do not directly receive inputs or produce outputs), process the signals from the input neurons in the threshold fashion described above and send the results to the neurons of the output layer. The neurons of the output layer then process the signals from the hidden layer, again in the threshold fashion described above, and thus produce the output of the network.

Even though the computations that are carried out by the individual neurons are very simple, properly set up multilayer perceptrons can compute very complicated

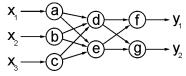


Figure 4. A simple three-layered perceptron with three inputs x<sub>1</sub>, x<sub>2</sub>, and x<sub>3</sub>; two outputs y<sub>1</sub> and y<sub>2</sub>; and seven neurons a through g.

functions. Indeed, it is fairly simple to prove that at most two hidden layers are necessary to approximate any Riemann-integrable function with arbitrary accuracy, even though this may require a fairly large number of neurons.

Multilayer perceptrons can be trained to approximate a function that is given only as a set of input/output patterns by a method that is called *error backpropagation*. The idea underlying this method is, in broad strokes, the following: The parameters of the network (connection weights and threshold values) are initialized to random values. Then the input patterns are processed with the network and the produced output is compared to the desired one. The error, usually the sum of the squared differences between the desired and actual outputs, is propagated backwards through the network and the parameters are adapted in such a way that the error gets smaller. This procedure is repeated until either the error is small enough or does not change anymore. Mathematically, the backpropagation procedure is a *gradient descent* on the error function. That is, repeatedly small steps are made in the direction in which the error gets smaller. This direction is determined by differentiating the error function.

The advantage of artificial neural networks is that they frequently produce the best results in w.r.t. accuracy compared to other methods. Their disadvantage is that the trained network is "black box." Since its "knowledge" is stored in the connection weights and threshold values, its computations are almost uninterpretable.

More detailed accounts of artificial neural network methodologies and especially of other network types and the tasks they are suited for can be found, for instance, in [21,22]. An interesting variant are neuro-fuzzy systems, which combine the learning ability of artificial neural networks with the comprehensiveness of fuzzy systems [23].

#### Genetic Algorithms

With *genetic algorithms* one tries to imitate the optimization process of biological evolution. Random mutations and recombinations of candidate solutions as well as quality dependent reproduction aim at finding (near-) optimal solutions of (combinatorial) optimization problems. Since most learning tasks can be reformulated as (combinatorial) optimization problems, genetic algorithms are almost universally applicable.

In general, a genetic algorithm approach consists of the following steps:

- 1. Find an appropriate way of encoding candidate solutions. An encoding of a candidate solution is called a *chromosome*.
- 2. Define a *fitness function* to assess the quality of a candidate solution.
- 3. Generate a random initial population of (encodings of) candidate solutions.
- 4. Evaluate the candidate solutions of the population.
- 5. Select the candidate solutions of the next generation according to their quality: The higher the fitness of a candidate solution, the higher the probability that it receives a child (a copy of itself in the next generation). The same candidate solution may be selected several times.
- 6. Modify the chromosomes representing the selected candidate solutions by applying *genetic operators* like *mutation* (randomly modify a small part of a chromosome) and *crossover* (randomly combine parts of two parent chromosomes).

- 7. Repeat steps 4 to 6 until some termination criterion is reached, for example the best individual has a given minimum quality, no improvement occurred during a certain number of steps, or a preset number of generations has been created.
- 8. The best-candidate solution of the last generation (or, if it is recorded, the best solution encountered during the whole process) is the solution found.

Which encoding of the candidate solutions is most suitable depends on the specific problem. Usually strings of characters are used in analogy to the sequential organization of information in biological chromosomes. Each position in the string corresponds to a *gene*, each character that may be at the position of a given gene to an *allele* (or *allelomorph*) of the gene. The practical advantage of such an encoding is that several standard genetic operators are available for string representations. For example, mutation can consist in randomly selecting a few positions in the string and altering the characters at these positions (see Figure 5, left side). Crossover can be defined as selecting a random cut point and exchanging the part of the strings on one side of the cut point (called a one-point crossover, see Figure 5, right side).

A simple method to achieve fitness-dependent selection is *fortune wheel selection*. A fortune wheel is set up on which each sector is associated with a candidate solution in the current population. The size of the sector reflects the fitness of the candidate solution. Each turn of the fortune wheel then selects one candidate solution for the next generation. Since better candidate solutions are associated with larger sectors, they have a better chance of getting selected (or getting selected more often).

More sophisticated genetic algorithms include *elitism*, i.e., that the best candidate solutions are always copied unchanged to the next generation, and techniques to avoid *crowding*, i.e. low diversity of the candidate solutions of a population, which can hinder improvements because of the limited amount of genetic material that is available.

The main advantages of genetic algorithms are that they are model-free (i.e., they do not presuppose a specific model of the domain under consideration) and that there are almost no limits to their applicability. Their disadvantages are that it often takes them a long time to find a solution and that their success can depend heavily on the chosen encoding of the problem. If the encoding is unsuitable, a genetic algorithm may even fail completely. This reduces the advantage of its being model-free, because it requires a lot of effort to find a good encoding, which can be just as costly as building a domain-specific model.

An extensive treatment of genetic algorithms and their variants, such as simulated annealing, genetic programming, and evolution strategies, is in [24-26].

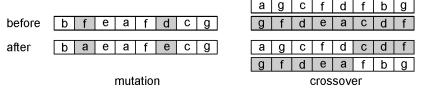


Figure 5. The genetic operators mutation and crossover.

#### References

- 1. Russell, S., and P. Norvig. 2003. *Artificial Intelligence: A Modern Approach*, 2<sup>nd</sup> ed. Upper Saddle River, NJ: Prentice Hall.
- Nilsson, N. J. 1998. Artificial Intelligence: A New Synthesis. San Francisco, CA: Morgan Kaufmann.
- 3. Genesereth, M. R., and N. J. Nilsson. 1987. *Logical Foundations of Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann.
- 4. Levesque, H. J., and G. Lakemeyer. 2001. *The Logic of Knowledge Bases*. Cambridge, MA: MIT Press.
- 5. Reiter, R. 2001. Knowledge in Action. Cambridge, MA: MIT Press.
- 6. Jensen, F. V. 2001. *Bayesian Networks and Decision Graphs*. New York, NY: Springer-Verlag.
- 7. Borgelt, C., and R. Kruse. 2002. *Graphical Models: Methods for Data Analysis and Mining*. Chichester, UK: J. Wiley and Sons.
- 8. Kruse, R., J. Gebhardt, and F. Klawonn. 1994. *Foundations of Fuzzy Systems*. Chichester, UK: J. Wiley and Sons.
- 9. Klir, G. J., and B. Yoan. 1995. Fuzzy Sets and Fuzzy Logic: Theory and Applications. Upper Saddle River, NJ: Prentice-Hall.
- 10. Zimmermann, H.-J. 1996. Fuzzy Set Theory and Its Applications. Dordrecht, Netherlands: Kluwer.
- 11. Kolodner, J. 1993. Case-Based Reasoning. San Mateo, CA: Morgan Kaufmann.
- 12. Watson, I. 1997. *Applying Case-Based Reasoning: Techniques for Enterprise Systems*. San Francisco, CA: Morgan Kaufmann.
- 13. Pearl, J. 1984. *Heuristics: Intelligent Search Strategies for Computer Problem Solving.* Reading, MA: Addison-Wesley.
- 14. Rayward-Smith, V. J., I. H. Osman, and C. R. Reeves, eds. 1996. *Modern Heuristic Search Methods*. New York, NY: J. Wiley and Sons.
- 15. Langley, P. 1995. *Elements of Machine Learning*. San Mateo, CA: Morgan Kaufmann.
- 16. Mitchell, T. 1997. Machine Learning. New York, NY: McGraw-Hill.
- 17. Lavrac, N., and S. Dzeroski. 1994. *Inductive Logic Programming: Techniques and Applications*. Upper Saddle River, NJ: Prentice Hall.
- Bergadano, F., and D. Gunetti. 1995. *Inductive Logic Programming*. Cambridge, MA: MIT Press.
- 19. Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.
- 20. Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth.
- 21. Haykin, S. 1994. *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice-Hall.
- Anderson, J. A. 1995. An Introduction to Neural Networks. Cambridge, MA: MIT Press.

- 23. Nauck, D., F. Klawonn, and R. Kruse. 1997. *Foundations of Neuro-Fuzzy Systems*. Chichester, UK: J. Wiley and Sons.
- 24. Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley.
- 25. Michalewicz, Z. 1996. Genetic Algorithms + Data Structures = Evolution Programs. Berlin, Germany: Springer-Verlag.
- Mitchell, M. 1996. An Introduction to Genetic Algorithms. Cambridge, MA: MIT Press.

# 3.5 Databases, Knowledge Discovery, Information Retrieval, and Web Mining

S. Hoche, I. Geist, L. Peña Castillo, and N. Schulz

Abstract. An abundance of digital information is now available, and large investments in data collection are being made in, for instance, the area of agribusiness. A successful exploitation of the gathered data, e.g., to extract valuable information, detect useful, frequent or extraordinary patterns, or to support complex decision processes, demands powerful means for storing, accessing and analyzing data. Database Management Systems (DBMSs) provide an efficient, integrated and standardized platform for data storage and access. Knowledge Discovery in Databases (KDD) aims at the semi-automatic discovery of useful information in large data collections usually stored in databases. Information Retrieval (IR) is concerned with gathering from unstructured and semantically fuzzy data, such as natural language texts, images, audio, or video, information relevant to a user-defined query. Web mining describes techniques to extract useful information from the World Wide Web. In this section, we present an overview of the state of the art of DBMSs and the emerging fields of KDD, IR, and web mining, and relate core methodologies to agricultural applications.

**Keywords.** Databases, Data mining, Information retrieval, Web mining.

#### 3.5.1 Introduction to Databases

Applications typically use their own methods and structures for storing data, and application developers require detailed expertise about, e.g., physical storage of or high-level interfaces to the data. A central data storage—a *database (DB)* —managed by an information system called a *database management system (DBMS)* allows applications to neglect details such as physical access paths. Furthermore, data are usually used in an application-dependent way, i.e. each application has an individual interest in the data and thus utilizes certain parts of the data in a certain way. Given a number of applications accessing some common files for data storage, the problem of *data redundancy* might appear.

A typical scenario causing redundancies is the following. A text-processing application manages texts, products, and addresses. Furthermore, there is bookkeeping soft-

ware that also stores account, product, and address information. A third system handles inventory management, and processes products and addresses. In this scenario data is stored redundantly: addresses are kept in all three systems.

A DBMS integrates all data into a single structured database, and can thus help to avoid redundancies. To this end, uniform access to the data is provided for all applications and users in form of a descriptive query language. Such a language specifies what data are accessed without any consideration of the access paths. This allows for an internal optimization of queries on large datasets.

As the data is integrated into a single database, multiple users and applications may access a database simultaneously. A DBMS supports the concurrent access of multiple users by introducing the concept of *transactions*, which are operations consisting of several database read and write actions, and their synchronization.

Since each application has its own requirements on the data, a concept of *data in-dependence* is needed. Data independence allows for the abstraction from the actual physical storage scheme and the introduction of a standardized query interface. A three-tier architecture realizing data independence comprises the following layers [1]:

- The *internal schema* describes the physical storage of the data.
- The middle tier specifies a logical, implementation independent view on the entire database. It is called the *conceptual schema*.
- The last layer consists of *external schemas* that define special views on the conceptual schema for different applications.

In addition, database management systems provide support with respect to the important concept of data safety and security. Firstly, powerful recovery and backup mechanisms assure the persistence of the data even in the case of an error. Secondly, access control techniques such as fine-grained user roles and rights prevent unauthorized access to the data

All these mechanisms are combined as follows: A database is a structured dataset that is managed by a DBMS. A DBMS consists of different software modules that handle a database. The *database system (DBS)* denotes the combination of a DBMS and a specific database.

#### Relational Databases

A conceptual schema defines a logical view of the data. The most popular conceptual data model is the *relational model* [2, 3]. It is used by many commercial database systems like Oracle 10g, Microsoft SQL Server, or IBM DB2. A relational database consists of a collection of tables. A table represents a relation. Figure 1 depicts the concepts of a relation. The table header defines the structure of the table and is called the relation schema. One row of the table is called a *tuple*. A relation is a set of all tuples, i.e. the body of the table. A column represents an attribute. The name of an attribute is given in the table header, and its values are stored in its column entries.

The relational model is a very simple data model. However, it makes use of additional *integrity constraints*. Integrity constraints are classified in *local* and *global constraints*. Local constraints describe conditions within a single table. For instance, a

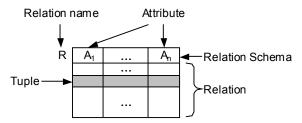


Figure 1. Relational data model.

unique key condition on an attribute specifies that the attribute values identify each tuple uniquely. Global constraints affect multiple tables. An example is the *foreign key constraint*, which requires that an attribute value present in one table also has to be contained in a second referred table. Altogether, a relation schema consists of a structure description and integrity constraints.

As described above, one of the advantages of DBMSs is the utilization of a descriptive query language. Different query languages have been proposed based on the relational model. These languages are required to be at least as powerful as the *relational algebra*, which comprises operations such as projection, selection, renaming, and join, as well as the set theoretic operations union, intersection, and difference [4]. *Selection extracts tuples from a relation that satisfy the condition defined in the selection statement. Projection allows the selection of columns, and <i>renaming* assigns a new name to a column in the query result. *Join enables the combination of tables over common attributes and values.* The most important relational query language is *SQL*, which is discussed in the next section.

#### SQL: Structured Query Language

The *structured query language* SQL is a relational query language that is supported by many vendors of commercial relational DBMSs. The current standard is SQL:2003 [5] that provides features for object-relational data processing, interfaces for data mining, information retrieval, and external data sources as well as the standard of processing XML data in SQL databases. We give only a brief overview about basic query concepts of SQL. Data or view definition features of SQL as well as advanced features are not discussed here. The interested reader is referred to the literature references, e.g., [5].

Here, we use an example to illustrate the basics of SQL. Assume the relations shown in Figure 2. The relation Milkprod holds all data about the milk production per cow in the recent years. Information about the animals is stored in the second relation Cows. Both relations are linked to each other via the common attribute No.

A basic SQL query consists of three clauses: select, from and where. The from clause specifies the relations used in the query. Attributes required in the result set, column renaming, and additional functions are specified in the select clause. A selection condition can be defined in the where clause of a query. Consider the simple query: "Return all cows, denoted as Cow\_no, with milk production of at least

Milkprod		
No	Milk	Year
110	3500	2003
111	4000	2003
110	4500	2004

Cows		
No	Price	Birth_date
110	1000	05-06-1997
111	1200	31-05-1999
113	800	11-05-2000

Figure 2. Example tables "Milkprod" and "Cows."

4000 liters in the year 2004." This query declared as an SQL statement is:

```
Select No as Cow_no, Milk
FROM Milkprod
WHERE Milk >= 4000 AND Year = 2004
```

A second example illustrates the combination of two relations via a join operation. Assume we want to know the correlation of price, age, and milk production per cow. Thus, we have to connect both tables based on the "No" attribute. One possibility to express a join in SQL is using the natural join keyword in the FROM clause. Thus, the relations Milkprod and Cows are joined via the common attribute "No". Finally, the required attributes are specified in the select statement:

```
SELECT No, Milk, Price, Birth_date
FROM Milkprod natural join Cows
```

Besides relational DBMSs, systems using other conceptual models exist. Spatial database systems support spatial queries, typical for maps, with special data models, for instance constraint databases [6]. Extensions to the relational model, e.g., the object-relational data model [7], allow the inclusion of extended query concepts such as data mining, information retrieval queries, and XML processing, as well as complex data types, into database management systems. Advanced concepts of SQL query features as well as extended SQL abilities like data mining and multimedia interfaces can be found in [1,4,8].

#### 3.5.2 Knowledge Discovery in Databases

Knowledge Discovery in Databases (KDD) aims at the discovery of useful information in large data collections usually stored in databases, for instance unusual or frequently occurring patterns or subgroups of certain properties, clusters, or rules providing a general description of the data or allowing future predictions of yet unknown objects of the same domain.

Nowadays, digital information is relatively easy to access and inexpensive to store, and rapidly growing amounts of data are collected. The number of the gathered objects is soaring as well as the number of features describing an object. Human abilities to analyze vast amounts of data lag far behind the technical means of data collection and storage. Valuable information coded in the data may remain undiscovered by conventional methods of analysis.

#### From Data to Knowledge

In traditional approaches, data analysis is centered on hypotheses formulated by some specialists. A hypothesis describes some expected properties of the data, and is, for example, translated into SQL queries to a database system. The DBS in response generates all records satisfying the hypothesized property. Such an approach to data analysis is not only slow and inefficient, but highly subjective. Analysis is biased by presumptions about the data at hand. Even worse, perspectives not covered by any of the hypotheses are not investigated, and significant patterns may remain undetected because the right query was not posed.

KDD is an automated approach to extraction of knowledge from data. It combines methods from the areas of machine learning, databases, statistics, and visualization [9,10]. The aim of KDD is not necessarily an exhaustive understanding of the data at hand, but rather the detection of valuable nuggets of information concealed within it [10]. KDD is defined as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [11]. The term *process* hints at the iterative nature of KDD (cf. next paragraph). The process being *nontrivial* means that the results cannot be obtained by some straightforward computational steps, but rather by applying sophisticated inference approaches. The claim that the identified patterns are *valid*, *novel*, *potentially useful*, *and understandable* implies that the discovered knowledge should apply to the data at hand to some high degree, lead to new insights, and be beneficial to the task at hand as well as described in a comprehensible manner

#### The KDD Process

KDD is an inherently interactive and iterative process that is depicted in Figure 3. It comprises several steps that are performed in collaboration between user and system, and which might have to be repeated several times until satisfactory results are obtained [9,11]. Process models differ slightly from one author to the other but always comprise the same elements (see for example [11, 12]).

- In a first *investigation* step, a thorough understanding of the application domain has to be developed, and the goal of the KDD task at hand has to be determined. Possible goals could, for example, be to detect a correlation between agricultural practices and yield, to identify market trends, changes in consumers' behavior, or potential future markets.
- Selection involves creating the target data set to undergo analysis. This step might comprise the integration of data from different sources, or the selection of a certain fraction of the given data. In a preprocessing step, strategies for handling noise, outliers, and missing data have to be applied to the data.
- Next, the number of records and features in the resulting data set can, if necessary, be reduced in a *transformation* step.
- The goal of the KDD process identified in the first step can now be matched to a particular data mining method. The structure of the data may already suggest a specifically useful approach. Model parameters have to be determined, and the actual analysis, or *Data Mining* step, can commence.

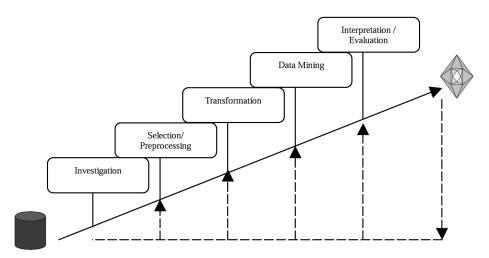


Figure 3. The KDD process: From data to knowledge.

 Finally, the mined patterns have to be evaluated and interpreted. Some or all of the previous steps may have to be repeated until the initially defined objective is met.

## Data Mining

Data mining is sometimes used synonymously with KDD. However, we will here follow the common approach [9, 11] and define Data mining as one step in the KDD process detailed in the preceding paragraph in which data are analyzed on the basis of some data mining algorithms.

Over the past years, data mining workbench systems, such as Weka [13], Enterprise Miner [14] and Clementine [15], have been successfully put into practice in a wide range of scientific (e.g., geophysics [16, 17], medicine [17-19]) and commercial areas (e.g., fraud detection [17], investment [11], risk management [19], telecommunication [17]). For a more complete overview about data mining techniques, systems, and applications, the interested reader is referred to [13,20,21].

The systems and underlying techniques designated to accomplish the data mining step of the KDD process differ depending on the knowledge discovery goal. Discovery goals can be broadly categorized as *prediction* or *classification*, and *description* [11].

Prediction or classification techniques aim at constructing classification schemes from empirical data that can be employed to predict the behavior of yet unknown objects of the same domain. Based on the attributes of the given data, hypotheses in the form of functions mapping a data item to one of several predefined classes [22] are automatically generated. The objective is to categorize the data at hand as well as future data points as accurately as possible. Predictive techniques comprise methods such as decision trees, support vector machines, neural nets, rule learners, and probabilistic nets.

In contrast, descriptive systems identify data regions of particular local interest and present the discovered patterns in a human-understandable form. Popular descriptive techniques include subgroup discovery, clustering, change and deviation detection, dependency modeling, and summarization.

## Applications of Data Mining to Agriculture

In agriculture, KDD has only recently emerged as an essential technology [23]. Little et al. [24] describe a project demonstrating the potential utility of KDD in the administration of the United States Department of Agriculture (USDA) crop insurance program. Their approach aims at detecting suspicious planting and harvesting patterns in more than one million records from an USDA database containing information about agricultural practices, cropping type (irrigated vs. non-irrigated, grain vs. silage), acres planted and acres harvested, regional characteristics, and meteorology. KDD methods are applied to identify patterns of exceptionally small yields compared to the total crop planted, which cannot be explained by a regional climatological event, and which thus might indicate misuse of the USDA crop insurance program.

In the *Data Mining for Site-Specific Agriculture* project of the Illinois Council on Food and Agricultural Research (C-FAR) [25], KDD is applied in the area of precision agriculture and variable rate application practices in an effort to improve crop yields. The objective of the project is to predict the spatial characteristics of crop yield based on interactions between spatial and temporal characteristics of weather, fertilizer, seed variety, soil properties, planting date, cropping and management history.

Canteri et al. [26] explore the usefulness of data mining methods on precision agriculture databases which, because of their size and complexity, cannot be efficiently analyzed by traditional methods. They investigate successful techniques to relate crop yield and physical-chemical soil properties.

#### 3.5.3 Information Retrieval

Information retrieval (IR) deals with searching information in collections of unstructured and semantically fuzzy data, such as natural language texts, images, audio, or video. While data retrieval systems deal with structured data stored, for example, in a relational database, and retrieve all items containing the keywords defined in a query, information retrieval systems (IRSs) aim at retrieving information about a subject rather than data that satisfy a given query. In fact, queries in information retrieval are not crisp, and an IRS aims at retrieving all information that might be relevant to a query describing the user's information need. Relevance is the central aspect in IR.

In the following, we focus on classical text information retrieval and introduce basic concepts underlying retrieval models as well as measures for the evaluation of the retrieval performance. More detailed information on IR can be found in the books by van Rijsbergen [27], Salton and McGill [28], and Beaza-Yates and Riberio-Neto [29].

#### The IR Process

The information retrieval process, which is depicted in Figure 4, comprises the *preprocessing* of the documents stored in a given data collection, their *retrieval*, and the *ranking* of relevant documents.

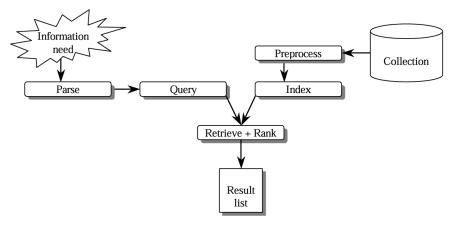


Figure 4. The information retrieval process.

In the preprocessing phase the documents' content is summarized by keywords. Thereafter, an *index* of the text is built, i.e., a data structure that supports a fast search in a large document collection. A common index structure is the *inverted file*, which consists of a vector containing all distinct keywords in the collection, and for each keyword a list of all documents in which it occurs.

Once the index is built, the retrieval process can be initiated by processing the user-defined query. The query is the system's representation of the user's information need. The initially built index structure allows a fast query processing.

The retrieved documents are then ranked according to their relevance to the given query, and the ordered result list is presented to the user. In the next section, we will describe the single steps of the information retrieval process in more detail.

# Preprocessing and Indexing

In classical text information retrieval, each document is described by a set of representative keywords, called the *index terms*. Indexing requires that documents with the same content are described by the same keywords. Such keywords can be either extracted automatically or specified by an expert.

The relevance of an index term for describing a document's content depends on the number of times it occurs in all the searched documents. An index term that appears in many documents is not very significant. However, one present in only a small fraction of the searched collection narrows down the number of relevant documents, and is therefore more meaningful. For that reason, a weight that quantifies its importance for describing a document's content is assigned to each index term.

The preprocessing of documents includes the following operations:

- Lexical analysis of the text with the goal to identify the words in the text, and to handle hyphens, digits, punctuation marks, and the case of letters;
- Elimination of *stop words* (e.g., "the" and "to") to filter out words without significance to the classification of the documents;

- Stemming of words to reduce a word to its root form, and to allow the retrieval of documents containing syntactic variations of query terms;
- Selection of *index terms* to determine which words will be used as index terms;
- Construction of a term categorization structure, such as a *thesaurus*, to expand the query with related terms.

After the preprocessing step, the indexed documents can be queried. The issued query is evaluated based on the underlying retrieval model. In what follows, we will discuss several information retrieval models

## Information Retrieval Models

The three classical models in information retrieval are the Boolean, vector and probabilistic models, which we will briefly present now. For a more detailed discussion, the interested reader is referred to [29].

The *Boolean model* is a simple retrieval model based on classical set theory and Boolean algebra. A query is specified as a Boolean expression with a crisp semantics, and the retrieval strategy is based on Boolean logic without any notion of a grading scale. Thus, the Boolean model performs a data rather than an information retrieval task. It returns *true* if an index term is present in a document, and *false* otherwise. The weights of the index terms are binary, and each document is either relevant or not.

The *vector model* [30] is based on the similarity of multi-dimensional vectors which reflect the occurrence of index terms in the query and the searched documents. More specifically, a document's relevance for a query is evaluated based on the cosine of the angle between the index term vectors of the document and the query. The retrieved documents are ranked according to their relevance, and documents might be retrieved even if they only partially match a query.

The *probabilistic model* [31] estimates the probability that a term appears in a document, or that a document satisfies the information need and is therefore relevant for a given query. The retrieved documents are ranked based on the probability values.

In general, the Boolean model is considered as the weakest classical model, since it cannot recognize partial matches [28]. Whether the vector model surpasses the probabilistic model is an ongoing discussion.

Over the past years, alternative models for each type of classical model have been proposed. The Boolean model has been enhanced by the *extended Boolean model* and the *fuzzy model* [29]. The *generalized vector model* [32] and the *latent semantic indexing model* [33] are based on the classical vector model. The *inference network* and *the belief network* are advancements of the classical probabilistic model [29].

# Ranking

After the retrieval process, the ordered list of relevant documents is presented to the user. In the vector and the probabilistic model, the documents' ranking order is based on their relevance for the given query. In the Boolean model, ranking is not possible, since a document is either in the result set or not, and all retrieved documents are equally relevant.

At this point the user can interactively refine the query by marking documents as relevant or non-relevant. Based on the user's judgment the query is reformulated, e.g., represented as a new vector of index terms in the vector model. Then, the IRS retrieves all documents that are relevant according to the refined query. This iterative query refinement process is called *relevance-feedback* [34].

#### Retrieval Performance Evaluation

It is possible that non-relevant documents are retrieved by an IRS or that relevant documents are not retrieved. For that reason there exist certain measures to determine the performance of an IRS. The standard measures of IR performance are *recall* and *precision*.

Consider a query q, and a set  $D_r$  of documents in a given collection that are relevant to q, where  $|D_r|$  is the number of documents in  $D_r$ . Based on the query q, the system retrieves an answer set  $D_a$  with  $|D_a|$  documents.  $|D_{ra}|$  is the number of documents in the answer set  $D_a$  that are relevant to q, i.e.  $|D_{ra}|$  is the cardinality of the set  $D_{ra} = D_a \cap D_r$ , as shown in Figure 5.

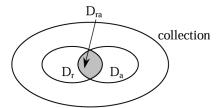


Figure 5. Precision and recall for a document collection.

The *recall*  $\bf R$  expresses how many of the relevant documents in the collection are retrieved, and is defined as the number of relevant documents retrieved compared to the total number of relevant documents in the collection:  $\bf R = |D_{ra}| / |D_{r}|$ .

The *precision* **P** describes how many of the retrieved documents are relevant, and is defined as the number of relevant documents retrieved compared to the total number of documents retrieved:  $P = |D_{ra}| / |D_a|$ 

Ideally, precision and recall should both be 100%. However, the two measures are interdependent, and an increase in recall usually results in a decrease in precision. Thus, IR systems attempt to simultaneously maximize precision and recall.

# Applications of Information Retrieval to Agriculture

In the agricultural sector, information retrieval plays an important role in the dissemination of information and knowledge about agricultural practice, market trends, price of agro-forestry products, international trade laws, legal documents, etc.

Otuka [35] applies IR techniques to share and reuse farming knowledge and experience among farmers and advisers. Textual documents describing both successful and unsuccessful farming cases are stored in a web-based system, which can be queried for information pertinent to a specific problem. The retrieval system's problem-solving abilities are improved by additionally enabling the system to retrieve agricultural spe-

cialists for a given problem, where specialists are characterized in the system by terms occurring in their publications [36].

Hoa [37] describes the use of scientific, technological, and economic information to build up an infrastructure for the agricultural development process in Vietnam. He illustrates the use of IR in the Information Centre for Agriculture and Rural Development (ICARD) of Vietnam, which maintains several databases on crop varieties, fertilizers and agro-chemicals, domestic and foreign agricultural production, and on international prices of rice, wheat, corn, coffee, rubber, and fertilizer at some main markets. Retrieved information is applied to support policy making, strategic planning and decision processes, to seek export markets for agricultural products, and to enlarge investment cooperation and joint ventures between foreign countries.

## 3.5.4 Web Mining

Since its invention in 1989 as an Internet-based hypermedia initiative for global information sharing, the World Wide Web (also called the web or WWW) has had an explosive growth. The number of web users has grown at an unknown but presumably exponential rate; the information sources available on the WWW have proliferated; and the web sites, programs, and technology have been in continuous change. All these aspects have created the need for automatic tools and techniques to help the user find the desired information, and to analyze the structure and usage of the web. Thus, web mining, or the application of data mining techniques to the World Wide Web, has been the focus of several research projects in the last years. Scime [38] provides a record of current research and practical applications in web mining.

Web mining has been broadly defined as "the discovery and analysis of useful information from the World Wide Web" [39] or more precisely as "the extraction of interesting and potentially useful patterns and implicit information from artifacts or activity related to the World Wide Web" [40].

Web mining research can be classified into the categories of web content mining, web structure mining, and web usage mining [41]. Web content mining concerns the discovery of useful information from the content of web sites. Web structure mining focuses on the web's structure and the links between the web sites, and web usage mining studies the access patterns of web users.

#### Web Content Mining

Web content mining is the process of extracting knowledge from the content of web documents, including texts, images, audio, and video, or from the description of such documents. It focuses on the automatic search of information sources available online. Web content mining techniques go beyond keyword extraction used mostly in the first generation of search engines on the web.

Within web content mining, the systems can be classified based on their mining strategy into systems that directly extract information from the content of the documents, and systems that improve the search results of other tools, such as search engines and web spiders [40]. In addition, one can differentiate between agent-based and database approaches to web content mining [39].

Database approaches, such as WebLog [42] and ARANEUS [43], combine standard database querying mechanisms and data mining techniques to access and analyze information from the web, and focus on the integration of heterogeneous and semi-structured data from the web into more structured data collections, such as relational databases.

Agent-based systems can, on behalf of a particular user, autonomously or semiautonomously search for and organize relevant information from the web. They can be classified into three categories. Intelligent search agents, such as OCCAM [44], FAQ-Finder [45], and ShopBot [46], search for relevant information using characteristics of a particular domain and possibly a user profile. Information filtering agents, such as HyPursuit [47], use information retrieval techniques to automatically retrieve, filter, and categorize web documents. Personalized web agents, such as WebWatcher [48], attempt to learn user preferences and to discover the appropriate web documents. Here, we briefly describe some web content mining projects:

- Occam [44] is an information-gathering engine. The user can specify the required information as a database query, and Occam tries to use its knowledge about various sites to derive an action plan to obtain the information.
- FAQ-Finder [45] is an automated question-answering system that uses files of Frequently Asked Questions (FAQs) available on the web. The user poses a question to the system about any topic, and FAQ-Finder attempts to find the FAQ file most likely to yield an answer, searches within that file for similar questions, and returns the given answers to the user.
- ShopBot [46] is a comparison-shopping agent. It receives as input the home pages of online stores, learns how to buy from these sites, and is then able to visit these sites, obtain product information and summarize the results to the user.
- WebWatcher [48] is a tour guide agent for assisting users to browse the web.
   Based on the information the user seeks, WebWatcher accompanies the user from page to page, highlights links believed to be relevant, and learns from experience.
- The *CiteSeer* [49] project at the NEC Research Institute provides algorithms, techniques, and software to implement digital libraries. All these are implemented in the NEC Research Index (http://citeseer.ist.psu.edu/), which crawls the web to locate scientific articles, extracts information such as the citations, citation context, article title, authors, etc., and performs full-text indexing and autonomous citation indexing.

#### Web Structure Mining

Web structure mining is the process of inferring information about web pages from the web's link structure. Links between web pages can generally be interpreted as indicators for relevance or quality [50]. The rationale is that a web page that is frequently referred to is likely to be more important than a page that is seldom referenced. Accordingly, the number of web pages a document refers to may indicate the

richness or topic diversity of this document. Thus, a document with a large number of links is likely to be a good source of information.

Web structure mining has applications in search, browsing, and traffic estimation. For example, *PageRank* [51] is a global ranking of all web pages based on their location in the web's link structure, and provides the basis of the popular search engine Google (http://www.google.com/).

The HITS (Hyperlink-Induced Topic Search) algorithm [50], which searches a collection of web pages for *authority pages* (highly referenced web documents) and for *hub pages* (documents that provide links to authority pages), has been used in various web applications.

The web structure mining project *Clever* [52] incorporates several algorithms that classify a web resource based on an analysis of its link structure. The *SALSA* [53] algorithm is a stochastic approach to web structure mining based on random walks on graphs derived from the link structure between web documents.

#### Web Usage Mining

Every web server records and maintains a log entry for every single access it gets. The log files constitute a huge collection of structured information. Web usage mining, also known as web log mining, is the process of automatic discovery of interesting usage patterns from logs kept by web servers [40].

The knowledge obtained by web usage mining can be used to enhance the quality of service provided by web servers, to customize web sites to users, to improve the design and navigation of web sites, to target users for electronic commerce, and to support marketing decisions. There are several research and commercial projects concerned with web usage mining, such as <code>WebSIFT</code> [52], <code>WebLogMiner</code> [40], and <code>WebUtilization Miner</code> (WUM) [54]. Web usage mining generally comprises the following three steps:

- 1. *Preprocessing*, which consists of data cleaning, i.e. the elimination of irrelevant or noisy data, and in grouping the data into useful data abstractions;
- 2. *Pattern discovery*, which is the application of methods and algorithms to extract knowledge from the data. Some of the methods used are statistical analysis, association rules, clustering, and classification;
- 3. *Pattern analysis*, which consists of understanding the patterns or rules found in step 2, and in filtering out uninteresting patterns.

# Applications of Web Mining to Agriculture

Web mining techniques form a vital basis for information and knowledge dissemination systems and decision support systems for the agricultural sector. Pan [55] discusses methods and techniques of collecting and managing information resources from the World Wide Web to increase information dissemination for various users in the agricultural sector.

Gandhi [56] describes a specific market information system which automatically collects information from online sources, processes it and widely disseminates it to farmers, retailers, wholesalers and the government.

Kurlavicius and Kurlavicius [57] introduce a web-based agribusiness decision support system that assists in strategic planning and operative decision making in agribusiness by searching for an optimal solution for a given agricultural problem. The approach uses both web mining and information retrieval techniques. Information agents automatically search the WWW for relevant documents, and categorize them. Domain information agents extract domain specific knowledge from the collected information

# References

- 1. Date, C. J. 2000. *An Introduction to Database Systems*. 7<sup>th</sup> ed. Reading, MA: Addison-Wesley Publishing Company.
- 2. Codd, E. 1970. A relational model for large shared data banks. *Communication of the ACM* 13(6): 377-387.
- 3. Codd, E. 1982. Relational database: A practical foundation for productivity. *Communication of the ACM* 25(2): 109-117.
- 4. Elmasri, R., and S. B. Navathe. 2000. *Fundamentals of Database Systems*. 3<sup>rd</sup> ed. Reading, MA: Addison-Wesley.
- 5. Melton, J., ed. 2003. Database Languages—SQL. ISO/IEC 9075-1:2003.
- 6. Kuper, G., L. Libkin, and J. Paredaens. 2000. *Constraint Databases*. Heidelberg, Germany: Springer-Verlag.
- 7. Stonebraker, M., P. Brown, and D. Moore. 1999. *Object-Relational DBMSs Tracking the Next Great Wave*. 2<sup>nd</sup> ed. San Francisco, CA: Morgan Kaufman Publishers.
- 8. Ullman, J. D., and J. Widom. 1997. *A First Course in Database Systems*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- 9. Mannila, H. 1996. Data mining: Machine learning, statistics, and databases. *Proc.* of the 8<sup>th</sup> International Conference on Scientific and Statistical Database Management, 1-6.
- Keim, D., and H. Kriegel. 1996. Visualization techniques for mining large databases: A comparison. *IEEE Transactions on Knowledge and Data Engineering*, Special Issue on Data Mining 8(6): 923-938.
- 11. Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth. 1996. From data mining to knowledge discovery in databases. *AI Magazine* 1996: 37-54.
- 12. Mannila, H. 1997. Methods and problems in data mining. *Proc. of International Conference on Database Theory*, 41-55.
- 13. Witten, I.H., and E. Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. 2<sup>nd</sup> ed. San Francisco, CA: Morgan Kaufmann.
- 14. http://www.sas.com/technologies/analytics/datamining/miner/
- 15. http://www.spss.com/spssbi/clementine/
- 16. Fayyad, U., D. Haussler, and P. Stolorz. 1996. KDD for Science data analysis: Issues and examples. *Proc. of the 2<sup>nd</sup> International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 50-56.

- 17. Han, J., R. B. Altman, V. Kumar, H. Mannila, and D. Pregibon. 2002. Emerging scientific applications in data mining. *Communications of the ACM* 45(8): 54-58.
- 18. Lloyd-Williams, M. 1997. Discovering the hidden secrets in your data—The data mining approach to information. SCGISA & RRL.net Workshop on Health and Crime Data Analysis.
- 19. Apte, C., B. Liu, E. P. D. Pednault, and P. Smyth. 2002. Business applications of data mining. *Communications of the ACM* 45(8): 49-53.
- 20. Dunham, M. H. 2002. *Data Mining: Introductory and Advanced Topics*. Englewood Clifts, NJ: Prentice Hall, Inc.
- 21. Larose, D. T. 2004. *Discovering Knowledge in Data: An Introduction to Data Mining*. New York, NY: John Wiley.
- 22. Weiss, S. I., and C. Kulikowski. 1991. Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning, and Expert Systems. San Francisco, CA: Morgan Kaufmann.
- 23. Persidis, A. 2000. Data mining in biotechnology. *Nature Biotechnology* 18: 237-238.
- 24. Little, B., W. Johnston, A. Lovell, S. Steed, V. O'Conner, G. Westmorland, and D. Stonecypher. 2001. Data mining U.S. corn fields. *Proc. of the First SIAM International Conference on Data Mining* 1: 99-104.
- 25. http://www.gis.uiuc.edu/cfardatamining/default.htm
- Canteri, M. G, B. C. Ávila, E. L. dos Santos, M. K. Sanches, D. Kovalechyn, J. P. Molin, and Gimenez. 2002. Application of data mining in automatic description of yield behavior in agricultural areas. *Proc. of the World Congress of Computers in Agriculture and Natural Resources*, 183-189.
- 27. van Rijsbergen, C. J. 1975. *Information Retrieval*. London, UK: Butterworths.
- 28. Salton, G., and M. J. Gill. 1983. *Introduction to Modern Information Retrieval*. New York, NY: McGraw-Hill Book Co.
- 29. Baeza-Yates, R., and B. Ribeiro-Neto. 1999. *Modern Information Retrieval*. Harlow, UK: Addison-Wesley.
- 30. Salton, G. 1971. The SMART Retrieval System–Experiments in Automatic Document Processing. Englewood Cliffs, NJ: Prentice-Hall Inc.
- 31. Robertson, S. E., and K. S. Jones. 1976. Relevance weighting of search terms. *J. American Society for Information Science* 27: 129-146.
- 32. Wong, S. K. M., W. Ziarko, and P. C. N. Wong. 1985. Generalized vector space model in information retrieval. *Proc. of the 8<sup>th</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.*
- 33. Dumais, S. T. 1991. Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments and Computers* 23: 229-236.
- 34. Rocchio, J. J. 1971. Relevance feedback in information retrieval. *The SMART Retrieval System–Experiments in Automatic Document Processing*, ed. G. Salton, 313-323. Englewood Cliffs, NJ: Prentice-Hall Inc.
- 35. Otuka, A. 1999. Case retrieval and management system for agricultural case base. *Proc. of the 2<sup>nd</sup> Conference of the European Federation for Information Technology in Agriculture, Food and the Environment.*

- 36. Otuka, A. 2000. Retrieval of specialists and their works for agriculture case base. *Proc. of the 2<sup>nd</sup> Asian Conference for IT in Agriculture.*
- 37. Hoa, T. T. 1998. Database for Agriculture in Information Centre for agriculture and rural development of Vietnam. *Agricultural Information Technology in Asia and Oceania*, 25-28.
- 38. Scime A. 2004. Web Mining. Idea Group Inc. (IGI).
- 39. Cooley, R., J. Srivastava, and B. Mobasher. 1997. Web mining: Information and pattern discovery on the World Wide Web. *Proc. of the 9<sup>th</sup> IEEE International Conference on Tools with Artificial Intelligence*.
- 40. Zaïane, O. R. 1999. Resource and knowledge discovery from the internet and multimedia repositories. PhD thesis. B.C., Canada: Simon Fraser University.
- 41. Kosala, R., and H. Blockeel. 2000. Web mining research: A survey. *ACM SIGKDD Explorations* 2(1): 1-15.
- 42. Lakshmanan, L., F. Sadri, and I. N. Subramanian. 1996. A declarative language for querying and restructuring the web. *Proc. of the 6<sup>th</sup> International Workshop on Research Issues in Data Engineering: Interoperability of Nontraditional Database Systems (RIDE-NDS'96)*.
- 43. Merialdo P., P. Atzeni, and G. Mecca. 1997. Semistructured and structured data in the web: Going back and forth. *Proc. of the Workshop on the Management of Semistructured Data*.
- 44. Kwok, C., and D. Weld. 1996. Planning to gather information. *Proc. of the 14<sup>th</sup> National Conference on Artificial Intelligence*.
- 45. Hammond, K., R. Burke, C. Martin, and S. Lytinen. 1995. Faq-finder: A case-based approach to knowledge navigation. *Working Notes of the AAAI Spring Symposium: Information Gathering from Heterogeneous, Distributed Environments*. Menlo Park, CA: AAAI Press.
- 46. Doorenbos, R. B., O. Etzioni, and D. S. Weld. 1997. A scalable comparison-shopping agent for the World-Wide Web. *Proc. of the 1<sup>st</sup> International Conference on Autonomous Agents*.
- 47. Weiss, R., B. Velez, M. A. Sheldon, C. Namprempre, P. Szilagyi, A. Duda, and D. K. Gifford. 1996. HyPursuit: A hierarchical network search engine that exploits content-link hypertext clustering. *Proc. of the 7<sup>th</sup> ACM Conf. on Hypertext*.
- 48. Joachims, T., D. Freitag, and T. Mitchell. 1997. WebWatcher: A tour guide for the World Wide Web. *Proc. of the 11<sup>th</sup> International Joint Conference on Artificial Intelligence*.
- 49. Lawrence, S., K. Bollacker, and C. L. Giles. 1999. Indexing and retrieval of scientific literature. *Proc. of the 8<sup>th</sup> International Conference on Information and Knowledge Management.*
- 50. Kleinberg, J. 1999. Authorative sources in a hyperlinked environment. *Journal of the ACM* 46(5): 604-632.
- 51. Page, L., S. Brin, R. Motwani, and T. Winograd. 1998. The PageRank Citation Ranking: Bringing order to the web. Technical Report, Stanford Digital Library Technologies Project.

- 52. Chakrabarti, S., B. E. Dom, S. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, D. Gibson, and J. Kleinberg. 1999. Mining the web's link structure. *IEEE Computer* 32(8): 60-67.
- 53. Lempel, R., and S. Moran. 2001. SALSA: The stochastic approach for link-structure analysis. *ACM Transactions on Information Systems* 19: 131-160.
- 54. Spiliopoulou, M. 1999. The laborious way from data mining to web mining. International *J. Computer Systems Science and Engineering, Special Issue on "Semantics of the Web"* 14: 113-126.
- 55. Pan, S. 2004. Approaches on collecting and integrating of agricultural network information resources. *AFITA/WCCA Joint Conference on IT in Agriculture*.
- 56. Gandhi, V.P. 2004. An IT based market information system for improving marketing efficiency of fruits and vegetables in India. *AFITA/WCCA Joint Conference on IT in Agriculture*.
- 57. Kurlavicius, A., and G. Kurlavicius. 2004. Agricultural information and farm management decision support by internet. *AFITA/WCCA Joint Conference on IT in Agriculture*.

# Mechatronics and Applications

#### 4.1 Automation and Control

J. K. Schueller

**Abstract.** Sensors, actuators, and controllers can form automation or control systems to control machines or systems in their desired tasks. Classical or modern control theory can aid design and analysis of many systems. The components and techniques are discussed and some examples presented.

Keywords. Automation, Control, Sensors, Actuators.

#### 4.1.1 Introduction

Machines or systems which have the capability to self-act or self-regulate are called automated. *Automation* allows these machines and systems to perform their tasks in a productive, efficient, reliable, and accurate manner without great amounts of human intervention. *Control* is the exercise of regulation, whether by machine or human intervention. Automation allows machines to control themselves. The development of information technologies has brought more capabilities to automation and control.

Although there have long been some automatic controls, such as the float regulators on ancient Greek water clocks, control theory and an engineering understanding of automation have been rather recent developments. James Watt's centrifugal fly ball governor for steam engines was an early development and a vital contributor to the Industrial Revolution. While some fundamental theory was developed in the 1800s by Maxwell, Lyapunov, and others, most control theory was developed in the 1900s in response to the needs of long-distance telephony, World War II military, and the aerospace industry.

Agricultural and biological engineers have used automatic controls and control theory in their efforts to get their machines and systems to respond properly within the complex biological/chemical/physical environments in which they must operate. The most famous early agricultural control example is the Ferguson System, developed by Harry S. Ferguson in the 1920s, which allowed tractors to vary implement soil working depth to maintain a constant load on the tractor. There are now many examples of successful agricultural and biological automatic control implementations. Fertilizer applicators mix and apply the right fertilizer mixture according to a variable rate map and GPS position location. Environmental controls of livestock buildings keep animals healthy and productive. Automated irrigation systems apply the correct amount of water when and where it is needed. These are some examples of how automatic controls are widely applied in agricultural engineering.

#### 4.1.2 Control Theory

One way control systems are classified is by whether they are *open-loop* or *closed-loop*. In open-loop systems a command is given to a system and it is assumed the system performs properly. Closed-loop systems compare the results or output of the system to the desired output and take appropriate corrective actions. Closed-loop systems generally exhibit more accurate performance, but cost more and may tend to be more unstable. Most systems to which control theory is applied are closed-loop.

Figure 1 shows an example of a closed-loop system. The input, which indicates the desired output, is compared to the sensed output and the error between those two is used to generate a command by the controller. The actuator then generates a control action, which causes the plant (the machine or system being controlled) to behave in the desired manner. An open-loop system will not have the sensor, which allows feedback of the output. Open-loop systems are therefore more sensitive to disturbances and system parameter variations because the resulting changes in the output are not sensed.

Another way control systems can be classified is as to whether they are sequential, continuous, or discrete.

Sequential control systems cause a machine or system to go through a set series of operations. They generally do not exercise much regulation within the operations. For example, a sequential control system may remove the milking machine when a cow is finished in a milking parlor, open the gate to let the cow out, close the gate after the cow is out and then let the next cow in. Sequential control analysis and design often makes use of ladder logic. Complex sequential control systems can be analyzed with Boolean algebra, truth tables, flowcharts, or state diagrams. Contemporary sequential control systems may use programmable logic controllers (PLCs) to allow the sequence of control to be easily modified in software rather than requiring hardware or connection changes.

Continuous control systems are the usual subject of control theory analysis [1-3]. Such systems are physical systems in which the input-output behavior of the system can be described by ordinary differential equations in time. Many of the components have a relationship that can be modeled as a constitutive equation of the relationship of a through variable to an across variable. For example, the current through an electrical component or the fluid flow through an orifice can be related to the voltage across or the pressure across that component.

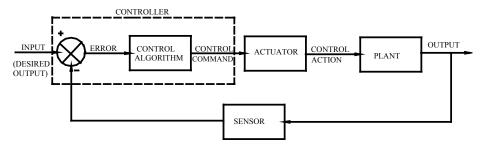
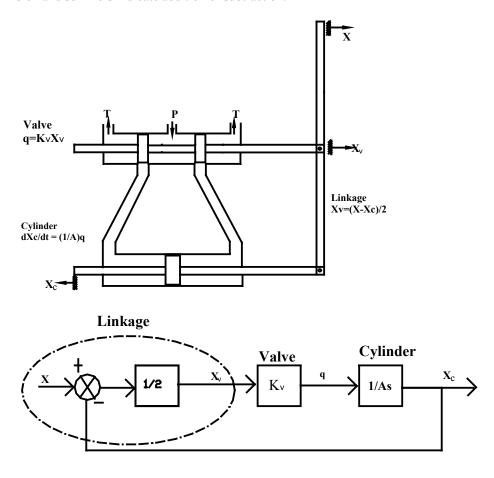


Figure 1. Closed-loop control system.

Physical systems can be modeled by writing the differential equations for the various components, be they electrical, mechanical, fluid, heat transfer, electromechanical, hydraulic, or some other system. The equations can then be transformed to the Laplace domain to get a transfer function which describes the dynamic input-output relationship.

A block diagram, such as Figure 2, with mathematical transfer functions replacing the component names can be generated to combine the various components into a complete system. Block diagram algebra can then be used to simplify the representation and calculate the overall transfer function. Blocks indicate multiplication or division and summers indicate addition or subtraction.



 $Xc/X = (\frac{1}{2})Kv (1/As)/(1+(\frac{1}{2})(Kv)(1/As)1) = 1/((2A/Kv)s+1)$ 

Figure 2. Block diagram and transfer function of a hydromechanical servo.

Most practical control theory, including all of it discussed here, is only applicable to linear systems, those that have a linear relationship between the input and the output. However, nonlinear systems can often be approximated by linearizing them at the points of normal operation. Although agricultural and biological systems may also be very complex, many times there are dominant behavioral characteristics. Many such systems can be modeled as approximately being a gain, delay, integrator, first-order, or second-order system. The output responses of such systems can be approximated by modeling the inputs as being impulses, steps, or ramps.

The types of systems studied by agricultural and biological engineers are often so complex or unknown that they cannot be modeled analytically. However, their response behavior can be determined experimentally. Such systems can be subjected to known inputs, such as steps, sinusoids, or random excitations, to find their behavioral characteristics. The transient, frequency, or stochastic responses allow a dynamic model to be developed. Care must be taken to avoid overfitting, exciting nonlinearities (especially saturations and dead zones), and exceeding dynamic and frequency ranges of the system or the instrumentation. But such step, swept-sine, and stochastic modeling can be very powerful.

Many contemporary systems are controlled by computers. If the computer control system is dynamically very fast compared to the system being controlled, it may be modeled as a continuous controller. However, many times the theory of *discrete*, or digital, control is used. This type of control recognizes that the computer only interfaces with the system at fixed instances in time. Most often, z-transforms are used in place of Laplace transforms as the classical discrete analytical tool.

Modern control theory is replacing classical control theory in many situations. Modern control theory uses state-space techniques in which the state of the system is modeled by equations which describe how the state variables change. Computers and the refinement of linear algebra have made modern control analyses tractable. Section 3.2 (Control and Optimization) of this handbook has more details on modern control theory as well as some of the contemporary advances in the areas of robust control and fuzzy control. Modern, robust, fuzzy, and other recently developed control techniques are areas of much current research, development, and application to practice. They have much promise in improving the performance and reliability of automated machines and systems.

# 4.1.3 Control and Automation Implementation

Automatic control was historically implemented by mechanical components. For example, Watt's governor used centrifugal force on rotating masses to move a linkage, which actuated a steam valve. In a similar manner, Ferguson's tractor draft control system used the force against a spring to cause a mechanical displacement, which controlled a hydraulic valve, in turn causing the draft-causing implement to be raised or lowered. Many systems still use mechanical elements in the control system. But the development of electronics and computers has widened the use of automatic controls and improved performance.

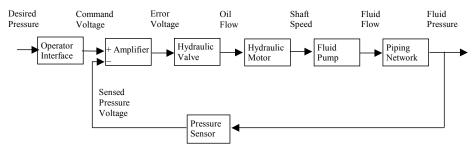


Figure 3. Pressure control system.

An example of the evolution of controls can be seen in the thermostatic temperature controls of buildings. Early fans, heaters, and air conditioning systems often used mechanical bimetallic strips with mercury switches to keep the temperature of buildings housing plants, animals, or humans at the desired temperature. The differential thermal expansion of the bimetallic strips repositioned a bubble of mercury in a curved tube, thereby activating an electrical circuit supplying power to the climate control equipment until the temperature changed to the desired setting and the bimetallic strips caused the circuit to be unpowered. Contemporary climate control systems may use computers, sensors, and actuators to achieve better and more sophisticated control. The principle of sensing the output, comparing it to the desired output, and taking corrective action remains whether the control system is mechanical or electronic.

In Figure 1, the plant is the object that is to be controlled. The machine or system must have an output, which can be measured, and a means for the behavior of the plant to be affected by the control action. The other components of Figure 1 are added to the system to complete the closed-loop control system. As an example, Figure 3 shows a pressure control system of the type used in liquid pesticide or fertilizer applications. The valve, motor, pump and piping networks form the plant to be controlled.

Sensors must accurately measure the output of the machine or system; otherwise the controller will take wrong actions and the output of the system will be accordingly wrong. Besides accuracy, sensors must have adequate resolution and range. In addition, the sensors should have fast dynamic response compared to the plant and the controller.

Analog sensors provide a voltage or current proportional to the plant's output. Sensors are commonly used to measure displacement (potentiometers, linear variable differential transformers, resolvers, capacitive sensors), acceleration (accelerometers), temperature (thermocouples, resistance thermometers, thermistors), strain (strain gauges), and many other quantities. Sensors can be used with mechanical components to measure another quantity. For example, strain gauges measure strain, but put on a diaphragm, a strain gauge can measure pressure.

As computers and digital electronics are used more in automatic controls, there is increased use of digital sensors. These may be analog sensors, which have integrated electronics to supply a digital output, or they may be inherently digital. For example, the speeds of shafts on many agricultural machines are determined by counting pulses over short time periods from variable reluctance sensors near the teeth of rotating

gears. Recent advances have produced new kinds of sensors. The global positioning system (GPS) can be used for large-scale displacements. Machine vision and optical sensors are becoming more powerful and more commonly used as sensors.

The output feedback from the sensor is compared to the desired output (input) in the controller. Based upon the error between the desired output and the actual output feedback, and often the time history of that error, the controller issues a command. Determining what command the controller should issue for different errors and error histories is a task for the engineer designing the control system.

Control systems may be implemented in electronic components without the use of computers. Many such systems use operational amplifiers to compare the input and output feedback and then generate a command proportional to the error. It is possible to add circuitry to make the command also partially proportional to the integral of the error or the derivative of the error. This common type of control is known as PID-proportional, integral, and derivative. Increasing the proportional gain (the amount of command generated per unit of error) will cause the system to respond faster and have less steady-state error, but it may also decrease stability and lead to more overshoot of the desired output. Adding integral control may remove the steady-state error, but decrease stability. Adding derivative control may stabilize the system, but make it more susceptible to noise and saturation. Tuning the controller to the best control settings may improve the system's performance without adding hardware.

Small computers are often used as controllers. Microprocessors, microcontrollers, and digital signal processors (DSPs) can quickly and efficiently implement simple or complex software control routines. Often the computers run simple programs which input the desired output and the actual output, calculate the appropriate control command, and then output the command in a continuous loop. The reliance on computers is often not obvious to the users of these embedded controllers. The software may be either interrupt-driven or program-driven.

Personal computers and larger computers can also be used for control applications. Depending upon the needs of the control system, the computers may either run conventional operating systems or operating systems specifically designed for real-time control. When such general-purpose computers are used, control interfacing becomes an issue. Analog sensor signals and actuators requiring analog commands imply that the control systems have analog-to-digital and digital-to-analog converters to interface the computers with the analog hardware. Digital sensors or actuators don't require such converters, but the communications connections or buses must be compatible.

Many controllers provide an output which can be immediately input into the plant. However, in many other cases the controller output is the not appropriate physical quantity (for example, current instead of force), does not have enough power, or is of inappropriate scale for the plant input. Actuators are often used to convert the control command into a control action that can influence the behavior of the plant. It is important that the actuator provide the proper control action when commanded. In addition, it must have significantly better dynamic response than the plant to avoid degrading closed-loop control system performance. Although actuators are seldom discussed in control theory, they are important in the types of systems encountered by agricultural

and biological engineers.

The most common type of actuator is electromechanical. A voltage or current input provides electrical power to a solenoid or motor. The actuator output is a force, torque, or displacement. Electromechanical actuators may be either linear or rotary. Due to the high forces or torques involved in many agricultural applications, electrohydraulic actuators are also common. The final component of the electrohydraulic actuator, either a hydraulic motor or a cylinder depending upon whether rotary or linear output is needed, follows an electrohydraulic valve in most applications. The performance of the valves in such systems is obviously important. It must also be remembered that the displacement of a hydraulic motor or cylinder is proportional to the integral of flow.

The control action from the actuator affects the plant and hopefully changes the plant's output to achieve the desired performance. As mentioned above, the plant is the object being controlled. The control system is designed according to the characteristics of the plant and the performance required or desired. The control systems are easier to design, and usually perform better, if the plant is time invariant, meaning its parameters do not change, and linear.

A wide variety of mathematical techniques are now available to aid the design of control systems. Common classical techniques are described in most control engineering textbooks and include root locus, pole placement, compensator, and frequency domain techniques. Modern control theory techniques often use optimization methodologies. Commercial software is available to perform most common analyses. When the system is significantly complex, nonlinear, or time variant, the difficulty of obtaining closed-form solutions usually leads to the use of commercial system simulation software to find the time-domain responses of various candidate systems to typical inputs.

# 4.1.4 Automation and Control in Agriculture and Related Fields

The wide variety of agricultural systems and the diversity throughout the world makes it difficult to generalize about the application of automation and control [4-6]. However, in many such applications of automation and control, the situation is difficult. The systems to be controlled may be a complex combination of physical, chemical, and biological components. Even the example of pH control system in Figure 4 has electrical, mechanical, and chemical components. Many agricultural and biological automation systems are located outdoors or in agricultural buildings where they may be subject to a wide range of atmospheric and other environmental conditions, such as temperatures, humidity, and vibrations. The systems are often installed in remote or rural locations where the maintenance and service infrastructure is sparse. The systems must be cheap, reliable, and easy for relatively unskilled human operators. It is a demanding task.

The use of automation and control can be controversial. Whether the local economic, social, and technical situation supports automation must be determined. This can be a special concern in developing countries where the reduction of labor usage by automation may not be desired. But where it fits, automation and control can often increase the quantity and quality of the food and fiber produced, while helping to protect the environment.

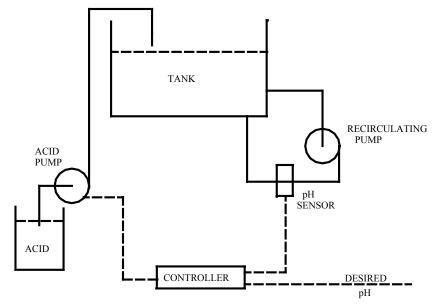


Figure 4. Example of a pH control system.

There are very many examples of automation and control being applied to agriculture and related fields. Some are simple and some are sophisticated. Table 1 lists some of the examples which can be found in other volumes of this Handbook. Many others can be found in books, papers, conference proceedings, and other literature.

The early control systems used on agricultural equipment were mechanical or hydromechanical, such as the Ferguson system and the self-leveling system for hillside combines. Now, electronic and computer controls dominate new designs. The integration of mechanical, electronic, and software elements used in most automation system is often known as *mechatronics*, especially in Europe and East Asia [7, 8].

Mechatronic systems depend on goals or commands entered into their control computers to guide their actions. Sensors also provide inputs on the state of the system and

Volume	Page	Controlled System
I	362	Irrigation
	498	Irrigation water delivery
II	304	Aquaculture
III	45	Diesel engine injection
	171	Ferguson system hitch
	309	Direct injection pesticide sprayer
	477	Greenhouse climate
	610	Precision agriculture application
IV	42	Grain dryer
	345	Cold storage refrigeration

Table 1. Some control systems in other volumes of the CIGR Handbook series.

its environment. The mechatronic systems then use software to decide on the appropriate signals to be output to the actuators. Such systems are very flexible in that simple software changes can change the system behavior. They also allow more complicated and sophisticated control algorithms.

One area in which mechatronics is achieving greater usage is the movement to *X-by-wire* in vehicles. X-by-wire systems replace mechanical or hydromechanical functions in a vehicle with a combination of mechanical, electronic, and software components. For example, conventional vehicle brakes may be replaced with a system in which there is no direct connection between the brake pedal and the brakes. Such systems require high reliability of the components and the overall system for safety.

Figure 5 shows a simplified schematic example of a system for steering a vehicle. The driver's positioning of the steering wheel is sensed and transmitted to a computer that then determines and communicates a command to another computer that controls the steering actuator. The actuator's position is closed-loop controlled by the second computer's sensing of the position of the vehicle wheels. Feedback can be supplied to the driver by the first computer through an actuator's effects on the steering wheel.

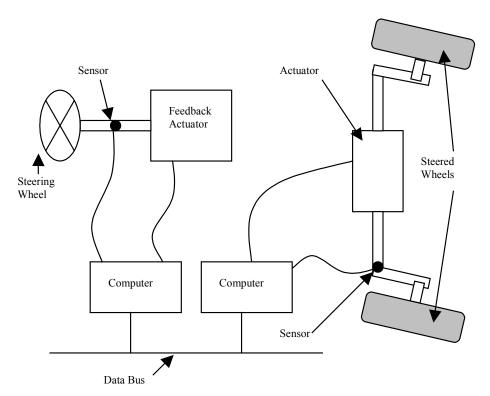


Figure 5. Simplified schematic of a vehicle steering-by-wire system.

Such a system is very flexible, due to the wide variety of algorithms which may be implemented in the computers and the other information which may be accessed from the data bus. For example, the steering system may have a variable ratio between the steering wheel and the angle of the vehicle wheels, which in turn changes with vehicle speed.

Since agricultural equipment operates in complex environments with weather, fields, plants, or animals that can vary widely, there is a great advantage to equipment with control systems that can respond to such variations. Historically, operating parameters of agricultural equipment, such as speeds and geometric clearances, were either fixed or adjusted infrequently by human operators. Automation and control systems now allow the operating parameters to be adjusted automatically in response to changing conditions to improve productivity, efficiency, and quality. In order for such control systems to be used, the machine must be capable of being adjusted. For example, fixed-ratio mechanical (such as belt, chain, or gear) drives might need to be replaced with variable-speed hydraulic or electrical drives with appropriate valving or drivers.

Agricultural equipment has evolved to accommodate such control systems [9,10]. A contemporary grain combine harvester is a good example. It may have automatic control of such items as header height, reel speed, travel speed, rotor speed, concave opening, and sieve opening. The harvester has to be designed with drives and actuators to permit control systems to do their jobs. The components of the control systems often allow more flexibility in the design and layout of agricultural equipment since electrical, and to a lesser extent hydraulic, power and signals can be transferred more easily from one part of the equipment to another than mechanical power and adjustments. For example, a rotating shaft goes straight from one component to another, but a hydraulic hose or electrical wire can bend along a convoluted path. Returning to the grain combine harvester example, the operator can now control the many functions from the operator station and the engine can be located far from power-consuming components.

Due to the demands of increased performance from agricultural equipment and the improvements in automation and control systems, especially sensors, actuators, and algorithms, automation and control systems will continue to become more prevalent in agricultural systems. The trend of networking control systems together under standards such as SAE J1939, DIN 9684, and ISO 11783 will continue to accelerate. More, better, and coordinated automation and control will contribute to better agricultural equipment.

#### References

- Dorf, R. C., and R. H. Bishop. 2005. Modern Control Systems, 10<sup>th</sup> ed. Upper Saddle River, NJ: Prentice-Hall.
- 2. Ogata, K. 2002. *Modern Control Engineering*, 4<sup>th</sup> ed. Upper Saddle River, NJ: Prentice-Hall.

- 3. Wells, R. L., J. K. Schueller, and J. Tlusty. 1990. Feedforward and feedback control of a flexible robotics arm. *IEEE Control Systems* 10(1): 9-15.
- 4. Cox, S. W. R. 1997. *Measurement and Control in Agriculture*. Oxford, UK: Blackwell Science.
- 5. Schueller, J. K. 1992. A review and integrating analysis of spatially-variable control of crop production. *Fertilizer Research* 33: 1-34.
- Searcy, S. W., ed. 1991. Automated Agriculture for the 21<sup>st</sup> Century. St. Joseph, MI: ASAE.
- 7. De Silva, C. W. 2005. *Mechatronics: An Integrated Approach*. Boca Raton, FL: CRC Press.
- 8. Histand, M. B., and D. G. Alciatore. 1999. *Introduction to Mechatronics and Measurement Systems*. New York, NY: McGraw-Hill.
- 9. Klenin, N. I., I. F. Popov, and V. A. Sakun. 1970. *Sel'skokhozyaistvennye Mashiney* (Agricultural Machines). Moscow, Russia: Kolos Publishers.
- 10. Srivastava, A. K., C. E. Goering, and R. P. Rohrbach. 1993. *Engineering Principles of Agricultural Machines*. St. Joseph, MI: ASAE.

# 4.2 Positioning and Navigation

H. W. Griepentrog, B. S. Blackmore, and S. G. Vougioukas

Abstract. This chapter covers some of the recent developments in positioning and navigation of agricultural vehicles. Reliable absolute or relative positioning of a vehicle is the basic requirement for manual and automated steering and essential for navigation of autonomous systems. Furthermore, an agricultural vehicle has to be able to perform several navigation modes within a field in order to succeed in performing a field operation.

**Keywords.** Positioning, Absolute positioning, Relative positioning, Sensor fusion, Navigation, Navigation modes.

# 4.2.1 Introduction

In the agricultural environment special characteristics appear for the navigation process of vehicles. This environment offers a very different set of circumstances to those encountered by a laboratory or indoor vehicle. A number of additional complications are raised [1]:

- Operating areas can be large and geographically separated;
- Ground surfaces are often uneven with varying tractive conditions;
- Depending on the operation, wheel slippage may be far from negligible;
- Environmental conditions (rain, fog, dust, etc.) may affect sensor observations;
- Low cost systems are required.

Fortunately, agricultural operations are carried out in semi-natural environments; a farm can generally be described by fields with known boundaries and crop plants

within a field are often arranged in particular structures (i.e., oriented crop rows and patterns). This *a priori* information can be used to improve the positioning and navigation tasks and in general increases the vehicle performance [2].

The absolute or relative position of a tractor is the basic requirement for manual navigation and automated steering and essential for navigation of autonomous vehicles. Applications in agriculture, such as asset surveying and precision farming, also rely on positional information. Geo-referenced tracking of production processes to allow a subsequent tracing becomes more important today due to food safety aspects with documentation and traceability to enhance consumers' trust.

There are two ways to define a position. The first way is to use an absolute coordinate system (e.g. a map projection such as UTM or WGS84) and define the position within this fixed frame of reference. The second way is to use a relative coordinate system. This is usually relative to the position and orientation of the tractor. Both use orthogonal Cartesian axes but the navigation directrix (which governs the movement and position of a point) may be either Cartesian (x and y or easting and northing) or vector (modulus and argument or heading and distance).

Furthermore, there are two main procedures within positioning operations to determine an unknown location, the trilateration and triangulation. By using trilateration the position of a vehicle is determined with distance measurements to the known points. In trilateration navigation systems are usually three or more transmitters mounted at known locations in the environment and one receiver on the rover. Global Positioning Systems (GPS) with absolute positioning are an example of trilateration. In triangulation there are three or more active transmitters mounted at known locations. A rotating sensor on board the vehicle measures the angles between the virtual line to the transmitter beacons and the vehicle's longitudinal axis. The unknown x-and y-coordinates and the unknown vehicle orientation can be computed based on these angle measurements.

#### 4.2.2 Positioning

#### Absolute Positioning

The most common form of absolute positioning is the *Global Navigation Satellite System (GNSS)*, which is a space-based microwave positioning system that uses trilateration between known positions of orbiting satellites. GNSS provides 24-hour three-dimensional position, velocity, and time information to a user anywhere on or near the surface of earth. Currently there are two systems available: the Global Positioning System (run by the NAVSTAR GPS Joint Program Office, http://gps.losangeles.af.mil) and the Global Orbiting Navigation Satellite System (GLONASS, operated by the Russian Federation). The European Commission decided in 2003 to build up a European GNSS which will be available in 2009 (GALILEO, http://europa.eu.int/comm/dgs/energy\_transport/galileo/intro/future\_en.htm).

The GPS system is the most commonly used GNSS today. The GPS is made up of the space segment, the control segment, and the users. The space segment consists of a constellation of at least 24 orbiting satellites.

GPS satellites transmit on two L-band carrier frequencies: 1.57542 GHz (L1) and 1.22760 GHz (L2). The L1 signal has a sequence encoded which contains two codes, a precision (P) code and a coarse/acquisition (C/A) code. The L2 carrier contains only P-code that is encrypted for military and authorized civilian users. The commercial GPS receivers utilize the L1 signal and the C/A code.

P-code users calculate their geocentric positions to about 5 m with a single handheld satellite receiver. The military has encrypted the P-codes and only authorized users can utilize it. As a consequence civilian users cannot observe the P-codes.

Selective Availability (SA) was used to dither the positional accuracy with single GPS receivers to 100 m horizontally and 156 m vertically at the 95 % level. On 2 May 2000, SA was switched off and the current horizontal accuracy is about 5 to 25 m at the 95% level.

# Differential GPS (DGPS)

Differential positioning (DGPS) can be conducted with either post- or real-time processing. The former is simpler and less expensive, while the latter is more complicated because of the need of a radio link. Differential corrections may take the form of measurement corrections or position corrections. With either approach the coordinates of one point, which is used as a reference station, must be known and available. The further a reference site is from a rover site, the more the errors at the two sites will differ and the less accurate the position determination using differential techniques will be.

To apply differential positioning in post-processing the logged data from the reference site and the rover are combined together on a computer. The proper differential corrections are computed and applied by using algorithms to match the exact time of the observations at the monitor receiver with the identical times from the rover receiver.

DGPS can provide data to an accuracy of a couple of meters in dynamic situations and even better while stationary. That improvement has a significant effect on GPS as a data resource: GPS can not only be used for coarse navigation of boats and planes. It becomes a universal measurement system capable of positioning points on a very precise scale.

Real-time DGPS uses a base station located over a known control point where it is continuously computing the difference between its known and its reported position. It then sends this correction information via radio transmitters or other satellites which broadcast the correction information. The user's GPS receiver needs to have a radio receiver or GPS receiver capable of reading correction information distributed by satellites, and will correct the recorded position recorded 1 or 2 s previously. The systems can provide real-time positioning at accuracies of 2 to 5 m. One drawback of real-time GPS systems is the time delay between when a measurement is made at the reference site and the time it takes to send and implement the correction at the rover site. This error will of course be avoided in the post-processing mode.

To improve GNSS performance to satisfy requirements, various satellite-based augmentations have been developed to transmit differential corrections to the users.

These cover large areas and the data receiver often is integrated into the GPS receiver. This type of augmentation system includes the European Geostationary Navigation Overlay Service (EGNOS). It is Europe's first step into satellite navigation. EGNOS will complement the military-controlled GPS and GLONASS systems. The correction data will improve the accuracy of the current services from about 20 m to better than 5 m. The EGNOS coverage area includes all European states but can be extended to non-European regions. Beginning in 2009 the EGNOS infrastructure will be integrated into Galileo. North America has a similar system called the Wide Area Augmentation System (WAAS) and Japan has the Multifunctional Transport Satellite Space-based Augmentation System (MSAS).

# Real-Time Kinematic GPS (RTK GPS)

Carrier-phase GPS is a method of position determination capable of providing centimeter positional accuracy. It is used extensively for surveying applications. The receivers are much more complex and expensive than the hand-held code-based GPS receivers. This technique rapidly became the means of conducting centimeter-accuracy surveying and navigation for a variety of practical and scientific purposes.

#### Pseudolites

GPS pseudolites or pseudo-satellites are ground-based transmitters that transmit GPS-identical signals in a local area. They can aid in GPS positioning in three main ways [3, 4]. First, they can be used to augment the GPS satellite constellation by providing additional ranging sources when the natural satellite coverage is inadequate. Second, pseudolites can be used as an aid to carrier-cycle ambiguity resolution when using carrier-phase differential GPS (CDGPS) for precise positioning. Third, pseudolites can be used to replace completely the GPS satellite constellation. This is generally done to emulate GPS indoor positioning or on extraterrestrial locations, e.g. positioning a vehicle on Mars [3].

# Relative Positioning

*Dead reckoning* is a mathematical procedure for determining the present vehicle location by using previous positions combined with known course and velocity information over a given time duration. Relative positioning in general can be seen as dead reckoning because it always refers, relatively, backwards to known positions.

Odometry is a widely used relative positioning method of indoor mobile robots. Due to accumulating error characteristics it is used only for short-term navigational tasks. It is the integration of incremental motion information from wheel rotation and/or steering orientation over time. Many papers are available that document the extensive research carried out in this area [5]. For outdoor conditions several problems occur that contribute to often-unacceptable errors in the positional information. In outdoor applications errors mostly appear due to wheel slippage and 3-dimensional positioning tasks.

Inertial measuring units (IMU) use gyroscopes and accelerometers to determine rotation and acceleration. Based on these measured data a course and the current position relative to a known previous point—where the measuring started—can be calculated. IMUs for outdoor operation are often extended by electronic compasses that

measure the orientation of the vehicle relative to the earth's magnetic field. IMUs achieve high positioning accuracies when they are integrated, but are still expensive. *Referenced Positioning* 

Landmarks are distinct features in an environment that a sensor can recognize. Landmarks can be of geometric shapes and they may include additional information such as particular patterns. Landmarks have a fixed and known position from where a vehicle can relatively localize itself. Before a vehicle can use landmarks for navigation, the characteristics of the landmarks must be known and stored in its controller memory. The accuracy of the calculated vehicle position depends then on how reliable landmarks are recognized and used for the localization process.

There are two types of landmarks: artificial and natural landmarks. The terms are defined by [6] as follows:

- *Natural landmarks* are those objects or features that are already in the environment and have a function other than vehicle navigation [7];
- Artificial landmarks are specially designed objects or markers that need to be placed in the environment with the sole purpose of enabling positioning.

In outdoor natural landmark navigation, the detection and matching of the characteristic features from the sensory inputs is the crucial problem. Computer vision is the most suitable sensor type for this task. Today, typical computer vision-based natural landmarks for navigation purposes in agriculture are crop rows. Another common feature recognized today is crop edges detected by a laser scanner-based proximity sensor (see Section 4.3 of this handbook). The automated operations based on these features are inter-row hoeing and steering of combine harvesters.

In artificial landmark positioning a simple configuration uses three or more detectors positioned around the workspace. A vehicle-mounted laser is swept horizontally and the time at which the beam was detected is communicated to the positioning system. By using triangulation the location of the vehicle can be determined. This system has the disadvantage of requiring a communication link between the vehicle and the detectors.

Systems that rely on a vehicle-mounted laser, when used on rough terrain, have the drawback of missing the targets. An alternative is to fix lasers in the field and mount the detector on the vehicle [8].

*Map-based* (or map-matching) *positioning* is a technique in which the vehicle uses its sensors to create a map of the unknown environment. This local map is then compared to a map previously stored in memory or requested and received from a connected GIS system. The robot can compute its actual position and orientation in the environment by matching the measured and prior stored environmental features [9].

#### Sensor Fusion

Under all operational conditions no single positional sensor alone can adequately provide the required information. Because of this lack of a single good method using either relative or absolute positioning, it is recommended to combine methods, at least one from each group. This process, based on different sensor types, is called *sensor fusion*. Most of the present navigation sensor integration techniques are based on Kal-

man filtering procedures, which represent one of the best solutions for multisensor integration. A Kalman filter is a linear and model based estimator which uses stochastic, recursive, weighted and least squares computing algorithms [10]. It has been proved for GPS parallel-tracking systems [11].

#### 4.2.3 Navigation

Most agricultural tasks need to be planned and optimized in advance, before the actual field operation will be executed [12, 13]. However, robot operation in the field is not deterministic and cannot be entirely planned, because the environment is only semi-known and semi-structured. Furthermore, the environment is dynamic; agricultural robots have to operate in the presence of other robots, humans, and/or animals.

For a navigation process, besides the position, information about velocity, attitude and heading, acceleration and angular rate is included in the problem. Although the positional data—relative and/or absolute—is one of the most important information, the knowledge about the vehicle's attitude and behavior has also to be determined by sensors. While using sensor fusion the same information can then be used for either location or behavior determination.

Vehicles are currently classified into map-based and sensor-based systems. Map-based or automated guided vehicles (AGV) vehicles are not free to alter their planned navigation route. Therefore these systems are almost fully deterministic in their behavior. The positional information is the most important data; otherwise, no acceptable behavior is possible. Sensor-based or self-guided vehicles (SGV) perform differently and rely mostly on their own local environmental awareness. Sophisticated motion control, obstacle avoidance, pattern recognition, and autonomous navigation are the basic functions to achieve safe, reliable, and accurate operations for these mobile agents. The positional information is one among other important data to describe the vehicle's attitude and its closer environment.

Furthermore, AGV and SGV mobile vehicle types can also be described as Cartesian map-based and relative sensor-based systems [14]. A hybrid system comprising a combination of both is necessary when the behavioral actions of vehicles shall become acceptable. This hybrid system then combines adaptive and goal-oriented control. Both vehicle types require navigation tasks, where, besides the position, velocity, attitude and heading, acceleration and angular rates are included in the process.

#### Navigation Tasks

For an AGV to move from A to B (navigation) a planned route is needed, including the start and end positions. The positioning system together with the automatic steering keeps the vehicle on the route.

SGVs only need to know the target position and the current position. The actual route will be calculated depending on the information from sensor systems about obstacle avoidance and distance left to the target position.

Especially for unknown environments, the navigation mode has to become more advanced. Such hybrid systems have been proposed; they consist of not only a sensor fusion but also behavior fusion [12, 15]. The main components of the system are an obstacle avoider, a goal seeker, a navigation supervisor, and an environment evaluator.

This navigator is able to perform successfully in various unknown or partially known environments, and has satisfactory ability in tackling moving obstacles [12].

An agricultural robot has to be able to perform several navigation modes within a field in order to succeed in performing a field operation.

# Navigation Mode Changer

For hybrid systems complex navigation tasks in dynamic environments require that certain elementary behaviors are activated and deactivated, or that some of their parameters are changed as the vehicle interacts with the environment and progresses in different stages of the agricultural operation. In hybrid architectures, this task is assigned to a mediator process, which acts as a hybrid automaton [16], or a discrete event system [17], i.e. a vehicle with each node corresponding to a distinct behavior.

# Navigation Modes

Vougioukas et al. [12] proposed a mode changer for a robotic agricultural operation. It uses eight simple navigation tasks: initialization, calibration, path planning, path tracking, watch-and-wait, obstacle avoidance, failure, and completion (Figure 1).

The path-tracking mode is purely deterministic, while the obstacle-avoidance mode is completely reactive. The function of the navigation task is to move the vehicle along a path that is computed by a path planner before the vehicle starts moving, while avoiding any obstacles that may exist somewhere on the path. The navigation tasks were implemented and tested with a particular platform [12]. In the following a description of the implementation of the operating modes and the mode changer is given.

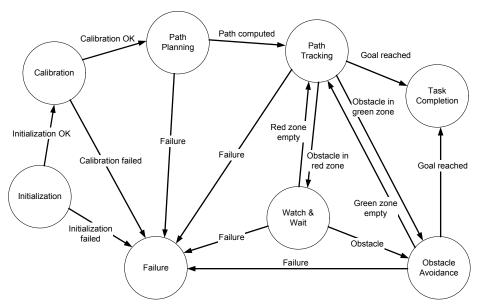


Figure 1. Mode changer: Different navigation tasks for field operations [12].

- *Task-initialization mode*—In this operating mode the robot launches and connects to all the actuator and sensor objects and performs all the necessary memory allocation for the active objects. If any problem that would prohibit the continuation of the execution of the task is encountered an appropriate message is sent to the mode manager.
- Task-calibration mode—In this mode a number of calibration procedures that involve robot motion are performed. More specifically, the compass is calibrated using filtered heading information from the GPS, while the robot moves along a small predefined distance. Also, in this mode, if any problem that would prohibit the continuation of the execution of the task is encountered an appropriate message is sent to the mode manager.
- Path-planning mode—Generally, in this mode a collision-free path that connects its current position and orientation to a desired (goal) position and orientation is computed for the vehicle, based on the task requirements (e.g., field coverage), on detailed knowledge of the field and vehicle geometry, and on the position and geometry of any existing obstacles. Path planning should take into account kinematic and dynamic constraints introduced by the vehicle and any implement it is carrying. In the case of multiple feasible paths, optimization criteria, such as minimum time travel, minimum fuel consumption, etc., can be used [18,19].
- Path-tracking mode—Path tracking or route following is the operation in which
  the vehicle follows a predetermined route. It constitutes a major research area in
  robotics and autonomous agricultural vehicles. The task is nontrivial, especially
  in the case of non-holonomic vehicles. A path can be defined as a sequence of
  waypoints connected via straight-line segments.
- Obstacle-avoidance mode—Obstacle avoidance is the operation in which the vehicle keeps moving in a desired direction while avoiding static and dynamic obstacles. The obstacle-avoidance operating mode is entered when some range sensor detects an obstacle in the "green zone" of the robot. Its implementation is based on the virtual force field (VFF) method [20]. This method combines the evidence grids (or occupancy grids) for obstacle representation with the potential field method for real-time obstacle avoidance [21].
- Watch-and-wait mode—This mode is entered when an obstacle suddenly appears close to the robot. In this case, the robot will stand still for a period of time, reading its range sensors, and wait until the obstacle leaves or is removed. If this does not happen, an appropriate message will be sent to the mode changer, which will result in entering the task failure mode.
- Task-failure mode—This operating mode can be reached from all other modes except for the task-completion mode. The actions taken in it depend greatly on the type of task being executed. In this particular case study the robot is commanded to stay still, all active objects are freed from memory, data logging files are closed, and informative error messages are printed on the standard output.
- *Task-completion mode*—In this mode the robot is commanded to stay still, all active objects are freed from memory, and data logging files are closed.

# References

- 1. Hague, T., J. A. Marchant, and N. D. Tillett. 2000. Ground based sensing systems for autonomous agricultural vehicles. *Computers and Electronics in Agriculture* 25: 11-28.
- Tillett, N. D., T. Hague, and S. J. Miles. 2002. Inter-row vision guidance for mechanical weed control in sugar beet. *Computers and Electronics in Agriculture* 33: 163-177.
- 3. LeMaster, E. 2003. GPS on the web—Applications of GPS pseudolites. *GPS Solutions* 6: 268-270.
- 4. Barnes, J., C. Rizos, J. Wang, D. Small, G. Voigt, and N. Gambale. 2003. Locata: The positioning technology of the future? *Proc. SatNav 2003 the 6<sup>th</sup> International Symposium on Satellite Navigation Technology Including Mobile Positioning and Location Services*.
- 5. Borenstein, J. 1996. Measurement and correction of systematic odometry errors in mobile robots. *IEEE Trans. Robot. Autom.* 12(6): 869-880.
- 6. Borenstein, J., H. R. Everett, L. Feng, and D. Wehe. 1997. Mobile robot positioning—Sensors and techniques. *J. Robotic Systems* 14(4): 231-249.
- 7. Olson, C. F. 2002. Selecting landmarks for localization in natural terrain. *Autonomous Robots* 12(2): 201-210.
- 8. Shmulevich, I., G. Zeltzer, and A. Brunfeld. 1989. Laser scanning method for guidance of field machinery. *Trans. ASAE* 32(2): 425-430.
- 9. Yang, H., K. Park, J. G. Lee, and H. Chung. 2000. A rotating sonar and a differential encoder data fusion for map-based dynamic positioning. *J. Intelligent and Robotic Systems* 29(3): 211-232.
- 10. De Schutter, J., J. De Geeter, T. Lefebvre, and H. Bruyninckx. 1999. Kalman filters—A tutorial. Leuven, Belgium: Katholieke Universiteit Leuven.
- 11. Han, S., Q. Zhang, and H. Noh. 2002. Kalman filtering of DGPS positions for a parallel tracking application. *Trans. ASAE* 45(3): 553-560.
- 12. Vougioukas, S., S. Fountas, B. S. Blackmore, and L. Tang. 2004. Navigation task in agricultural robots. *Proc. International Conference on Information Systems and Innovation Technologies in Agriculture, Food and Environment* (Thessaloniki, Greece), 55-64.
- 13. Blackmore, B. S., S. Fountas, S. Vougioukas, L. Tang, C. G. Sørensen, and R. N. Jørgensen. 2004. A method to define agricultural robot behaviours. *Proc. Mechatronics and Robotics*.
- 14. Freyberger, F., and G. Jahns. 2000. Symbolic course description for semiautonomous agricultural vehicles. *Computers and Electronics in Agriculture* 25(1-2): 121-132.
- 15. Ye, C., and D. Wang. 2001. A novel navigation method for autonomous mobile vehicles. *J. Intelligent and Robotic Systems* 32: 361-388.
- 16. Egerstedt, M., K. Johansson, K. Lygeros, and S. Sastry. 1999. Behavior based robotics using regularized hybrid automata. *Proc. IEEE Conference on Decision and Control (CDC '99)*.

- 17. Kosecka, J., and R. Bajcsy. 1994. Discrete event systems for autonomous mobile agents. *J. Robotics and Autonomous Systems*.
- 18. Sørensen, C. G., T. Bak, and R. N. Jørgensen. 2004. Mission planner for agricultural robotics. *Proc. AgEng, Leuven, Belgium*.
- 19. Stoll, A. 2003. Automatic operation planning for GPS-guided machinery. *Proc.* 4<sup>th</sup> European Conference on Precision Agriculture ECPA, ed J. V. Stafford, 657-664. Berlin, Germany: Wageningen Academic Press, Wageningen, NL.
- 20. Borenstein, J., and Y. Koren. 1989. Real-time obstacle avoidance for fast mobile robots. *Transactions on Systems, Man and Cybernetics* 19(5): 1179-1187.
- 21. Khatib, O. 1986. Real-time obstacle avoidance for manipulators and mobile robots. *International J. Robotics Research* 5(1): 90-98.

# 4.3 Autonomous Vehicles and Robotics

B. S. Blackmore and H. W. Griepentrog

**Abstract.** This section covers some of the recent developments in vehicle automation ranging from various forms of driver-assisted steering through totally autonomous vehicles. Commercial and research examples are given with a description of how behavioral robotics can be applied to agriculture.

**Keywords.** Autonomous tractor, Automatic steering, Agricultural robotics, Behavioral robotics.

#### 4.3.1 Driver Assistance

Various driver assistance aids are available to help reduce the complexity and difficulty of field operations as well as to help improve efficiency. These aids fall into two main groups: assistance with steering and automated implement tasks.

To improve the field efficiency, overlaps and skips of operations should be kept to a minimum. The distance between the working envelope of the implement and the previously treated area is usually judged by eye and relies on the skill and experience of the driver. As implements get wider, this task becomes more difficult, so different marking systems have been developed, such as using a disc coulter to mark the correct driving distance while cultivating and using foam markers while spraying.

# Steering Assist: Crop Edge Detection

As combine harvester sizes have grown, the corresponding header widths have increased, so that header widths of 10 m are not unusual. Judging the distance between the divider and the crop edge becomes difficult especially as it is can be 5 m to one side and slightly ahead of the driver. One method of overcoming this has been to use a

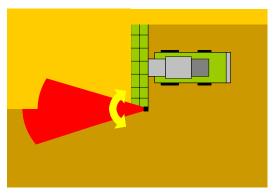


Figure 1. A laser scanner used to distinguish the edge of a standing crop.

forward-looking laser scanner mounted over the divider. The laser scanner sweeps an arc of over the edge of the standing crop and the range to each point at one degree intervals are calculated. As the standing crop is higher than the cut stubble, the crop edge can be distinguished. The distance between the crop edge and the divider can be calculated and fed into the steering control system to keep the header full.

# Steering Assist: Light Bar or Graphic Display

Driving in a straight line, at the correct distance and parallel to a previous track, has always been important to minimize skips and overlaps of field operations. Soilengaging discs and foam markers have been used, but the combination of higher-accuracy positioning systems and simple driver interfaces have enabled a semi-skilled driver to produce a skilled output. Given that the vehicle position can be measured more accurately than the driver can judge, the driver records two points A and B down the initial crop row, called the *AB line*. The working width of the implement is entered and the system calculates an infinite series of tracks parallel to the AB line. The driver watches the light bar and uses it to select the next track and to keep the tractor on course. Some commercially available systems now offer the ability to automatically steer the tractor while in the straight part of the route. Manual steering is used to turn at the end of the rows.

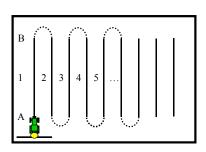




Figure 2. The AB line and the infinite (not shown) parallel tracks and a John Deere tractor using the Autotrac steering system.

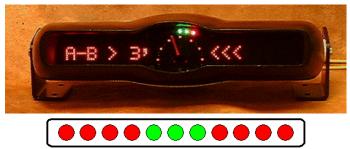


Figure 3. Light bars. Detailed indicator (above, ©Trimble); simple LED arrangement (below).

The light bar itself consists of a horizontal line of LEDs that represent the deviation of the desired position from the actual tractor position. The driver steers the tractor so that the central LEDs remain lit. Green LEDs can be used for the central dead band while red LEDs show the desired path to the left and right. More advanced versions give heading guidance and error as well as direction to the next feature (used for soil sampling).

# Driver Assistance: Self-Guiding Inter-Row Weeder

Mechanical inter-row weeding can be very efficient when carried out by skilled operators, but it is difficult to keep high accuracy over long periods of time. Lateral crop-row position data can be extracted from a camera mounted above the crop by binarising the image into soil and green, and using the knowledge that the crop was planted in rows, to find the best regression line to approximate the center of the row. These lines can be extrapolated backward to the weeder, which can be shifted laterally to run in the center of the row spacing.

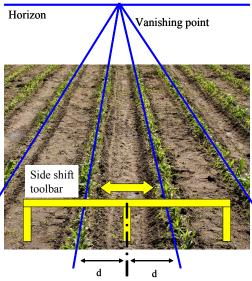


Figure 4. Crop row regression lines and an intra-row weeder schematic.

#### 4.3.2 Automatic Steering

There are a number of commercially available systems that can be retrofitted to conventional tractors that allow the steering function to be automated, while the driver attends to other tasks. Although these systems may remove part of the steering task from the operator, they cannot be considered to be autonomous, as the driver must carry out many other tasks apart from steering.

Figure 5 shows a flowchart of an automatic steering and implement control system. The positioning and orientation of the tractor is assessed by combining absolute position from an RTK GPS, relative accelerations from the IMU (fiber-optic gyroscope), and ground speed and distance from the odometry, by using a Kalman filter. The Kalman filter is used to ameliorate the errors from the positional sensors by creating an individual probabilistic error function for each sensor and adjusting a weighting function in favor of the most appropriate one. A computer in the cab (with an operator interface) sends steering and implement-control messages to a control CPU on the tractor and a job computer on the implement via a Control Area Network (CAN) bus. The tractor control CPU is tightly coupled to a number of closed-loop feedback systems such as steering, tractor speed, gearbox control, engine speed, linkage control, etc. The implement job computer is dedicated to the specific implement and controls its particular function whether it is spraying, weeding, plowing, etc. At present the tractor and operator control the implement but in the future the implement may well control the tractor.

As all of the functions, such as the proposed route and application rates, have been defined before the job starts, it can be considered a deterministic process.

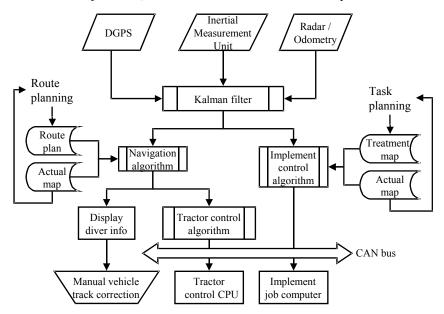


Figure 5. Flowchart of the basic tractor and implement control functions.

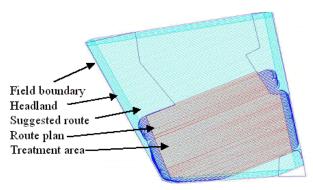


Figure 6. Route plan and treatment map for an automatically steered tractor.

A georeferenced field boundary is imported into the route planning software and the tractor and implement parameters are defined. Some parameters, such as turning circle, working speeds, implement width, etc., are extracted from a predefined database. A working direction is defined, usually along the length of the field (similar to the A-B line above) and the software creates a set of suggested linear routes on the headlands and in the main part of the field at the working width distance apart. The user then identifies the particular route required and the places where different operations or applications should occur. This route plan and treatment map is then transferred to the tractor controller as in Figure 5. The tractor is driven manually for a few seconds to calibrate the Kalman filter and then to the start position before being switched into automatic mode. The tractor then follows the route, correcting for any sensed positional or steering errors. Implement control is started when the tractor has reached the boundary of a treatment area. A log file is recorded showing all actual routes and treatments [1].

#### 4.3.3 Autonomous Tractors

In the past, agricultural engineers have developed many ways to automate agricultural tasks, and the goal of developing an autonomous tractor has almost become the "holy grail" of agricultural engineering with hundreds of papers and patents dating from the 1920s. Although many of these systems were successful in terms of automating particular tasks, none have been able to deal with the real-world complexity of the agricultural environment to become truly autonomous.

A clear distinction can be made between an automatically steered tractor and an autonomous tractor. An automatically steered tractor, as described above, needs an operator to attend to unknown object avoidance, safety, and other non-automated tasks; an autonomous tractor must be capable of working without an operator. It is obvious, from a human perspective, that we need more intelligent control of the tractor but what we humans find easy is often difficult to achieve in a computer. Furthermore, intelligence is difficult to define and can only be compared to human intelligence. Another approach is to define the actions of the tractor in terms of tasks and behaviors.

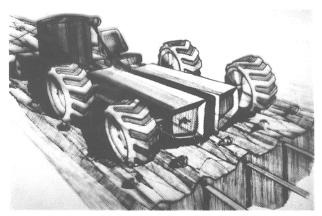


Figure 7. An early concept drawing of a wire-following tractor.

Many researchers working in robotics consider behavior-based robotics to be the most appropriate way to develop truly autonomous vehicles. In this way a definition of autonomous tractor behavior can be expressed as "sensible long-term behavior, unattended, in a semi-natural environment, while carrying out a useful task."

This sensible long-term behavior is made up of a number of parts. First, sensible behavior, which at the moment is device-independent, needs to be defined. Alan Turing defined a simple test for artificial intelligence [2], which is in essence: If a machine's behavior is indistinguishable from a person's then it must be intelligent. We cannot yet develop an intelligent machine but we can make it more intelligent than it is today by defining a set of behavior modes that make it react in a sensible way (defined by people) to a predefined set of stimuli in the form of an expert system. Second, it must be able to carry out its task over prolonged periods, unattended. When it needs to refuel or resupply, it must be capable of returning to base and restocking. Third, safety behaviors are important. The operational modes of the machine must make it safe to others as well as to itself, but it must be capable of safely deactivating when subsystems malfunction. Catastrophic failure must be avoided, so multiple levels of system redundancy must be designed into the vehicle. Fourth, because the vehicle is interacting with a complex semi-natural environment (in horticulture, agriculture, parkland and forestry uses), it must use sophisticated sensing and control systems to be able to behave correctly. Many projects in the past have found ways to simplify the environment to suit the vehicle, but the approach should now be to embed enough intelligence within the tractor to allow suitable emergent behavior to work in an unmodified environment [3].

# Purposeful Autonomous Behavior

The operation of an autonomous vehicle can be divided into two parts: tasks and behaviors. The *task* is what the tractor has been instructed to do: navigate, plow, seed, etc. The way in which it carries out the task is then called the *behavior*. Tasks and

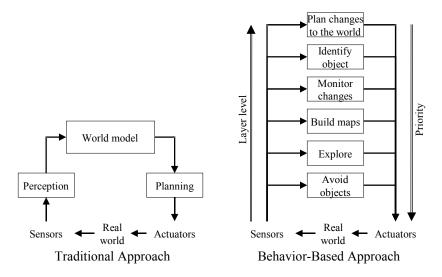


Figure 8. Comparison between the traditional and behavior-based approaches.

behaviors can be determined before an operation starts (*deterministic*) but the tractor requires the ability to react to new or unknown situations. This requires a *reactive* response to a changing *context*. The combination of both forms is called a *hybrid system*. Some low-level tasks and behaviors can be brought together to form new higher-level behaviors that may not always give the expected results. These are called *emergent* behaviors. Higher-level behaviors can be said to *subsume* lower, more primitive, behaviors [4]. Further reading on behavior-based robotics can be found in [5].

#### 4.3.4 Sensing Systems

For a vehicle to be able to interact with its environment in a sensible manner, it must be able to sense and understand its local proximity or environment. There are two main tasks: the need to identify characteristics about a particular target or point of interest (not covered here), and the ability to sense nearby objects that may become obstacles as it navigates. These sensors give proximity data relative to the vehicle. The absolute position of the relative data can be calculated given the absolute position of the vehicle and the pose of the vehicle. A number of non-contact range finders are available but the two most commonly used are ultrasonic range finders and laser scanners.

# Sensing Proximity by an Ultrasonic Ring

An *ultrasonic ring* is made up from a number of individual ultrasonic rangefinders set at fixed angles to give coverage around the vehicle (Figure 9). Each rangefinder emits a directed ultrasound "chirp" and the time taken to pick up the returned echo is proportional to the distance to the reflecting object.

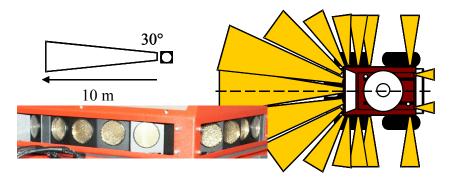


Figure 9. An ultrasonic ring comprised of 17 ultrasonic rangefinders.

At programmable intervals, each unit transmits an ultrasonic burst, waits until the ringing in the unit has stopped, and amplifies the returned signal. Amplification is increased over time to compensate for the reduced energy that has been attenuated over distance. Distances are calculated from the elapsed time between transmission and reception and the velocity of sound in air (which is taken as a constant). The operating range is between 20 cm and 10 m with a 30° dispersion angle [6].

These are cheap and reliable rangefinders but are prone to signal loss when sensing ultrasonic absorbent materials such as soft fabrics.

# Sensing Proximity by Laser Scanners

Laser scanners are used to detect an intersecting surface profile from a laser plane. One laser measurement system (Figure 10) emits a pulsed rotating laser beam at 75 Hz through 180° and the distance to each point is calculated at 1° intervals. The range depends on the reflectivity of the object being sensed, but it varies between 30 m and 150 m. Resolution is nominally 10 mm but the statistical error increases to 40 mm at the upper limits of its range. Output of the distance measurements is via an RS232 serial communications port [7].

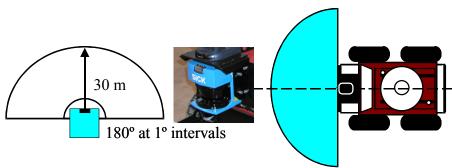


Figure 10. Laser scanner used for forward proximity sensing.

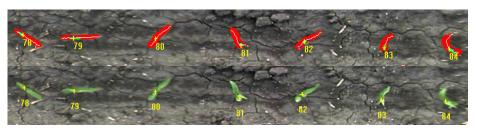


Figure 11. Augmented reality: Video images with plant count, likely stem position, and calculated leaf area [8].

## Sensing Obstacles

An object becomes an obstacle when it is likely to interfere with navigation. Most robotic systems under development assume a 2.5-dimensional world. That means all obstacles are considered to have infinite height and must be avoided. Consider that in a 3-D world a small rock is not an obstacle if a spray boom can pass safely over the top, but it is an obstacle if the tractor attempts to run over it. This type of 3-D interaction is difficult to achieve.

#### Sensing Targets

A target or point of interest is the position or object that the task is working on. This can take the form of a waypoint, crop row, or even an individual plant. Although RTK GPS can give absolute positions to the centimeter level, which may be all that is needed for some operations, specialized relative sensors are needed to identify individual targets. Cameras and machine vision techniques are often used for this purpose (see Figure 11).

# 4.3.5 Multiple Vehicles

When one autonomous tractor has been developed, it will be a relatively easy task to combine a number of them to be able to increase the work rates. Three levels of interaction have been identified.

- Coordination of multiple vehicles can be carried out centrally. Each vehicle is working independently and does not necessarily know about other vehicles but has their own task to carry out. An example would be where each vehicle would be carrying out a different task in different fields.
- Cooperation is where multiple vehicles are working in the same field and are
  aware of each other and what others are doing. If three vehicles were carrying
  out the same task, such as mechanical weeding in the same field, then each vehicle should know which rows other vehicles are working in before they select a
  new row to start in. It would not make sense for two vehicles to come head-tohead in the same row. Real time communications between vehicles on a peer-topeer basis would be needed.
- Collaboration is where multiple vehicles could share the same task at the same time. An example would be for multiple vehicles to pull a large trailer that one vehicle could not pull on its own. This is a very difficult situation to manage effectively.

# 4.3.6 Agricultural Research Vehicles

#### Robotra

Robotra (Figure 12) was designed as a tilling robot at the Institute of Agricultural Machinery in Saitama, Japan since 1993. It is a commercial tractor that has been retro-fitted with a range of positioning systems (RTK GPS, surveying grade laser range-finder, odometry, digital compass, and inertial measurement) and control systems to interface with the tractor to allow high levels of automation [9].



Figure 12. Robotra, an autonomous research tractor.



Figure 13. A small purpose-built sensing platform for crops and weeds (©DIAS).

# Autonomous Platform and Information System (API)

A small four-wheel drive, four-wheel steer platform was produced as a student project [10] and later modified to take color cameras for weed detection and hyperspectral cameras for crop health parameters. This omnidirectional platform has a small ground footprint, high crop clearance and good maneuverability that make it an ideal crop scouting platform.

#### Demeter

A New Holland 2250 windrower was retrofitted with DGPS, INS, odometry, and two cameras used to grab images of the crop in front of the machine. An image processing system was used to extract the relative position of the edge of the crop. This



Figure 14. Demeter (image from http://www-old.rec.ri.cmu.edu/projects/demeter/index.shtml).



Figure 15. Autonomous windrowing (image from http://www-old.rec.ri.cmu.edu/projects/demeter/index.shtml).

gave a relative directrix for the harvester to follow. Reliability was improved by integrating the multiple relative positioning systems and the absolute positioning systems (GPS) to remove accumulated positional offsets as well as primary guidance. Rudimentary object avoidance algorithms where incorporated into the image processing [11].

# References

- 1. Glasmacher, H. 2002. AGRO NAV Plan—Software for planning and evaluation of the path and work of field robots. *Automation Technology for Off-Road Equipment*, ed. Q. Zhang, 405-411.
- 2. Turing, A. 1950. Computing machinery and intelligence. *Mind* 59: 433-60.
- 3. Blackmore, B. S., H. Have, and S. Fountas. 2001. A specification of behavioral requirements for an autonomous tractor. *Proc. of the 6<sup>th</sup> International Symposium on Fruit, Nut and Vegetable Production Engineering,* eds. M. Zude, B. Herold, and M. Geyer, 25-36. Potsdam-Bornim, Germany: Institut für Agrartechnik Bornim e.V.
- 4. Brooks, R. A. 1986. A robust layered control system for a mobile robot. *J. Robotics and Automation* RA-2: 14-23.
- 5. Arkin, R. C. 1998. Behavior Based Robotics. Cambridge, MA: MIT Press.
- 6. Polaroid. 1993. *Polaroid ultrasonic developer's kit. PXW6431*. Polaroid Corporation. Available at: www.polaroid-oem.com.
- SICK. 1998. Laser Measurement System LMS 2000. Laser Measurement Technology, SICK optics, Sebastian-Kneipp-Straße 1, D-79183 Waldkirch, Germany. Available at: www.sick.de.
- 8. Tang, L., Tian, L., and B. L. Steward. 2000. Color image segmentation with genetic algorithm for in-field weed sensing. *Trans. ASAE* 46: 1019-1028.
- 9. Matsuo, Y., Yamamoto, S., and O. Yukumoto. 2002. Development of tilling robot and operation software. *Automation Technology for Off-Road Equipment*, ed. Q. Zhang, 184-189.
- 10. Madsen, T. E., and H. L. Jakobsen. 2001. Mobile robot for weeding. Unpublished MSc. thesis. Danish Technical University.
- 11. Pilarski, T., M. Happold, H. Pangels, M. Ollis, K. Fitzpatrick, and A. Stentz. 2002. The Demeter system for automated harvesting. *Autonomous Robots* 13: 19-20.

# Precision Agriculture

# 5.1 Image Sensing and Phytobiological Information

K. Omasa

Abstract. Phytobiological information on plants and soils is useful in order to optimize mechanized and sustainable agricultural systems. In this section, promising image-sensing techniques for obtaining phytobiological information are briefly introduced, and the concept of a phytobiological information system (PIS) for plant production and sustainable agriculture is described.

**Keywords.** Imaging, Fluorescence, Phytobiological IT, Plant functioning, Spectral image, Sustainable agriculture, Thermal image, 3-D.

#### 5.1.1 Introduction

Precision agriculture has developed with the computerization of agricultural production systems and the networking of computerized control systems [1-3]. In the intelligent plant production system of controlled greenhouses, information on plant responses, measured by sensors, is used to optimize the system. In particular, information on shapes, components, and functions of living plants obtained by image instrumentation is effectively used for diagnosis and control of production processes. Such an approach is known as the "speaking plant approach (SPA)" [4]. Recent advances in biotechnology and micropropagation demonstrate the importance of developing the SPA to the level of cells and tissues.

Meanwhile, interest in sustainable and environmental agricultural engineering is increasing [3,5,6]. New types of agricultural engineering, including precision farming, recycle-type agriculture, and controlled agroforestry, may harmonize increases in plant production with environmental conservation and remediation in order to cope with global environmental problems. Hierarchical remote sensing from satellites, aircraft, vehicles, etc., is a powerful tool for the new type of agricultural engineering because it provides useful spatial information on the functioning of plants and agroecosystems. For example, spatial information from remote sensing used with precision farming results in more accurate farm work, with localized on-demand fertilizing and localized control of weeds and pests [3,7]. In particular, precision farming using autonomous vehicles needs to join with close-up remote sensing to obtain phytobiological information on plants and soils. It is also necessary for hierarchical remote sensing to verify

remote sensing data from satellites and aircraft by exact phytobiological data from close-up remote sensing.

In this section, promising image-sensing techniques in image instrumentation and the close-up remote sensing for obtaining phytobiological information on plants and soils are introduced, and then the concept of the phytobiological information system (PIS) for using image sensing, including large-area remote sensing from satellites and aircraft, is described.

#### 5.1.2 Image-Sensing Techniques for Obtaining Phytobiological Information

Table 1 shows typical image-sensing techniques for obtaining phytobiological information on plants and soils. Trends in image sensing techniques include hyperspectral, three-dimensional, and active sensing. Ordinary multispectral imaging is available for obtaining phytobiological information on color, pigments, shape, and growth of individual plants and parts; changes in water status in plants and soils; and soil properties [8-13]. Hyperspectral image sensing is capable of resolving several hundred spectral bands in the region from visible light to short-wave infrared and may make it possible to provide more phytobiological information by analysis of continuous spectral properties, compared with ordinary multispectral analysis. Thermal infrared imaging (a passive spectral imaging method) is effective for early detection of plant stresses as well as for measurement of surface temperatures of plants and soils [9,10,14]. Image analysis of the energy balance on the leaf and canopy provides phytobiological information on stomatal response and evapotranspiration [15-17]. Fluorescence image sensing methods (which are active methods), such as spectral analysis of steady-state laser-induced fluorescence (LIF), provide phytobiological information on changes in cell walls bound by fluorophores and on bleaching of plant pigments [18,19]. Analysis of chlorophyll a fluorescence induction is used not only for early detection of patchy changes in stomatal aperture and photosynthetic activity caused by biotic and abiotic stresses, but also for information about the development of the photosynthetic apparatus of attached leaves [20,21]. Three-dimensional (3-D) surface imaging, such as stereo-pair images and laser scanners (scanning range finders), makes it possible to remotely estimate 3-D structure and growth of plants and canopies

Table 1. Image-sensing techniques for obtaining phytobiological information on plants and soils.

Image Sensing Techniques	Phytobiological Information
Multi- or hyperspectral image sensing	Color, shape, and growth of individual plants and
(near UV to near-infrared, including color)	parts, plant pigments, water status, soil properties
Thermal image sensing	• Temperature, evapotranspiration, stomatal response
• Fluorescence image sensing (LIF, Chl fluorescence, etc.)	<ul> <li>Bleaching of plant pigments, movement of fluorophores in mesophylls, photosynthetic system</li> </ul>
• 3-D surface image sensing (stereo, shape-from-x, laser scanner)	<ul> <li>3-D surface structure and biomass of plants and canopy</li> </ul>
• 3-D light microscopic imaging	• 3-D structure and functions of cells and tissues
• CT (X-ray CT, MRI, optical CT, etc.)	<ul> <li>3-D structure and content, transfer and metabolism of biochemical components of/in tissues and plants</li> </ul>

[22-25]. The *confocal laser-scanning microscope system* (a newly computerized light microscope system with a large working distance) and *CT* (computed tomography) techniques also provide information on 3-D structure and functions of and in cells, seedlings, and plants [10,26-29].

## 5.1.3 Multispectral or Hyperspectral Image Sensing

Reflectance spectra of plants and soils in the visible to short-wave infrared region (400 to 2,500 nm) include a large amount of phytobiological information [8,12,30,31]. Typical reflectance spectra of healthy and dead (dry) cucumber leaves and wet and dry loamy soils are shown in Figure 1. Absorption of photosynthetic pigments such as chlorophylls, carotenes, and xanthophylls in the leaf dominates the visible region from 400 to 700 nm [8,30,31]. Although each of the pigments has absorption maxima in the 300- to 500-nm region, only chlorophyll absorbs in the red region of 600 to 700 nm as well as the blue region. Leaf reflectance drastically increases in the region from 690 to 740 nm (red-edge) and keeps a high value to 1,300 nm (near-infrared), although the change in soil reflectance is slight. The near-infrared region is influenced by cellular structure and refractive indexes within the leaf. Absorption of water dominates the near-infrared and mid-infrared regions of 900 to 2,500 nm in healthy leaves and wet soil; and major absorption bands of water occur at 1,450 nm and 1,940 nm, and the minor absorption bands appear near 960 and 1,200 nm. Other biochemical components such as carbohydrates (starch and cellulose), protein, lignin, N, P, K, and Mg have absorption maxima in short-wave infrared region of 900 to 2,500 nm, although the molecular functional groups contributing to this are primarily limited to C-H, O-H and N-H [30]. However, note that optimum reflectance bands for estimating effects of water stress on plants and contents of biochemical components are not necessarily the absorption maxima [9,30,32].

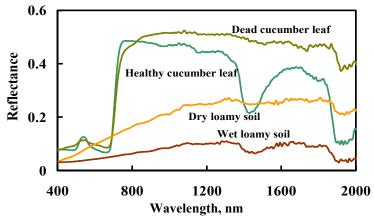


Figure 1. Typical reflectance spectra of healthy and dead (dry) cucumber leaves and wet and dry loamy soils.

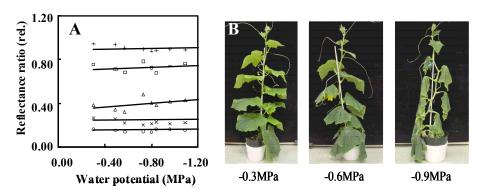


Figure 2. Changes in the reflectance band ratio of recoverable cucumber leaves with water potential above −1.2MPa (A) and the plants (B) [32]. Symbols in A: x = band ratio of 550 nm/850 nm; o = 680/850; + = 1.200/850; Δ = 1.450/850; □ = 1.650/850.

Spectral analysis of several bands with special reflectance features in the region from visible to near-infrared has been widely used for multispectral image sensing [9-11,33]. The use of the near-infrared region with high leaf reflectance allows easy separation of plants from the background soil. Spectral analysis of the visible region including color information (hue, saturation, lightness) is effective for separation of leaf, petal, fruit, and other parts of plants and their growth analyses. Leaf water status has been estimated using spectral reflectance [32,34-36]. The band ratio of 1,450 nm to 850 nm showed a good linear correlation ( $R^2 = 0.91$ ) over a wide range of water potentials above -7.1 MPa, which was the water potential of a dry leaf, but only slight changes occurred in the band ratio of recoverable leaves with water potential above -1.2 MPa (Figure 2) [32].

Changes in plant pigment content have been assessed by ratios of spectral reflectance bands. For example, chlorophyll a content was estimated at  $R^2 = 0.90$  by using a band ratio of 550 nm and 900 nm [9]. The change in red-edge also depends on chlorophyll content. Therefore, these analyses have been used to detect symptoms of nutrient deficiency and other injuries that cause changes in chlorophyll content [30,31]. Figure 3 shows a hyperspectral camera joined to a portable scanning range finder (lidar) and typical spectral reflectance and band ratio images of wheat plants cultured under different amounts of nitrogen fertilizer. The spectral reflectance images were measured at 1-nm intervals by the hyperspectral camera. As shown in Figure 1, the spectral reflectance in the visible region from 400 to 700 nm was small because of absorption by photosynthetic pigments in the leaf. In particular, the reflectance showed the minimum at the absorption maximum (680 nm) of chlorophyll a. In Figure 3, the band ratio of 550 nm to 900 nm had a good linear correlation with chlorophyll a content in the wheat field, and chlorophyll a content depends on the amount of nitrogen fertilizer, so this band ratio is effectively used to find the optimal amount of fertilizer.

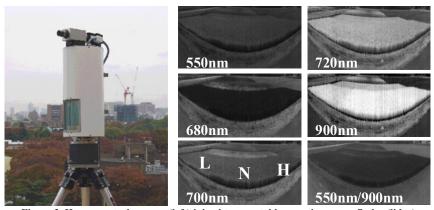


Figure 3. Hyperspectral camera (left) joined to a portable scanning range finder (lidar) and typical spectral reflectance and band ratio images (right) of a field of wheat plants cultured under different amount of nitrogen fertilizer (L = low, N = normal, H = high).

Some vegetation indexes obtained from the red and the near-infrared spectral regions made it possible to estimate net primary production (NPP) and leaf area index (LAI) as well as to extract vegetation information from the background, such as soil [8,31,37]. Hyperspectral analysis may be able to provide more phytobiological information on productivity and stresses of plants, biochemical and mineral components in living plants and soils, and classification of species, parts of plants, and soil types.

# 5.1.4 Fluorescence Image Sensing

Fluorescence image sensing is an active sensing technique. Phytobiological information is obtained by measuring the fluorescence emitted from fluorophores in living plants under irradiation with actinic light [18-21].

Figure 4 shows an excitation and emission matrix of steady-state fluorescence of a healthy cucumber leaf. When the leaf is irradiated with visible light (about 380 to 600 nm), fluorescence with very strong intensity in the spectral range of about 660 to 770 nm is emitted from chlorophyll *a*. Blue and green fluorescence is emitted from cell wall-bound phenolics, compounds in vacuoles and other fluorophores excited by ultraviolet rays and reabsorbed by photosynthetic pigments such as chlorophylls, carotenes, and xanthophylls [19,38]. Therefore, spectral analysis of steady-state fluorescence in the ultraviolet to red region (300 to 800 nm) has been used for early detection of changes in plant pigments, cell structure, and membranes [18,19,38,39]. Laser-induced fluorescence (LIF) may be effective for remote monitoring of the above-mentioned change [17,39]. Use of extrinsic fluorescent probes makes it possible to observe structures and movements of components in cells and tissues [26]. However, we should pay attention to the ambiguity of fluorescence and its dependence on the environment.

Chlorophyll *a* fluorescence also has been used for investigation of the photosynthetic system, and is a powerful tool for noninvasive analysis of photosynthesis [40,41]. Image sensing of chlorophyll fluorescence quenching in leaves gives non-uniform photosynthetic responses in adjacent tissues and can allow identification of

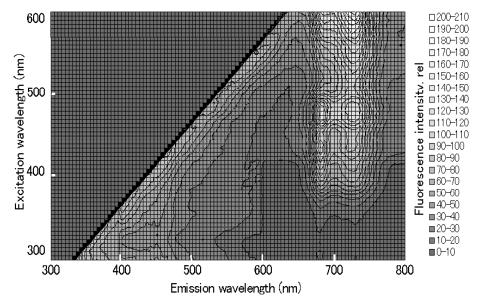


Figure 4. Excitation and emission matrix of steady-state fluorescence of a healthy cucumber leaf.

sites of inhibition within the photosynthetic apparatus. Originally developed by Omasa et al. [42] and Daley et al. [43], the techniques are used for early detection of changes in patchy stomatal response and photosynthetic activity caused by abiotic stresses such as air pollutants, low concentration O<sub>2</sub>, water deficit, UV light, chilling, agricultural chemicals and biotic stresses [42-49].

Figure 5 shows a photograph and images of nonphotochemical quenching (NPQ) and quantum yield ( $\Phi_{PSII}$ ) of PSII electron transport of an attached cucumber leaf 2 days after application of  $^{1}/_{1000}$  diluted solution of an herbicide (Nekosogi-ace) in soil. The result indicates decreased electron transport from PSII caused by DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea) in the herbicide at sites near veins with no visible injury 2 days after the treatment and consequently inhibition of *trans*-thylakoid proton

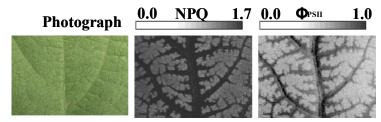


Figure 5. Photograph and images of NPQ and  $\Phi_{PSII}$  of an attached cucumber leaf at 2 days after application of an herbicide (Nekosogi-ace) in soil [49].

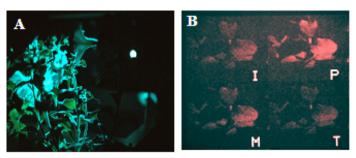


Figure 6. LIF imaging system (A) and images (B) of chlorophyll fluorescence transients [45].

gradient formation and decrease in CO<sub>2</sub> assimilation. The techniques can also be used to analyze and diagnose the development of the photosynthetic apparatus of attached leaves and cultured tissues [50-52]. Recently, field-portable imaging systems [48,53] for NPQ analysis and an LIF imaging system (see Figure 6) for remotely measuring chlorophyll fluorescence transients have been developed.

## 5.1.5 Thermal Image Sensing

Water evaporates from mesophyll cell walls in the substomatal cavity and diffuses into the atmosphere through the stomata and boundary layers of leaves and canopies. Carbon dioxide (CO<sub>2</sub>) for photosynthesis, as well as air pollutants, enters the leaf in the opposite direction [54,55]. When the thermal environment (air temperature, humidity, radiation, air current, etc.) is maintained relatively constant, leaf temperature provides phytobiological information such as stomatal response, transpiration, and absorption of CO<sub>2</sub> (photosynthesis) and air pollutants [9,10,14-17].

Thermography systems often have been used to remotely measure changes in temperature of plants and canopies (including soil) as a surrogate for stomatal conductance (= 1/stomatal resistance) and gas exchange [9,14,17,56-58]. In the latter half of the 1970s, a thermography system joined with a computer was developed for image analysis of leaf temperature [14,56]. Consequently, Omasa et al. [15,59,60] quantitatively evaluated spatial distributions of stomatal resistance, transpiration rates and absorption rates of air pollutants all over the attached leaf from leaf temperature. Recently, such a quantitative study has been noticed as a research field of thermal image sensing although it is difficult to analyze quantitatively the energy balance on all leaves of plants [16].

It is also very difficult to spatially evaluate stomatal conductance and transpiration rates of plants and canopies under growing conditions in the field. However, a thermal image can provide information for early detection of plant stresses (see Figure 7), because stomatal closure occurs before the appearance of visible injury, and for screening of plants with high growth under steady-state thermal environments [9,17]. Helicopter-borne remote sensing using a thermal camera was effective for early detection of environmental stress of canopies [17]. Microscopic thermal images provided information on responses of stomata at sites between veins of rice plants [61].

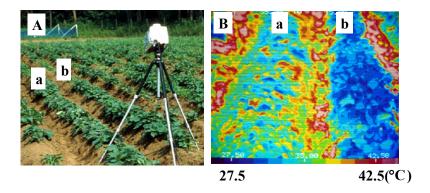


Figure 7. Effects of stomatal closure on leaf temperature of sweet potato plants in a plowed field. A, photograph; B, thermal image; a, closed stomata; b, open stomata. Air temperature 30°C.

#### 5.1.6 Three-Dimensional Image Sensing

The three-dimensional (3-D) image-sensing technology for measurement of surface architecture can be broadly divided into two categories: *passive techniques*, which reconstruct a 3-D shape image (i.e., range image or depth image) from stereo-paired images or from monocular images by shape-from-x methods, and *active techniques*, which obtain a 3-D shape image by irradiance of objects with electromagnetic waves such as laser light [22-25,27]. Active technology provides exact 3-D images of objects without clear texture, compared to passive technology, and may give more useful phytobiological information, although it is usually more expensive.

Figure 8 shows a passive 3-D color CCD camera developed by us and a 3-D texture mapping image of a flowering plant calculated from a series of nine color images obtained by changing focus planes of the camera using a modified shape-from-focus

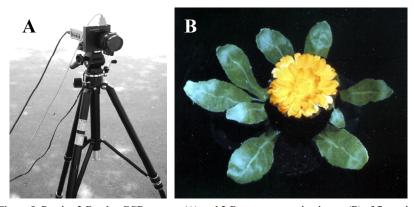


Figure 8. Passive 3-D color CCD camera (A) and 3-D texture mapping image (B) of flowering (pot marigold) plant calculated by a modified shape-from-focus (MSF) algorithm.

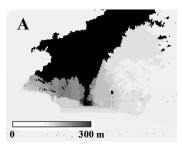




Figure 9. Range image (A) of street trees and 3-D view (B) of a potted plant measured by scanning range finders.

(MSF) algorithm [29]. This algorithm was well-suited for processing images of objects with clear texture but it was ill-suited for those with a glossy texture. This limitation was mitigated by illumination of objects with a checked pattern.

Figure 9 shows a range image of street trees and a 3-D view of a potted plant measured by scanning range finders (see Figure 3) using laser light of near-infrared range with large reflectance of leaves. The range images were obtained by measuring time-of-flight of laser light for the street trees and by a principle of triangulation for the potted plant. Omasa et al. [24] also estimates 3-D canopy structure and ground surface using a newly developed helicopter-borne scanning lidar system capable of scanning the entire canopy with a laser beam of small footprints (below tens of centimeters).

Meanwhile, the confocal laser-scanning microscope (CLSM) has been generally used to obtain information on the 3-D architecture of cells and tissues at high magnification [26,27]. In this system, the 3-D image is typically constructed by stacking numerous two-dimensional (2-D) images, which are obtained at constructive confocal planes. The CLSM has fluorescence imaging capability, especially using extrinsic fluorescent probes, and it can provide information on 3-D architecture and movements of biochemical components in cells and tissues as monochromatic or pseudo-color images. However, using the CLSM for in situ observation of cells and tissues over a wide magnification range under natural growing conditions is difficult. This problem results because, in this situation, the laser is operated at a narrow working distance and must be adjusted, thereby affecting the physiological reactions of the target cells. We have therefore developed a new computerized CCD video light microscope system with a wide working distance for obtaining 3-D natural color measurements of shape and growth of intact plants/cells under various growing conditions and over a wide magnification range [29]. This system was applied to 3-D measurement of intact petunia seedlings after providing a water supply to the seed.

Magnetic resonance imaging (MRI) and x-ray CT systems provide information on 3-D structure and functions of/in cells, seedlings, plants, and harvests. For example, root systems, difference in water contents in the soil, and water movements in the plants were three-dimensionally measured by MRI and x-ray CT [10,28,61]. These systems are also used for structural and functional research of harvests such as fruits and vegetables (Figure 10).

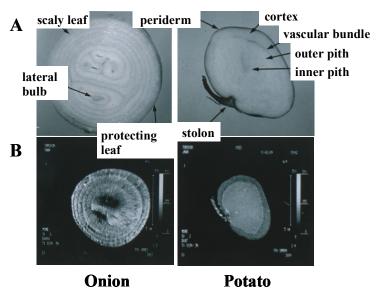
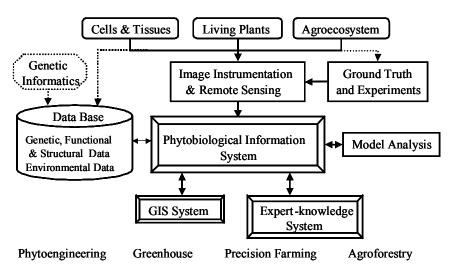


Figure 10. Section photographs (A) and MRI images (B) of onion and potato.

## 5.1.7 Phytobiological IT

Figure 11 shows a conceptual flow of phytobiological IT for plant production and sustainable agriculture. A phytobiological information system (PIS) is a management system that functions in database and model analysis. This system is connected with



Plant Production and Sustainable Agriculture

Figure 11. A conceptual flow of phytobiological IT for plant production and sustainable agriculture.

other geographic information systems (GIS) and expert-knowledge systems via network systems such as the Internet. In the system, phytobiological information on cells to agro-ecosystems obtained by imaging sensing techniques (including wide-area remote sensing described in Section 5.2) should be used, along with results of model analysis and phytobiological and environmental information obtained by other methods, to achieve both increasing plant production and optimizing use of water, fertilizers, and chemicals for sustainable agriculture. This information may also be used for improvements in farm management, in training of farmers and researchers, and in capability of machinery and control systems for phytoengineering, greenhouses, and precision farming. Results of model analysis using remote sensing data may provide a guide for agroforestry planning. Joining genetic informatics to this information may make it useful in the fields of genetic screening and eco-biomonitoring.

## References

- 1. Hashimoto, Y., G. P. A. Bot, W. Day, H.-J. Tantau, and H. Nonami, eds. 1993. *The Computerized Greenhouse*. San Diego, CA: Academic Press.
- 2. King, R. E., and N. Sigrims, eds. 2001. Special issue on computational intelligence in crop production. *Comput. Electr. Agr.* 31(1).
- 3. De Baerdemaeker, J., A. Munack, H. Ramon, and H. Speckmann. 2001. Mechatronic systems, communication, and control in precision agriculture. *IEEE Contr. Syst. Mag.* 21: 48-70.
- 4. Hashimoto, Y. 1989. Recent strategies of optimal growth regulation by the speaking plant concept. *Acta Hortic*. 260: 115-121.
- 5. Dixon, J., and M. McCann, eds. 1997. *Precision Agriculture in the 21<sup>st</sup> Century*. Washington, D.C.: National Academy Press.
- 6. Hegg, R. 1999. Sustainable and environmental engineering. *CIGR Handbook of Agricultural Engineering*, Vol. III, eds. B. A. Stout, and B. Cheze, 585-598. St. Joseph, MI: ASAE.
- Auernhammer, H., and J. K. Schueller. 1999. Precision farming. CIGR Handbook of Agricultural Engineering, Vol. III, eds. B. A. Stout, and B. Cheze, 598-616. St. Joseph, MI: ASAE.
- 8. Myers, V. I. 1983. Remote sensing applications in agriculture. *Manual of Remote Sensing*, 2<sup>nd</sup> ed., Vol. II, ed. R. N. Colwell, 2111-2228. VA: Amer. Soc. Photogrammetry.
- 9. Omasa, K., and I. Aiga. 1987. Environmental measurement: Image instrumentation for evaluating pollution effects on plants. *Systems and Control Encyclopedia*, ed. M. G. Singh, 1516-1522. Oxford, UK: Pergamon Press.
- 10. Omasa, K. 1990. Image instrumentation methods of plant analysis. *Modern Methods of Plant Analysis*, eds. H. F. Linskens, and J. F. Jackson, 203-243. Berlin, Germany: Springer-Verlag.
- 11. Kondo, N., and K. C. Ting, eds. 1998. *Robotics for Bioproduction Systems*. St. Joseph, MI: ASAE.

- 12. Ben-Dor E., J. R. Irons, and G. F. Epema. 1999. Soil reflectance. *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, 3<sup>rd</sup> ed., Vol.3, ed. A. N. Rencz, 111-188. New York, NY: John Wiley and Sons.
- 13. Hashimoto, Y., H. Murase, T. Morimoto, and T. Torii. 2001. Intelligent system for agriculture in Japan. *IEEE Contr. Syst. Mag.* 21(5): 71-85.
- 14. Hashimoto, Y., T. Ino, P. J. Kramer, A. W. Naylor, and B. R. Strain. 1984. Dynamic analysis of water stress of sunflower leaves by means of a thermal image processing system. *Plant Physiol*. 76: 266-269.
- 15. Omasa, K., and J. G. Croxdale. 1992. Image analysis of stomatal movements and gas exchange. *Image Analysis in Biology*, ed. D. P. Häder, 171-197. Boca Raton, FL: CRC Press.
- 16. Jones, H. G. 1999. Use of thermography for quantitative studies of spatial and temporal variation of stomatal conductance over leaf surface. *Plant Cell Environ*. 22: 1043-1055.
- 17. Omasa, K. 2002. Diagnosis of stomatal response and gas exchange of trees by thermal remote sensing. *Air Pollution and Plant Biotechnology*, eds. K. Omasa, H. Saji, S. Youssefian, and N. Kondo, 343-359. Tokyo, Japan: Springer-Verlag.
- 18. Lichtenthaler, H. K., ed. 1996. Vegetation stress. *Special Issue: J. Plant Physiol.* 148: 599-644.
- Kim, M. S., C. L. Mulchi, J. E. McMurtrey, C. S. T. Daughtry, and E. M. Chappele. 2002. Assessment of environmental plant stresses using multispectral steady-state fluorescence imagery. *Air Pollution and Plant Biotechnology*, eds. K. Omasa, H. Saji, S. Youssefian, and N. Kondo,. 321-341. Tokyo, Japan: Springer-Verlag.
- 20. Govindjee, J. A., and L. Nedbal. 2000. The chlorophyll fluorescence imaging and its application in plant science and technology. *Photosynthetica* 38: 481-482.
- 21. Omasa, K., and K. Takayama. 2002. Image instrumentation of chlorophyll *a* fluorescence for diagnosing photosynthetic injury. *Air Pollution and Plant Biotechnology*, eds. K. Omasa, H. Saji, S. Youssefian, and N. Kondo, 287-308. Tokyo, Japan: Springer-Verlag.
- 22. Suk, M., and S. M. Bhandarkar. 1992. *Three-dimensional Object Recognition from Range Images*. Heidelberg, Germany: Springer-Verlag.
- 23. Russ, J. C. 1995. *The Image Processing Handbook*, 2nd ed. Boca Raton, FL: CRC Press.
- 24. Omasa, K., Y. Akiyama, Y. Ishigami, and K. Yoshimi. 2000. 3-D remote sensing of woody canopy heights using a scanning helicopter-borne lidar system with high spatial resolution. *J. Remote Sens. Soc. Jpn.* 20: 394-406.
- 25. Omasa, K., Y. Urano, H. Oguma, and Y. Fujinuma. 2002. Mapping of tree position of *Larix leptolepis* woods and estimation of diameter at breast height (DBH) and biomass of the trees using range data measured by a portable scanning lidar. *J. Remote Sens. Soc. Jpn.* 22: 550-557.
- 26. Rigaut, J. P., S. Carvajal-Gonzales, and J. Vassy. 1992. Confocal image cytometry-quantitative analysis of three-dimensional images obtained by

- confocal scanning microscopy. *Image Analysis in Biology*, ed. D. P. Häder, 109-133. Boca Raton, FL: CRC Press.
- 27. Gu, M. 1996. Principles of Three Dimensional Imaging in Confocal Microscopes. River Edge, NJ: World Scientific.
- 28. Kramer, P. J., and J. S. Boyer. 1995. *Water Relations of Plants and Soils*, 162-164, 244-245. San Diego, CA:Academic Press.
- 29. Omasa, K. 2000. 3-D color video microscopy. *Image Analysis: Methods and Applications*, 2<sup>nd</sup> ed., ed. D. P. Häder, 257-273. Boca Raton, FL: CRC Press.
- 30. Wessman, C. A. 1990. Evaluation of canopy biochemistry. *Remote Sensing of Biosphere Functioning*, eds. R. J. Hobbs, and H. A. Mooney, 135-156. New York, NY: Springer-Verlag.
- 31. Ustin, S. L., M. O. Smith, S. Jacquemound, M. Verstraete, and Y. Govaerts. 1999. Geobotany: Vegetation mapping for earth sciences. *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, 3<sup>rd</sup> ed., Vol.3, ed. A. N. Rencz, 189-248. New York, NY: John Wiley and Sons.
- Fujino, M., R. Endo, and K. Omasa. 2002. Nondestructive instrumentation of water-stressed cucumber leaves—Comparison between changes in spectral reflectance, stomatal conductance, PSII Yield and shape. *Agr. Inform. Res.* 11: 161-170.
- 33. Chappelle, E. W., M. S. Kim, and J. E. McMurtrey, III. 1992. Ratio analysis of reflectance spectra (RARS): An algorithm for the remote estimation of the concentrations of chlorophyll *a*, chlorophyll *b*, and carotenoids in soybean leaves. *Remote Sens. Environ.* 39: 239-247.
- 34. Hunt, Jr. E. R., and B. N. Rock. 1989. Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sens. Environ*. 30: 43-54.
- 35. Bowman, W. D. 1989. The relationship between leaf water status, gas exchange, and spectral reflectance in cotton leaves. *Remote Sens. Environ*. 30: 249-255.
- 36. Inoue, Y., S. Morinaga, and M. Shibayama. 1993. Non-destructive estimation of water status of intact crop leaves based on spectral reflectance measurements. *Jpn. J. Crop Sci.* 62: 462-469.
- 37. Running, S. W. 1990. Estimating terrestrial primary productivity by combining remote sensing and ecosystem simulation. *Remote Sensing of Biosphere Functioning*, eds. R. J. Hobbs, and H. A. Mooney, 65-86. New York, NY: Springer-Verlag.
- 38. Lang, M., F. Stober, and H. K. Lichtenthaler. 1991. Fluorescence emission spectra of plant leaves and plant constituents. *Radiat. Environ. Bioph.* 30: 333-347.
- 39. Buschmann, C., G. Langsdorf, and H. K. Lichtenthaler. 2000. Imaging of the blue, green, and red fluorescence emission of plants: An overview. *Photosynthetica*. 38: 483-491.
- 40. Krause, G. H., and E. Weis. 1991. Chlorophyll fluorescence and photosynthesis: The basics. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 42: 313-349.
- 41. Govindjee, J. A. 1995. Sixty-three years since Kautsky: Chlorophyll *a* fluorescence. *Aust. J. Plant Physiol.* 22: 131-160.

- 42. Omasa, K., K. Shimazaki, I. Aiga, W. Larcher, and M. Onoe. 1987. Image analysis of chlorophyll fluorescence transients for diagnosing the photosynthetic system of attached leaves. *Plant Physiol.* 84: 748-752.
- 43. Daley, P. F., K. Raschke, J. T. Ball, and J. A. Berry. 1989. Topography of photosynthetic activity in leaves obtained from video images of chlorophyll fluorescence. *Plant Physiol.* 90: 1233-1238.
- 44. Omasa, K., S. Maruyama, M. A. Matthews, and J. S. Boyer. 1991. Image diagnosis of photosynthesis in water-deficit plants. *IFAC Workshop Ser.*, 1991 *No.1*, 383-388. Oxford, UK: Pergamon Press.
- 45. Omasa, K. 1998. Image instrumentation of chlorophyll *a* fluorescence. *SPIE* 3382: 91-99.
- 46. Rolfe, S. A., and J. D. Scholes. 1995. Quantitative imaging of chlorophyll fluorescence. *New Phytol.* 131: 9-79.
- 47. Siebke, K., and E. Weis. 1995. Assimilation images of leaves of *Glechoma hederacea*: Analysis of non-synchronous stomata related oscillations. *Planta*. 196: 155-165.
- 48. Osmond, C. B., P. F. Daley, M. R. Badger, and U. Lüttge. 1998. Chlorophyll fluorescence quenching during photosynthetic induction in leaves of *Abutilon striatum* Dicks, infected with abutilon mosaic virus, observed with a field-portable imaging system. *Bot. Acta.* 111: 390-397.
- 49. Takayama, K., E. Goto, and K. Omasa. 2000. Diagnosis of photosynthetic injury caused by agricultural chemicals using chlorophyll fluorescence imaging. *Proc. XIV Memorial CIGR World Congress* 2000, 1436-1441.
- 50. Croxdale, J. G., and K. Omasa. 1990. Chlorophyll *a* fluorescence and carbon assimilation in developing leaves of light-grown cucumber. *Plant Physiol*. 93: 1078-1082.
- 51. Croxdale, J. G., and K. Omasa. 1990. Patterns of chlorophyll fluorescence kinetics in relation to growth and expansion in cucumber leaves. *Plant Physiol*. 93: 1083-1088.
- 52. Omasa, K. 1992. Image diagnosis of photosynthesis in cultured tissues. *Acta Hortic*. 319: 653-658.
- 53. Daley, P. F. 1995. Chlorophyll fluorescence analysis and imaging in plant stress and disease. *Can. J. Plant Physiol.* 17: 167-17
- 54. Monteith, J. L. 1973. *Principles of Environmental Physics*. London, UK: Edward Arnold.
- 55. Jones, H. G. 1992. *Plants and Microclimate*, 2nd ed. Cambridge, UK: Cambridge University Press.
- 56. Omasa, K., F. Abo, Y. Hashimoto, and I. Aiga. 1980. Measurement of the thermal pattern of plant leaves under fumigation with air pollutant. *Res. Rep. Natl. Inst. Environ. Stud. Jpn.* 11: 239-247.
- 57. Horler, D. N. H., J. Barber, and A. R. Barringer. 1980. Effects of cadmium and copper treatments and water stress on the thermal emission from peas (*Pisium sativum* L.): Controlled environment experiments. *Remote Sens. Environ.* 10: 191-199.

- 58. Inoue, Y., B. A. Kimball, R. D. Jackson, P. J. Pinter Jr., and R. J. Reginato. 1990. Remote estimation of leaf transpiration rate and stomatal resistance based on infrared thermometry. *Agr. Forest Meteorol.* 51: 21-33.
- 59. Omasa, K., Y. Hashimoto, and I. Aiga. 1981. A quantitative analysis of the relationships between SO<sub>2</sub> or NO<sub>2</sub> sorption and their acute effects on plant leaves using image instrumentation. *Environ. Control Biol.* 19: 59-67.
- 60. Omasa, K., Y. Hashimoto, and I. Aiga. 1981. A quantitative analysis of the relationships between O<sub>3</sub> sorption and its acute effects on plant leaves using image instrumentation. *Environ. Control Biol.* 19: 85-92.
- 61. Omasa, K., and R. Funada. 1996. Image instrumentation of living plants. *BioSci.* and *Industry* 54: 545-546, 569-571.
- 62. Onoe, M., J. W. Tsao, H. Yamada, H. Nakamura, J. Kogure, H. Kawamura, and M. Yoshimatsu. 1983. Computed tomography for measuring annual rings of a live tree. *Proc. IEEE* 71: 907-908.

# 5.2 Remote Sensing from Satellites and Aircraft

K. Omasa, K. Oki, and T. Suhama

**Abstract.** This section introduces useful sensors for remote sensing from satellites and aircrafts using hyperspectral, hyperspatial, active, and 3-D observations. We also introduce recent advances in agricultural remote sensing including applications in sustainable agriculture such as precision farming, agroforestry, and land conservation.

**Keywords.** Hyperspectral sensor, Land conservation, Land use, Lidar, Precision farming, Remote sensing.

#### 5.2.1 Introduction

Remote sensing from satellites and aircraft has been widely used for applications in agriculture [1-3]. In particular, passive optical sensors mounted on aircraft, Landsat, and SPOT (the French Systeme Probatoire d'Observation de la Terra) have allowed applications such as prediction of crop production and land use change. Recent advances in agricultural remote sensing are applications in sustainable agriculture, such as precision farming, agroforestry, and land conservation. These closely relate to applications in forest, ecosystem, hydrology, and environmental management [3-5].

Meanwhile, technical trends in remote sensing from satellites and aircraft are hyperspectral, hyperspatial, active, and 3-D observations [5-11]. Although ordinary satellite optical sensors, such as the Landsat Thematic Mapper (TM) and SPOT High Resolution Visible (HRV), have been limited to less than ten spectral channels, the Hyperion on Earth Observing-1 (EO-1) launched by NASAin November 2000 provides a high resolution hyperspectral imager capable of resolving 220 spectral bands (from 0.4 to 2.5  $\mu$ m) with a 30-m spatial resolution. A hyperspatial QuickBird satellite launched in October 2001 provides panchromatic (PAN) images with a spatial resolution of

about 0.6 m. The use of active remote sensors such as lidar and SAR can provide more useful information than can passive sensors. For example, the Vegetation Canopy Lidar (VCL), which will be launched in the near future, will be able to measure 3-D ground surface and vegetation canopy with 1-m elevation accuracy by Nd:YAG diodepumped pulse lasers of a 25-m footprint [12]. Recent advances in aircraft remote sensing are similar to those in satellite remote sensing. Hierarchical remote sensing, which joins image sensing described in Section 5.1 with remote sensing from satellites and aircraft, may be submitted for more effective use in sustainable agriculture.

In this section, therefore, useful remote sensing sensors mounted on satellites and aircraft, and several applications in agriculture, are described.

### 5.2.2 Sensors for Remote Sensing

The major kinds of information obtained by remote sensing from satellites and aircraft is summarized in Table 1.

#### Remote Sensors on Satellites

MSS, TM and ETM+ (Landsat, NASA, USA)

The *Multispectral Scanner (MSS)* is an optical sensor mounted on Landsat 1 to Landsat 5. Landsat 1, with the MSS, was launched in 1972 and has verified the effectiveness of remote sensing from space. The MSS has four bands in the visible to the near-infrared region: 0.5 to 0.6  $\mu$ m (band 4), 0.6 to 0.7  $\mu$ m (band 5), 0.7 to 0.8  $\mu$ m (band 6), and 0.8 to 1.1  $\mu$ m (band 7). The spatial resolution is approximately 80 m. Furthermore, the MSS mounted on Landsat 3 is equipped with a thermal infrared band (10.4 to 12.5  $\mu$ m) with a spatial resolution of 240 m.

The *Thematic Mapper (TM)* is an optical sensor mounted on Landsat 4 and Landsat 5. It has seven bands, which are 0.45 to 0.52  $\mu$ m (band 1), 0.52 to 0.60  $\mu$ m (band 2), 0.63 to 0.69  $\mu$ m (band 3), 0.76 to 0.90  $\mu$ m (band 4), 1.55 to 1.75  $\mu$ m (band 5), 10.40 to

#### Table 1. Major information obtained by remote sensing.

Multi- or hyperspectral remote sensing (visible region to near-infrared region)

- Land (farmland, forests, natural vegetation, etc.):
   Terrain, land cover, land use, vegetation indices, plant species, phenology, biomass, crop yield, visible injuries, canopy biochemistry, soil types, chemical and physical properties of soils, fertilizer application, water status, snow and ice, water resources
- Hydrosphere (does not include atmosphere):
   Coral shelves, terrain of shallows, water pollution, plankton
- Atmosphere:

Clouds, water vapor, fog, aerosols, dust, air pollution

Thermal remote sensors

Thermal radiation, temperature of land and water area, evapotranspiration (including transpiration of plants), soil moisture, atmospheric temperature, ocean currents, clouds

Lidar

Terrain, 3-D canopy structure, biomass, aerosols, air pollutants, advective diffusion

SAR

Terrain, surface structure, clouds, rainfall

12.50 µm (band 6), and 2.08 to 2.35 µm (band 7). The spatial resolution of these bands, except band 6 (120 m), is 30 m. In comparison with the MSS, the TM is improved regarding the increased number of spectral bands and spatial resolution. As a result, the accuracy of land cover classification has been improved. Furthermore, by using these new wavelength ranges, such as band 3 and band 4, various vegetation indices for estimating vegetation conditions have been developed.

The Enhanced Thematic Mapper Plus (ETM+) is a sensor mounted on Landsat 7, launched in 1999. It is an improved version of the TM sensor on Landsat 4 and 5. The advantage of ETM+ is the improvement in spatial resolution of the thermal infrared band (band 6, spatial resolution of 60 m) and the addition of the panchromatic band (0.50 to 0.90  $\mu$ m, 15 m). In the future, applications using the thermal infrared band will become popular in the field of agriculture.

## Hyperion (EO-1, NASA, USA)

The *Hyperion*, mounted on Earth Observing-1 (EO-1) launched in 2000, is a high resolution hyperspectral imager capable of resolving 220 spectral bands (0.4 to 2.5  $\mu$ m) with a 30-m spatial resolution. It is collecting hyperspectral scenes over the course of its mission in coordination with the ETM+ on Landsat 7. Both Hyperion and ETM+ survey the same ground areas. Detailed comparisons of the Hyperion and ETM+ images may expand future applications in agriculture, forest, mining, geology, and environmental management.

## ASTER (EOS AM-1, NASA, USA)

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is an optical sensor mounted on EOS AM-1 launched in 1999. The ASTER has 3 bands (0.52 to 0.86  $\mu m$ ) with a spatial resolution of 15 m, 6 bands (1.60 to 2.43  $\mu m$ ) with a spatial resolution of 30 m, and 5 bands (8.125 to 11.65  $\mu m$ ) with a spatial resolution of 90 m. The ASTER is expected to be useful in various fields such as underground resourcing, environmental monitoring, and volcanic activity monitoring.

#### HRV and HRVIR (SPOT, CNES, France)

The *High Resolution Visible (HRV)* is an optical sensor mounted on SPOT 1, 2, and 3, and the *High Resolution Visible Infrared (HRVIR)* is mounted on SPOT 4. The SPOT was designed by the Centre National d'Etudes Spatiales (CNES), France. SPOT 1 was launched in 1986, SPOT 2 in 1990, SPOT 3 in 1993, and SPOT 4 in 1998. The HRV on SPOT 1 to 3 has two modes, namely multispectral and panchromatic modes. In the multispectral mode, there are three bands of 0.50 to 0.59  $\mu$ m (band 1), 0.61 to 0.68  $\mu$ m (band 2), and 0.79 to 0.89  $\mu$ m (band 3) with the spatial resolution of 20 m. In the PAN mode, the spatial resolution is 10 m of 0.51 to 0.73  $\mu$ m wavelength. The modifications from HRV to HRVIR on SPOT 4 are the improvement in spatial resolution of band 2 (10 m) and the addition of short-wave infrared band from 1.58 to 1.75  $\mu$ m (band 4, 20 m). The remotely sensed imagery measured by these sensors provides vegetation maps, lithologic maps, and detailed ground surface information. Furthermore, SPOT 5 was launched in 2002. It was improved to a spatial resolution of 10 m in band 1 to band 3 and of 2.5 m and 5 m in the PAN band from 0.48 to 0.71  $\mu$ m.

## IKONOS (Space Imaging, USA)

The *IKONOS* is an earth-observation satellite launched by Space Imaging/EOSAT in 1999. The major advantage of IKONOS is its high spatial resolution. For example, the spatial resolution of PAN (0.45 to 0.90  $\mu m$ ) and multispectral [0.45 to 0.52  $\mu m$  (blue), 0.52 to 0.60  $\mu m$  (green), 0.63 to 0.69  $\mu m$  (red), and 0.76 to 0.90  $\mu m$  (Near-IR)] bands is 0.82 m and 3.3 m, respectively. The high spatial resolution images measured by IKONOS are used for applications in urban, forest, and agricultural areas like airborne remote sensing images. These are also expected to be useful in precision farming.

## QuickBird (DigitalGlobe, USA)

The *QuickBird* of DigitalGlobe was launched in 2001. The spectral bands of QuickBird are similar to those of IKONOS. However, QuickBird has higher spatial resolution. The spatial resolution of PAN (0.45 to 0.90  $\mu$ m) and multispectral [0.45 to 0.52  $\mu$ m (blue), 0.52 to 0.60  $\mu$ m (green), 0.63 to 0.69  $\mu$ m (red), and 0.76 to 0.90  $\mu$ m (Near-IR)] bands are 0.61 m and 2.44 m, respectively. The QuickBird is expected to have effective use in urban, forest, and agriculture applications, like the IKONOS.

## AVHRR (TIROS-T and NOAA, NOAA, USA)

Initially the *Advanced Very High Resolution Radiometer (AVHRR)* was mounted on TIROS-N, launched in 1978, and later the AVHRR was also mounted on NOAA. The basic specifications of AVHRR have not changed, and the AVHRR has been continually utilized for earth observation. The current AVHRR has five bands, which are 0.58 to 0.68  $\mu$ m (band 1), 0.725 to 1.1  $\mu$ m (band 2), 3.55 to 3.93  $\mu$ m (band 3), 10.3 to 11.3  $\mu$ m (band 4), and 11.5 to 12.5  $\mu$ m (band 5). The AVHRR provides spatial resolution of approximately 1.1 km and a swath 2399 km wide. As NOAA orbits the globe 14 times daily, the AVHRR observes the same area twice a day. The wide-swath images measured by AVHRR are useful for various purposes, for example, for obtaining changes in cloud cover and land cover, vegetation index, and surface temperature. Consequently, these provide valuable information on surface properties and weather over regional scales.

## MODIS (EOS AM-1 and PM-1, NASA, USA)

The Moderate Resolution Imaging Spectroradiometer (MODIS) is an optical sensor mounted on EOS AM-1 and PM-1, launched in 1999 and 2002, respectively. The MODIS has 21 bands (0.4 to 3.0  $\mu$ m) and 15 bands (3.0 to 14.5  $\mu$ m). The spatial resolutions are 250 m (2 bands), 500 m (5 bands), and 1,000 m (29 bands). The MODIS can observe any site on the earth surface every 2 days. In the future, the wide-swath image measured by MODIS will be used effectively for earth observation like the AVHRR.

# SAR (JERS-1, Japan, ERS-1/2, EU, RADARSAT, Canada)

The *Synthetic Aperture Radar (SAR)* is an active sensor, which is used to observe physical properties, roughness, and inclination of the ground level using microwaves. The observation is almost independent of weather conditions, and it can be done through clouds. In particular, applications in vegetation classification of forests, grasslands, and agriculture areas have been studied using L (15 to 30 cm), C (3.75 to 7.5

cm), and X (2.4 to 3.75 cm) bands mounted on JERS-1, ERS-1/2, and RADARSAT, respectively.

Remote Sensors on Aircraft

Aircraft remote sensing has the advantages of flexible use and high spatial resolution in comparison with satellite remote sensing. For example, although the Landsat captures images of a region every 16 days, only several images are useful within a year; aircraft remote sensing can provide more images that could be used in analysis of rapid changes of seasons. Therefore, for development of agricultural remote sensing, it is necessary to use both satellite and aircraft remote sensing.

AVIRIS (JPL, NASA, USA)

The Airborne Visible Infrared Imaging Spectrometer (AVIRIS) is a hyperspectral image sensor developed by the Jet Propulsion Laboratory (JPL). The AVIRIS covers the range of 0.4 to 2.5  $\mu$ m in 224 contiguous bands approximately 10 nm wide and allows 20-m spatial resolution. The hyperspectral bands are expected for use in more exact classification of land cover and vegetation, diagnosis of crops, and identification of chemical compositions of soils.

CASI (ITRES Research Ltd, Canada)

The Compact Airborne Spectrographic Imager (CASI) is a hyperspectral sensor like the AVIRIS. Features of the CASI are a push-broom charge coupled device (CCD) with high spatial resolution (0.5 to 10 m) and an adjustable spectral range between 0.4 and 1.0  $\mu m$  of up to 288 programmable spectral bands at 1.9 nm intervals.

AISA (Spectral Imaging, USA)

The Airborne Imaging Spectroradiometer for Application (AISA) is a hyperspectral sensor with a 2-m spatial resolution, developed by Spectral Imaging Ltd. The wavelength range is 0.43 to 1.0  $\mu$ m and 512 spectral bands are the maximum. The AISA sensor head includes a fiber optic probe (FODIS) for real-time monitoring of downwelling solar irradiance to calculate the apparent reflectance of the earth's surface.

ADS40 (LH Systems, USA)

The Airborne Digital Sensor (ADS40) was developed by LH Systems and the German Aerospace Center. It has forward-, nadir-, and backward-looking linear CCD arrays to provide high spatial-resolution panchromatic images. The three line stereo images are captured simultaneously, and the triple overlap of the images provides high-quality DTM (digital terrain model). The ADS40 has further arrays for multispectral data (blue, B: 0.43 to 0.49  $\mu$ m, green, G: 0.535 to 0.585  $\mu$ m, red, R: 0.61 to 0.66  $\mu$ m and near-infrared, NIR: 0.835 to 0.885  $\mu$ m). The multispectral data can be used for applications in crop and land use analysis with a high spatial resolution of 20 cm.

ALTM (Optech, Canada)

The Airborne Laser Terrain Mapper (ALTM) is a series of airborne lidars (light detection and ranging), manufactured by Optech, for 3-D range measurement and terrain mapping. The ALTM 1025 has a scanning range finder with a range resolution of 1 cm and a range accuracy within 15 cm. The wavelength and beam divergence of the used laser (Nd:YAG) are 1,064 nm and 1.2 mrad (or 0.25 mrad). The frequencies of

pulse repetition and scan are 25,000 Hz and 25 Hz, respectively. The sampling data per one scan are 1,000 points. This system is expected to be useful in estimating the 3-D structure of woody canopies and vegetation biomass as well as terrain mapping.

In addition to those mentioned above, numerous remote sensors for aircraft, such as multi- and hyperspectral optical imagers, lidars, SARs, are produced by government agencies and private companies. The images measured by aircraft remote sensors require complicated rectification of aircraft tilt, shift, and flying height. Data of aircraft geometry from on-board GPS and *Inertial Measurement Units* (GPS/IMU) are used to rectify aircraft sensor images precisely.

# 5.2.3 Analysis of Land Use and Nitrogen Flow Using Data of Landsat and GIS Classification of Land Use with LANDSAT MSS

The eutrophication of lakes and rivers is closely related with increases in population and industrial and agricultural activities [13,14]. Therefore, relationships between eutrophication and land use change have been analyzed using Landsat and GIS data.

Figure 1 shows the basins of rivers flowing into two lakes and changes in land use in the basins from 1979 to 1990 [15]. The lakes lie 60 to 90 km northeast of Tokyo. Lake Kasumigaura is the second largest lake in Japan with a water area of 220 km² and a basin area of 2,135 km². The water depth is shallow, ranging from 0 to 7 m with an average of 4 m. In the lake, eutrophication is caused by influent nutrients such as nitrogen and phosphorous from the basins, and consequently algal blooms grow during the summer. As shown in Figure 1, the basin of each river was isolated using DEM (digital elevation model). The land cover was classified into five categories (urban area, paddy field, forest area, other crop land, and water area) using LANDSAT MSS data observed in 1979, 1984, and 1990. The urban area has rapidly increased along roads around railway stations and big cities, and consequently other areas have decreased.

# Analysis of Nitrogen Flow from Different Land Use Areas to Lakes

The inflow of nitrogen from the basin to the lake is a major factor in eutrophication. Therefore, the nitrogen flow from different land use area to the lake is analyzed by the following.

Annual nitrogen flow  $(L_i, kg/y)$  from a basin area (i) of one river to the lake is given by:

$$L_i = 10^{-3} C_i \cdot Q_i \tag{1}$$

where  $C_i$  (mg/l) is the annual mean value of total nitrogen concentration in the river near the mouth, and  $Q_i$  (m<sup>3</sup>/y) is the annual water flow of the river.

A ratio of  $L_i$  to the total area  $(A_i, \text{km}^2)$  of the basin i is expressed by:

$$\frac{L_i}{A_i} = \sum_{j=1}^{4} (U_j \cdot r_{ij})$$
 (2)

where j (= 1 to 4) is for each category of urban area, paddy field, forest area, or other crop land;  $U_j$  (kg·km<sup>-2</sup>·y<sup>-1</sup>) is the coefficient called the runoff load factor of each category j; and  $r_{ij}$  is the ratio of area ( $A_{ij}$ , km<sup>2</sup>) of each category j to  $A_i$  in the basin i.

The  $C_i$  was calculated from data measured at ten large rivers.  $Q_i$  was estimated from annual precipitation and run-off ratios of each river basin.  $A_i$  and  $r_{ij}$  were calculated from results in Figure 1. Because  $U_j$  for only four categories are unknown, these were estimated by minimizing E in the following equation.

$$E = \sum_{i} \left( \frac{L_i}{A_i} - \sum_{j=1}^{4} (U_j \cdot r_{ij}) \right)^2$$

Figure 2 shows estimated runoff load factors  $(U_j)$  of total nitrogen for four categories. The runoff load factors for forest area and paddy field were almost zero. However, those for the urban area and other crop land showed large values. The decrease in

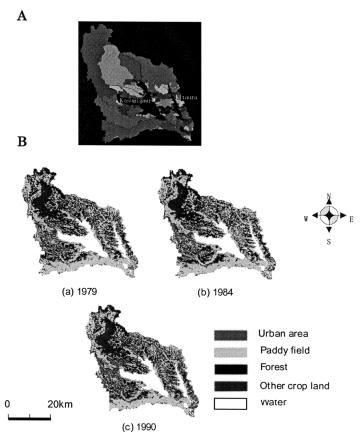


Figure 1. Basins of rivers flowing into lakes Kasumigaura and Kitaura in Japan (A) and changes in land use in the basins (B). Each area in (A) represents the isolated basin of each river [15].

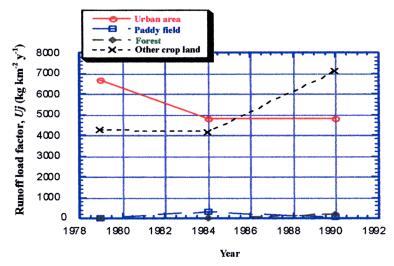


Figure 2. Differences in runoff load factor of total nitrogen among urban areas, paddy fields, forests, and other crop fields [15].

urban area runoff from 1979 to 1984 may be due to the spread of sewers. The increase in other crop field runoff from 1984 to 1990 may be caused by expansion of pig farming. The difference in precipitation may also have an affect on the runoff load. It is concluded that the extension of urban areas and crops fields other than paddy fields increases the eutrophication of Lake Kasumigaura through the rise of nitrogen flow to the lake.

## 5.2.4 Estimation of Wheat Growth Using Aerial Hyperspatial Data

Oversupply of fertilizers and agrichemicals in agriculture causes pollution of soil and water. Precision farming is a concept for increasing plant production and optimizing use of water, fertilizers, and agrichemicals [16-18]. The spatial distributions of these materials in farmland are not uniform. Therefore, hierarchical remote sensing, which joins image sensing (described in Section 5.1) with remote sensing, is expected to be one of the important tools for precision farming.

Figure 3 shows a normalized differential vegetation index (NDVI) image calculated from aerial hyperspatial images measured by the above-mentioned ADS40 (LH Systems, USA). The wheat field is the same place as that shown in Figure 3 of Section 5.1. In each test area (8  $\times$  8  $m^2$ ) of the field, nitrogen fertilizer at 0.0, 0.1, or 0.2 kg/m² was applied before seeding in the previous year. The multispectral images of four bands (R, G, B and NIR) with a spatial resolution of about 20 cm were measured in the following spring during the growing period of wheat crops. The NDVI was calculated using R (0.610 to 0.660  $\mu$ m) and NIR (0.835 to 0.885  $\mu$ m) images by the following equation:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{3}$$

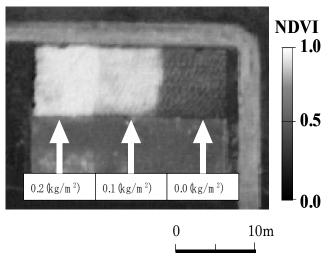


Figure 3. NDVI image of wheat field with applications of different amounts of nitrogen fertilizer.

The NDVI is a multipurpose index of contents of plant pigments, especially chlorophyll *a*, relating to photosynthesis, net primary production (NPP), and leaf area index (LAI) [1,19-26]. It ranges between –1 and 1. The increase of chlorophyll *a*, NPP, and LAI during the adult phase before the reproductive phase lowers red-light reflectance against high near-infrared reflectance, and consequently, the NDVI increases.

As shown in Figure 3, two test areas with the fertilizer application had high growth rates and high values of NDVI. However, the value of the right area without the fertilizer application was as low as that of grass area under the test area. The soil area of the left side and the asphalt road of the upper and right borders showed lower values than other areas. The road edge between the test area and the asphalt road showed values between the grass area and the test area with high growth rate. Although wheat crops of the middle area grew normally, those of the left area with the highest value of NDVI had a low tolerance to wind. This result suggests noteworthy points in use of NDVI.

# 5.2.5 Analysis of Farmland Using Aerial Hyperspectral Data

Hyperspectral data are useful for analyzing spectral properties of various crops and other land covers such as soils in farmland [27-31]. Figure 4 shows spectral properties of four areas covered by three crops (watermelon, maize, and taro) and loamy soil in farmland in the Miura Peninsula, about 50 km south of Tokyo. The spectral properties were obtained from aerial hyperspectral images, of 70 bands in a range of 0.43 to 0.90 µm and 2 m spatial resolution, measured by the AISA (Spectral Imaging, USA). The Miura Peninsula is a typical area of suburban agriculture in Japan. In the farmland, which is divided into small fields, many vegetable crops such as watermelon, maize, taro, cabbage, pumpkin, and Japanese radish are cultured year-around because of warm weather conditions.

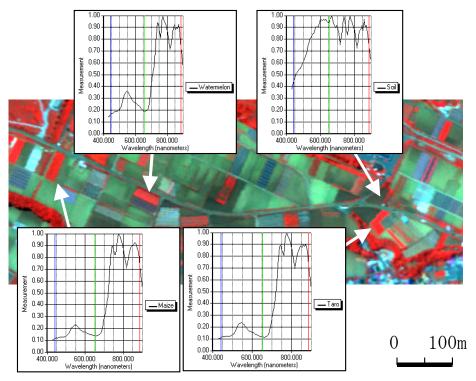


Figure 4. Spectral properties of three crop fields (watermelon, maize, and taro plants) and loamy soil in farmland, obtained from aerial hyperspectral images.

In Figure 4, the spectral properties of crops represent large absorptions caused by photosynthetic pigments in the visible range: a steep increase in the red-edge region from 0.69 to 0.74  $\mu$ m continuing high to 0.9  $\mu$ m, similar to the leaf reflectance spectra shown in Figure 1 of Section 5.1. The properties of loamy soil also are similar to those in Figure 1 and these are clearly different from those of vegetable crops. However, there are several drops in the high reflectance region beyond 0.74  $\mu$ m. These are caused by O<sub>2</sub> (at 0.76  $\mu$ m) in the atmosphere; by H<sub>2</sub>0 (at 0.96  $\mu$ m) in the atmosphere, crops and soils; and by other factors. The analysis of spectral properties in the visible to near-infrared region may provide more useful information on contents of pigments, water, minerals, and nutrients, as well as crop species, production, and growth conditions. Therefore, hyperspectral remote sensing is expected to be a key tool for precision farming.

## 5.2.6 3-D Remote Sensing of Terrain and Forests Using Aerial Lidar Data

Aerial scanning lidar is an emerging *active* remote sensing technology for direct 3-D terrestrial observation [7,9-11]. Figure 5 shows 3-D images of terrain (ground surface) and a woody canopy obtained by a helicopter-borne scanning lidar (ALTM 1025).

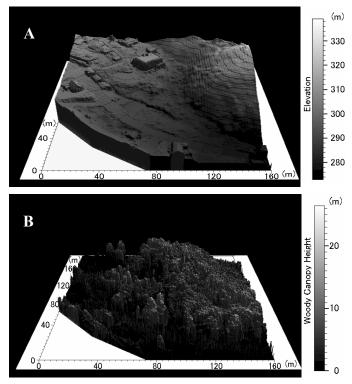


Figure 5. 3-D images of terrain (A, ground surface) and woody canopy (B) obtained by a helicopter-borne scanning lidar with high spatial resolution [9].

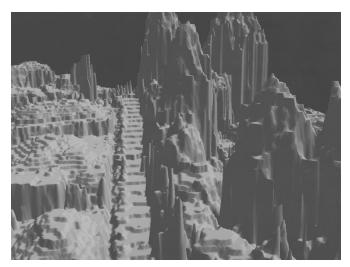


Figure 6. 3-D close-up view of first pulse mode DEM of a place in Figure 5 [9].

special model, Optech Co., Canada and Aero Asahi Co., Japan) with a high spatial resolution (about 33 cm mesh). The terrain of a valley was estimated with an accuracy of about 15 cm by interpolation of DEM data measured using a last-pulse mode, measuring the longest elapsed time between the emitted pulse of laser and the returned pulse. The woody canopy height was calculated by subtracting the terrain from DEM data measured using a first-pulse mode measuring the shortest elapsed time. The laser-derived tree heights were in error by less than 47 cm (RMSE = 19 cm) for coniferous trees and 40 cm (RMSE = 12 cm) for broadleaf trees.

Figure 6 shows a 3-D close-up view of first pulse mode DEM of a place in Figure 5. In left area of the place, the land is been turned into housing lots. There are several trees on the right. A straight narrow path with a flight of stairs is in the center. These stairs in the image were caused by digitization of the image (quantization error = about 15 cm).

We are studying farmland and public garden planning, forest management, carbon stock (biomass) estimation of forests by using this system and portable scanning lidars in addition to passive 2-D remote sensing data [9,11,32]. In the near future, these systems will be used effectively to assess forest development relating to forestation, agroforestry, and water resource management.

#### References

- 1. Myers, V. I. 1983. Remote sensing applications in agriculture. *Manual of Remote Sensing*, 2<sup>nd</sup> ed., Vol.III, ed. R. N. Colwell, 2111-2228. VA: Amer. Soc. Photogrammetry.
- 2. Steve, M. D., and J. A. Clark, eds. 1990. *Applications of Remote Sensing in Agriculture*. London, UK: Butterworths.
- 3. Owe, M., and G. D'Urso, eds. 2002. *Remote Sensing for Agriculture, Ecosystems, and Hydrology III*. Bellingham, WA: SPIE.
- 4. Hobbs, R. J., and H. A. Mooney, eds. 1990. *Remote Sensing of Biosphere Functioning*. New York, NY: Springer.
- 5. Rencz, A. N. 1999. *Remote Sensing for the Earth Sciences*. New York, NY: John Wiley and Sons.
- 6. Campbell, B. J. 1996. *Introduction to Remote Sensing*. New York, NY: The Gulford Press.
- 7. Flood, M., and B. Gutelious. 1997. Commercial implications of topographic terrain mapping using scanning airborne laser radar. *Photogramm. Eng. Remote Sens.* 63: 327-366.
- 8. Henderson, F. M., and A. J. Lewis, eds. 1998. *Principles and Applications of Imaging Radar*. New York, NY: John Wiley and Sons.
- 9. Omasa, K., Y. Akiyama, Y. Ishigami, and K. Yoshimi. 2000. 3-D remote sensing of woody canopy heights using a scanning helicopter-borne lidar system with high resolution. *J. Remote Sens. Soc. Jpn.* 20: 394-406.
- 10. Lefsky, M. A., W. B. Cohen, G. G. Parker, and D. J. Harding. 2002. Lidar remote sensing for ecosystem studies. *BioSci*. 52: 19-30.

- 11. Omasa, K., G. Y. Qiu, K. Watanuki, K. Yoshimi, and Y. Akiyama. 2003. Accurate estimation of forest carbon stocks by 3-D remote sensing of individual trees. *Environ. Sci and Tech.* 37: 1198-1201.
- Dubayah, R., J. B. Blair, J. L. Bufton, D. B. Clark, J. Jaja, R. Knox, S. B. Luthcke, S. B. Prince, and J. Weishampel. 1997. The vegetation canopy lidar mission. *Proceedings of Land Satellite Information in the Next Decade, II: Sources and Applications*, 100-112. Bethesda, MD: Am. Soc. Photogramm. Eng. Remote Sens.
- 13. Krenkel, P. A., and V. Novotny 1980. *Water Quality Management*. London, UK: Academic Press.
- 14. Lewis Jr., W. M., J. F. Saunders III, D. W. Crumpacker Sr, and C. Brendecke. 1984. *Eutrophication and Land Use*. New York, NY: Springer-Verlag.
- 15. Oki, K., Y. Yasuoka, and M. Inamura. 1998. Estimation of annual total nitrogen load to lake from basin with remote sensing. *Proceedings the 19<sup>th</sup> Asian Conference on Remote Sensing, Philippines*, Q61-66. http://www.gisdevelopment.net/aars/acrs/1998/ps2/ps2006.shtml.
- 16. Hegg, R. 1999. Sustainable and environmental engineering. *CIGR Handbook of Agricultural Engineering*, Vol.III, eds. B. A. Stout, and B. Cheze, 585-598. St. Joseph, MI: ASAE.
- 17. Auernhammer, H., and J. K. Schueller. 1999. Precision farming. *CIGR Handbook of Agricultural Engineering*, Vol.III, eds. B. A. Stout and B. Cheze, 598-616. St. Joseph, MI: ASAE.
- 18. De Baerdemaeker, J., A. Munack, H. Ramon, and H. Speckmann. 2001. Mechatronic systems, communication, and control in precision agriculture. *IEEE Contr. Syst. Mag.* 21: 48-70.
- 19. Tucker, C. J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8: 127-150.
- 20. Jackson, R. D., P. N. Slater, and P. J. Pinter. 1983. Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres. *Remote Sens. Environ.* 15: 187-208.
- 21. Hinzman, L. D., M. E. Bauer, and C. S. T. Daughtry. 1986. Effects of nitrogen fertilization on growth and reflectance characteristics of winter wheat. *Remote Sens. Environ.* 19: 47-61.
- Tucker, C. J., W. W. Newcomb, S. O. Los, and S. D. Prince. 1991. Mean and inter-year variation of growing-season normalized difference vegetation index for the Sahel 1981-1989. *Int. J. Remote Sens.* 12: 1113-1115.
- 23. Fernandez, S., D. Vidal, and E. Simon. 1994. Radiometric characteristics of *Triticum aestivum* cv. Astral under water and nitrogen stress. *Int. J. Remote Sens.* 15: 1867-1884.
- 24. Reddy, G. S., C. L. N. Rao, L. Venkataratnam, and P. V. K. Rao. 2001. Influence of plant pigments on spectral reflectance of maize, groundnut and soybean grown in semi-arid environments. *Int. J. Remote Sens.* 22: 3373-3380.
- 25. Singh, R., D. P. Semwal, A. Rai, and R. S. Chhikara. 2002. Small area estimation of crop yield using remote sensing satellite data. *Int. J. Remote Sens.* 23: 49-56.

- 26. Silleos, N., K. Perakis, and G. Petsanis. 2002. Assessment of crop damage using space remote sensing and GIS. *Int. J. Remote Sens.* 23: 417-427.
- 27. Demetriades-Shah, T. H., M. D. Steven, and J. A. Clark. 1990. High resolution derivative spectra in remote sensing. *Remote Sens. Environ*. 33: 55-64.
- 28. Malthus, T. J., B. Andriew, F. M. Danson, K. W. Jaggard, and M. D. Steven. 1993. Candidate high spectral resolution infrared indices for crop cover. *Remote Sens. Environ.* 46: 204-212.
- 29. Carter, G. A. 1994. Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *Int. J. Remote Sens.* 15: 697-703.
- 30. Cochrane, M. A. 2000. Using vegetation reflectance variability for species level classification of hyperspectral data. *Int. J. Remote Sens.* 21: 1113-1115.
- 31. Kokaly, R. F. 2001. Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration. *Remote Sens. Environ.* 75: 153-161.
- 32. Omasa, K., Y. Urano, H. Oguma, and Y. Fujinuma. 2002. Mapping of tree position of *Larix leptolepis* woods and estimation of diameter at breast height (DBH) and biomass of the trees using range data measured by a portable scanning lidar. *J. Remote Sens. Soc, Jpn.* 22: 550-557.

# 5.3 Speaking Plant/Speaking Fruit Approaches

Y. Hashimoto, T. Morimoto, and J. De Baerdemaeker

Abstract. A skilled grower can deal well with various plants and yield good products using intuition and visual information, in those ways communicating with the plants. To optimize plant production scientifically, it is important to measure plants' physiological responses using sensors and then use that information for control. Such an approach is called the speaking plant approach (SPA), where environmental factors are the input variables and the plant responses are the output variables. The term speaking plant (SP) refers to plant responses that are measured using sensors. This approach is also applicable to storage processes and is called a speaking fruit approach (SFA). Measurement, identification, and optimization of plant (or fruit) responses, as affected by environmental factors, are important tasks for the SPA (or SFA). It is, however, difficult to sufficiently address these tasks because plant responses are quite complex and uncertain, but intelligent control techniques can facilitate such tasks. This section introduces SPA and SFA concepts and then presents an intelligent control technique for realizing the SPA and SFA. Finally, applications of the SP- or SF-based intelligent control technique for optimization of cultivating and storage processes are provided, aiming at qualitative improvement of plants and fruits.

**Keywords.** SPA, SFA, Sensors, Plant responses, Fruit responses, Measurement, Identification, Optimization, Environmental control, Speaking plant, Speaking fruit.

#### 5.3.1 Introduction

Increasing consumer wealth comes with increasing consumer demand for better quality agricultural products, often coupled with a desire to avoid products that use any chemicals. In order to yield better quality plants for this market, advanced control techniques have been applied to the greenhouse environment [1-3]. However, because the studied control variables have been only environmental factors, the optimization of plant growth has not been realized. This is probably due to the complexity of the physiological responses.

In order to optimize plant growth, it is efficient to measure the physiological responses of the plant and determine the current physiological status, and then use the information for environmental control for optimization. A grower should control the environment optimally, communicating with the plants by sensors. Such an approach is called the *speaking plant approach* (SPA) [4-8], where the environmental factors are defined as the input and the plant responses as the output. The term *speaking plant* (SP) refers to plant responses measured by sensors.

The concept of the *speaking fruit approach (SFA)* for optimizing the storage process of fruit has also been of interest [9,10]. At present, the environmental factors during storage are usually maintained constant. A constant low temperature is generally used, probably because the storage process is considered a static system. Under this condition, however, it would be difficult to realize any qualitative improvement of the fruit. Recently, researchers have demonstrated that a high temperature during the first few days after storage inhibited the ripening and ethylene production of fruit [11,12]. These results suggest that the storage process should be defined as a dynamic system and that a flexible control is useful for improving the fruit quality.

This section introduces the concepts of the SPA for the cultivation process and the SFA for the fruit storage process, and applies them to optimizing plant growth (in hydroponic cultivation) and fruit ripening during storage.

## 5.3.2 A Computer Control System Based on the SPA and the SFA

Figure 1 is a diagram of a computer control system for a total production system, consisting of cultivation systems for plants and storage systems for fruits, for realizing the concepts of the SPA and the SFA. The physiological responses of the plant during cultivation and the fruit during storage are optimally controlled by environmental factors based on these concepts. In the cultivation process, plant responses such as net photosynthetic rate and transpiration rate are measured by sensors in order to determine the current physiological status of the intact plant during cultivation. In the storage process, fruit responses such as water loss, respiration rate, and skin color are measured to understand the current physiological status of the fruit. Based on these plant and fruit responses, environmental factors in the greenhouse and storage house are optimally controlled, aiming at their qualitative improvement. Thus, the measurement of both plant responses and environmental factors and the optimization of plant responses are the main tasks for the SPA.

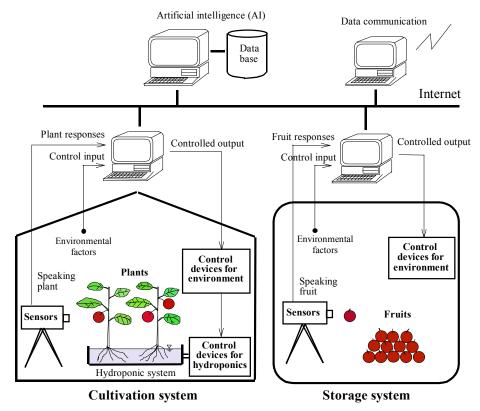


Figure 1. A decentralized computer control system for a total plant production system, consisting of a cultivation system and a storage system, based on the concepts of the SPA and the SFA.

#### Measurement Techniques

The first step to the SPA is to measure the plant responses of the intact whole plant using several types of sensors [13,14]. Fusing various types of information on plant response allows a correct estimation of the present physiological status of the plant. Non-destructive (or non-contact) sensing is also efficient for on-line control. From these findings, an image processing system is used for measuring plant growth [15,16], an infrared thermometer for the leaf temperature [4,14], an infrared CO<sub>2</sub> analyzer for the photosynthetic rate [5,13], surface electrodes for the bioelectric potential of the leaf [17], a stem capacitance meter for the water content in the stem, and ion-selective sensors for the nutrient uptake by the roots [18].

Figure 2 shows the fundamental plant responses as affected by irradiation, measured in a strictly controlled environment. When the plant was exposed to light, the leaf temperature at first rose because of sudden changes in the radiant energy, but later began to fall dramatically with increasing transpiration. The net photosynthetic rate began to increase and then reached a maximum value. During the same period, the

stem capacitance continued to decrease. After that time, however, both the transpiration and the net photosynthetic rate were severely suppressed due to the stomatal closure caused by water stress, which arose from excessive transpiration [19]. During the same period, therefore, the leaf temperature became much higher and the stem capacitance also increased. Later, both the transpiration and the net photosynthetic rate again increased and then reached their steady states. At the same time, both the leaf temperature and the stem capacitance gradually decreased and then reached steady levels. This is clearly because the stomata again opened with the improvement of water content in the plants. Thus, this shows that various plant responses are closely related to each other.

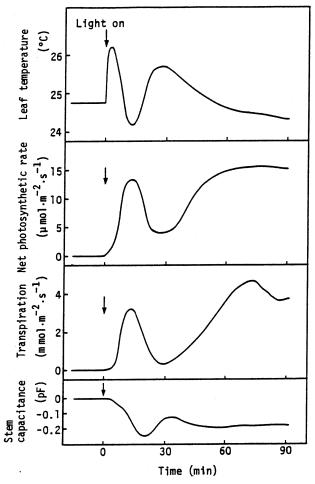


Figure 2. Fundamental plant responses (leaf temperature, net photosynthetic rate, transpiration, and stem capacitance) of the tomato plant as affected by irradiation (450  $\mu$ Em-2s-1), measured in a strictly controlled environment (25  $\pm$  0.1 °C and 60  $\pm$  2% RH).

## An Optimization Technique

De Baerdemaeker and Hashimoto [9] presented the introductory concepts of the SPA and the SFA at the Milano Congress of CIGR. Morimoto et al. [25] proposed a speaking plant- (or speaking fruit-) based control system for realizing the optimization of plant (or fruit) responses. The system (Figure 3) consists of a decision system and a feedback control system. The former determines the optimal set point trajectory of the environment (= optimal value), and the latter controls the environment on the basis of the optimal set points determined by the decision system [25,26]. In the decision system, the present plant responses as affected by environmental factors are first measured (Process A) and then identified using an identification technique (Process B). Furthermore, the optimal *l*-step set points of the environment that maximize the plant response (concerning quality) are searched for by simulation of the identified model using an optimization technique (Process C). Finally, the optimal set points are sequentially applied to the set point in the feedback control system (Process D). It will be shown that if these two procedures, identification and the search for an optimal value, are periodically repeated to adapt the time-variation of the cultivation or storage system, then two types of control performances, optimization and adaptation, can be satisfied.

Since plant (or fruit) control systems are usually quite complex and uncertain, intelligent approaches are useful for optimizing such systems [27,28]. Neural networks have the capability of identifying (learning) unknown nonlinear properties by their own learning ability [29]. Hirafuji [30] and Morimoto et al. [31] used neural networks to identify the physiological responses of the plant. Seginer and McClenden [32] used neural networks to determine the optimal temperature set point for greenhouse lettuce crops. Genetic algorithms can be used to find an optimal value (or at least a near-

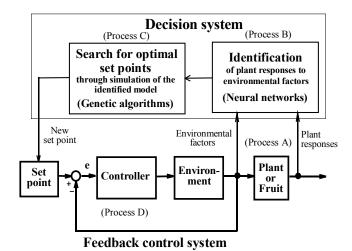


Figure 3. An SP- (or SF-) based control system, consisting of a decision system and a feedback control system for realizing the optimization of the cultivation (or storage) process.

global one) of a complex objective function, with a multi-point search procedure, by simulating the biological evolutionary process based on crossover and mutation in genetics [33]. In the decision system, therefore, genetic algorithms are used in the search for the optimal set points (Process C).

Thus, incorporating the concept of SPA (or SFA) with intelligent approaches allows the optimization of the plant production process to be successfully realized.

## 5.3.3 Application of the SPA for Optimization of a Hydroponic System

#### Optimization Problem I

Tomato plants (*Lycopersicon esculentum* Mill. cv. Momotaro) grown in a deep hydroponic culture were used for this experiment. The technique behind good fruit yields in tomato cultivation is to keep an optimal balance between vegetative growth (root, stem, and leaf growth) and reproductive growth (flower and fruit growth). In hydroponics, however, vegetative growth becomes more active than reproductive growth because the roots always exist in a suitable environment for ion uptakes. Active vegetative growth induces poor reproductive growth.

It is known that the balance between the two types of growth is often determined at the seedling stage. During the seedling stage, only stem growth, leaf growth, and root growth are visible. Luo and Kato [34] demonstrated that the S/R ratio (S = stem dry weight, R = root dry weight) is a good indicator for predicting future growth and that smaller values result in better yields. Based on our experiments, it was also found that larger stem growth resulted in poor flowering. However, in this case, because measuring leaf growth is much easier than measuring root growth, leaf growth was used as one of the predictors. It was also assumed that larger leaf growth is advantageous for promoting photosynthetic production of the plant because the area performing photosynthesis increases. From these findings, the ratio (TLL/SD) of total leaf length (TLL) to stem diameter (SD) was defined as a predictor for future plant growth, and the control was restricted to the seedling stage. Controls for maximizing TLL/SD may be of value only during the seedling stage.

Here, the control input is the nutrient concentration of the solution in hydroponics. Most skilled growers usually increase the nutrient concentration as the plants grow.

Let TLL(k)/SD(k) be a time series of TLL/SD, as affected by nutrient concentration NC(k) (k = 1, ..., N = sampling day, N = final day). The light intensity L(k) is also used as an input, but it is always kept constant in identifying the TLL(k)/SD(k).

For implementation, the control process (seedling stage:  $1 \le k \le N$ ) was divided into four steps: step 1, transplanting  $(1 \sim N_{1L})$ ; step 2, vegetative growth after transplanting  $(N_{1L}+1 \sim N_{2L})$ ; step 3, flowering of the first truss  $(N_{2L}+1 \sim N_{3L})$ ; and step 4, fruit setting for the first truss and flowering for the second truss  $(N_{3L}+1 \sim N)$ . An objective function was given by the average value of TLL/SD at the last step:

$$F_1(NC) = \sum_{k=N_{2r}+1}^{N} \frac{\frac{TLL(k)}{SD(k)}}{N - N_{3L} + 1}$$
(1)

The optimization problem here is to determine the optimal four-step set points of nutrient concentration,  $NC_1$ ,  $NC_2$ ,  $NC_3$ , and  $NC_4$ , that maximize  $F_1(NC)$ . The nutrient concentration was constrained to  $0.2 \le NC(k) \le 2.0$  (mS/cm) through preliminary experiments.

maximize 
$$F_1(NC)$$
 (2)  
subject to  $0.2 \le NC(k) \le 2.0 \text{ (mS/cm)}$ 

### An SP-Based Intelligent Control System for Optimization

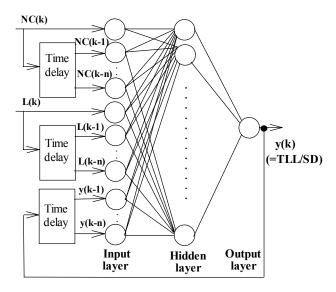
A speaking plant-based control system for realizing this optimization is shown in Figure 3. In the decision system, a neural network was first used for identifying the TLL/SD ratio to the nutrient concentration of the solution, and then a genetic algorithm was used for finding the optimal four-step set points of the nutrient concentration that maximize the objective function (1).

Figure 4 shows the decision system. Part (a) is a time-delay neural network used for identification (Process B of Figure 3). It consists of three layers. Cybenko [35] showed that a three-layer neural network with one hidden layer allowed any continuous function to be successfully identified. For dynamic identification, arbitrary feedback loops that produce time histories of the data are necessary elements of the network.

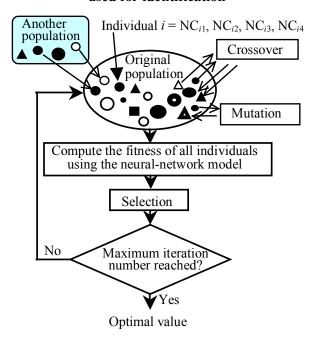
The current output y(k) = TLL(k)/SD(k) is estimated from both the historical input data  $[NC(k), ..., NC(k-n_1), L(k), ..., L(k-n_1)]$  and from the historical output data  $[y(k-1), ..., y(k-n_1)]$  (where  $n_1$  = number of system parameters, NC(k) = nutrient concentration, and L(k) = light intensity) [36]. The learning method was error back-propagation [37]. Figure 4(b) is the flow chart of the genetic algorithm used for determining the optimal value (Process C of Figure 3). Because the aim of this optimization is to determine the four-step set points of the nutrient concentrations, an individual is defined as the four-step set points of the nutrient concentration and each set point is coded as a six-bit binary string. A set of individuals is called a "population." The simple bound constraint was determined to be  $0.2 \le NC_l \le 2.0$  (mS/cm).

Individual 
$$i = NC_{i1}$$
,  $NC_{i2}$ ,  $NC_{i3}$ ,  $NC_{i4} = 100100$ , 001001, 001100, 101010

Fitness is given by Equation 1, which is the same as an objective function. An individual having the maximum fitness is regarded as an optimal value. The procedure is as follows. (1) An initial population consisting of several individuals is generated at random. (2) New individuals are added to the original population from another population to maintain diversity of the original population. (3) Crossover and mutation operators are applied to the individuals selected at random. Numbers of the crossover and mutation depend on the crossover rate and mutation rate, respectively. (4) The fitness values of all individuals are calculated using Equation 1, and their performances are evaluated. It is noted that the value of Equation 1 is obtained from simulation of the neural-network model. (5) Superior individuals are selected and retained for the next generation (*selection*) based on the elitist strategy. (6) Steps 2 to 5 are repeated until an optimal value can be obtained. An optimal value is given by an individual with maximum fitness.



# (a) A time-delay neural-network used for identification



## (b) Flow chart of the genetic algorithm

Figure 4. A decision system (Figure 3) consisting of (a) a neural network and (b) a genetic algorithm.

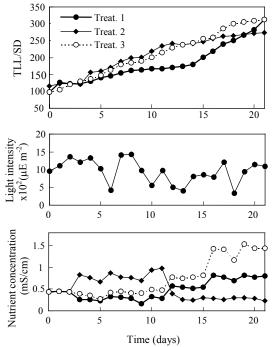


Figure 5. Observed daily changes in the TLL/SD (total leaf length/stem diameter) ratio of tomato plants grown in a deep hydroponic system, and the light intensity and nutrient concentration of the solution during the seedling stage.

#### Identification of the TLL/SD Ratio of Nutrient Concentration

First, the data for identification was obtained (Process A of Figure 3). Figure 5 shows the daily changes in the TLL/SD ratio of tomato plants, as well as the light intensity and nutrient concentration of the solution during the seedling stage. Three patterns of the TLL/SD ratio under three different nutrient concentration treatments are shown. Purwanto et al. [38] found that three or more data sets are necessary for dynamic identification. The light condition was arbitrary; only the nutrient concentration was changed to identify the TLL/SD response. It was found that the value of the TLL/SD ratio is markedly affected by nutrient concentration. Next, the TLL/SD ratio to the nutrient concentration was identified using the neural network (Process B). The system parameter number  $n_1$  and the hidden neuron number  $n_2$  of the neural network were determined through cross-validation to be 1 and 5. Through these procedures, a computational model could be obtained for predicting the behavior of the TLL/SD ratio under any combination of the four-step set points.

# Search for Optimal Four-Step Set Points (Process C)

Next, the optimal four-step set points of the nutrient concentration which maximize the objective function were searched for through simulation of the identified neural-network model using the genetic algorithm (Process C). Here the nutrient concentration was limited to the range 0.2 to 2.0 mS/cm. The optimal four-step set points ob-

tained here were a slightly higher level 1.4 mS/cm in the first step, a markedly lower level (0.3) in the second step, a slightly higher level (1.6) in the third step, and the maximum level (2.0) in the fourth step.

In hydroponics, because the roots of the plants always exist in a suitable environment for the uptake of any ions, vegetative growth is easily promoted. Active vegetative growth during the seedling stage will result in poor reproductive growth in the future [39]. Therefore, vegetative growth must be suppressed at the early seedling stage, before the flowering of the first truss. The low nutrient concentration in the second step seems to be effective in suppressing the excessive vegetative growth during the seedling stage. The high nutrient concentrations in the third and fourth steps are useful in accelerating reproductive growth.

## Optimal Control Performance (Process D)

Figure 6 shows the actual control performance of the TLL/SD ratio. The solid line shows the optimal control performance and the dotted line represents conventional

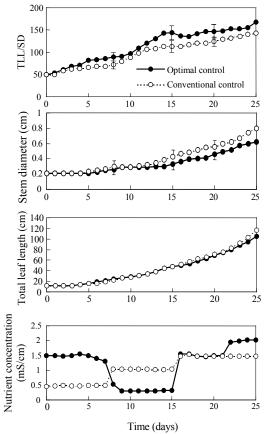


Figure 6. Actual optimal control performance of *TLL(k)/SD(k)* as affected by optimal four-step set points of nutrient concentration.

control performance. The conventional strategy is simply to increase the nutrient concentration in a stepwise fashion with the growth of the plants. A *t*-test was carried out to check the differences in growth. Comparing both control performances, the values of the TLL/SD ratio are 10%-15% higher with the optimal control than with the conventional control. This is because with the optimal control stem growth was significantly suppressed by the low nutrient concentration at the second step, while in both cases the leaf growth is not very different. Thus, the effectiveness of this control technique was confirmed experimentally.

# 5.3.4 Application of the SFA to the Fruit Storage Process

## Optimization Problem II

Constant and low temperatures (2-8°C) are normal for fruit storage. In recent years, however, it has been suggested that a flexible control of the storage environment is effective in improving the quality of fruit [40]. Researchers have demonstrated that the exposure of fruit to high temperatures (35-40°C) for several hours inhibits ethylene production and delays the ripening of fruit [11,12]. The delay of ripening is probably due to heat shock proteins (HSPs). It is known that the exposure of living organisms to heat stress generates several types of heat shock proteins to resist such heat stress [41,42]. It seems that the effective use of stress responses leads to the improvement of fruit quality.

Tomatoes (*Lycopersicon esculentum* Mill. cv. Momotaro) were used for the experiment. The aim of this study was to obtain the optimal operation of the heat stress application that minimizes the ripening of the tomato.

Let C(k) (k = 1, 2, ..., N) be time series of the color change, as affected by temperature T(k), at the time k, which is characterized by a cumulative response. An objective function,  $F_2(T)$ , was given by two evaluation factors: the amplitude of color change and its change rate.

$$F_2(T) = \alpha \cdot \sum_{k=N-3}^{N} C(k) + \beta \cdot \sum_{k=N-3}^{N} \{ C(k) - C(k-1) \}$$
 (3)

The control process was divided into eight steps. The optimization problem here is to determine the eight-step set points of temperature that minimize the objective function. The constraint of temperature was  $5 \le T(k) \le 35$ °C. The ripening of tomato was estimated from the skin color using a colorimeter (Minolta, CR-200b). A hue angle in the L.C.H. method was used for evaluating the skin color [43].

## An SF-Based Intelligent Control System for Optimization

A speaking fruit (SF)-based intelligent control system is much the same as the SP-based one. The neural network is first used for identifying the color change of the tomato fruits, as affected by temperature, and then the genetic algorithm is used for searching for the eight-step set points of temperature that minimize the objective function (Equation 3) through simulation of the identified model.

A time-delay neural network was used for the identification of the color change to the temperature. On the other hand, an individual used in the genetic algorithm application was given by the eight-step set points of the temperature ( $T_1$ ,  $T_2$ ,  $T_3$ ,  $T_4$ ,  $T_5$ ,  $T_6$ ,  $T_7$ , and  $T_8$ ). They were coded as a six-bit binary string. Here, the simple bound constraints were  $5 \le T_1 \le 35$ °C.

# Identification of the Color Change Response to Temperature (Processes A and B)

Figure 7 shows eight types of cumulative responses of the color change of tomatoes under different temperature operations during storage. In all cases, initial temperatures were 10°C. Treatments 1 and 2 include the heat stress of 35°C during the first day and the first two days, respectively, and then cooled to 5°C during the latter stages in both cases. Treatments 3, 5, and 6 are cases of constant temperatures at 5, 15, and 25°C, respectively. These patterns were determined so that the time history of the input satisfies the need for adequate change for better identification.

Next, these data were identified using a neural network. Six data sets (Treatments 1, 2, 3, 5, 6, and 7) were used for identification (= training) and two data sets (Treatments 4 and 8) for the validation of the model. The numbers of the system parameter and the hidden neurons were determined to be 1 and 4, respectively, through cross-validation.

#### Search for Optimal Eight-Step Set Points (Process C)

Next, the optimal eight-step set points of the temperature (= optimal heat treatment) were obtained through simulation of the identified neural network model using the genetic algorithm. An optimal value depended on the initial condition of the fruit and

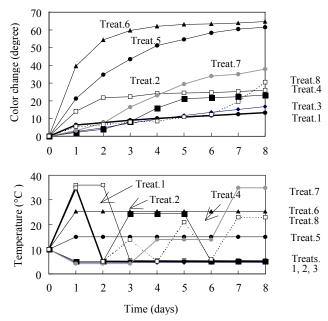


Figure 7. Observed cumulative responses of the color change of tomatoes as affected by temperature. A larger value of the color change means redder.

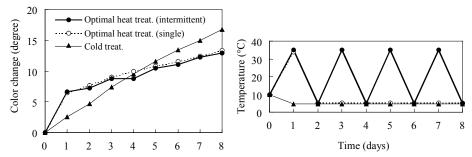


Figure 8. Actual optimal control performances of the color change of tomato fruit as affected by two types of optimal values (single and intermittent heat treatments).

The case of a normal cold treatment (5°C) is shown for comparison.

#### Optimal Control Performance (Process D)

Finally, the two types of optimal heat treatments were applied to a real system. Figure 8 shows the optimal control performance of the color change. It is clear that the color changes under the two optimal heat treatments are smaller than that under the cold treatment. There was no significant difference in the color change between the two optimal heat treatments.

Thus, it was shown that the speaking plant- and speaking fruit-based intelligent control techniques are suitable for the optimization of a total plant production system, which suggests one of the most important themes in precision agriculture.

#### References

- 1. Chalabi, Z. S., B. J. Bailey, and D. J. Wilkinson. 1996. A real-time optimal control algorithm for greenhouse heating. *Comput. Electron. Agric.* 15: 1-13.
- 2. Sigrimis, N., and N. Rerras. 1996. A linear model for greenhouse control. *Trans. ASAE* 39: 253-261.
- 3. Ioslovich, I., and I. Seginer. 1998. Approximate seasonal optimization of the greenhouse environment for a multi-state-variable tomato model. *Trans. ASAE* 41: 1139-1149.
- 4. Hashimoto, Y., T. Morimoto, and S. Funada. 1981. Computer processing of speaking plant for climate control and computer aided cultivation. *Acta Hortic*. 115: 317-325.
- 5. Hashimoto, Y., and T. Morimoto. 1985. Some speaking plant approach to the synthesis of control system in the greenhouse. *Acta Hortic*. 174: 219-226.

- 6. Hashimoto, Y. 1989. Recent strategies of optimal growth regulation by the speaking plant concept. *Acta Hortic*. 260: 115-121.
- 7. Tantau, H.-J. 1993. Optimal control for plant production in greenhouses. *The Computerized Greenhouse*, eds. Y. Hashimoto, G. P. Bot, W. Day, and H. -J.Tantau, 139-152. San Diego, CA: Academic Press.
- 8. Challa, H., and G. van Straten. 1993. Optimal diurnal climate control in greenhouses as related to greenhouse management and crop requirements. *The Computerized Greenhouse*, eds. Y. Hashimoto, G. P. Bot, W. Day, and H.-J. Tantau, 119-137. San Diego, CA: Academic Press.
- 9. De Baerdemaeker, J., and Y. Hashimoto. 1994. Speaking fruit approach to the intelligent control of the storage system. *Proc. of 12<sup>th</sup> CIGR World Congress, Milano* (1): 190-197.
- Hashimoto, Y., T. Morimoto, and J. De Baerdemaeker. 1995. New approach to total production systems based on an intelligent control. Preprints I<sup>st</sup> IFAC Workshop on Control Applications in Post-Harvest Processing and Technology, 151-156.
- 11. Biggs, M. S., R. William, and A. Handa. 1988. Biological basis of high-temperature inhibition of ethylene biosynthesis in ripening tomato fruit. *Physiol. Plant.* 72: 572-578.
- 12. Lurie, S. 1998. Postharvest heat treatments. Postharv. Biol. Technol. 14: 257-269.
- 13. Hashimoto, Y., B. R. Strain, and T. Ino. 1984. Dynamic behavior of CO<sub>2</sub> uptake as affected by light. *Oecologia* 63: 159-165.
- 14. Hashimoto, Y., P. J. Kramer, B. R. Strain, and H. Nonami. 1990. *Measurement Techniques in Plant Science*. USA: Academic Press.
- 15. Eguchi, H., and T. Matsui. 1977. Computer control of plat growth by image processing II. Pattern recognition of growth in on-line system. *Environ. Control in Biol.* 15: 37-45.
- 16. Hashimoto, Y., T. Ino, P. J. Kramer, A. W. Naylor, and B. R. Strain. 1984. Dynamic analysis of water stress of sunflower leaves by means of a thermal image processing system. *Plant Physiol*. 756: 266-269.
- 17. Uchida, T., Y. Nakanishi, and T. Sakano. 1991. Measurement of bioelectric potential on the surface of Spinach Lamina. *Proc.* 1<sup>st</sup> IFAC/ISHS Workshop on Mathematical and Control Applications in Agriculture and Horticulture, 373-378.
- 18. Baily, B. J., B. G. D. Haggett, A. Hunter, W. J. Albery, and L. R. Svanberg. 1988. Monitoring nutrient film solutions using ion-selective electrodes. *J. Agri. Eng. Res.* 40: 129-142.
- 19. Shiraishi, M., Y. Hashimoto, and S. Kuraishi. 1978. Cyclic variations of stomatal aperture observed under the scanning electron microscope. *Plant and Cell Physiol.* 19: 637-645.
- Morimoto, T., W. Purwanto, J. Suzuki, and Y. Hashimoto. 1997. Optimization of heat treatment for fruit during storage using neural networks and genetic algorithms. *Comput. Electron. Agric.* 19: 87-101.
- 21. Chen, H., and J. De Baerdemaeker. 1993. Effect of apple shape on acoustic measurements of firmness. *J. Agric. Eng. Res.* 56: 253-266.

- 22. De Belie, N., S. Schotte, J. Lammertyn, B. M. Nicolai, and J. De Baerdemaeker. 2000. Firmness changes of pear fruit before and after harvest with the acoustic impulse response technique. *J. Agric. Eng. Res.* 77: 183-191.
- 23. Jancsók, P. T., L. Clijmans, B. M. Nicolai, and J. De Baerdemaeker. 2001. Investigation of the effect of shape on the acoustic response of conference pears by finite element modelling. *Postharvest Biol. Technol.* 23: 1-12.
- 24. De Belie, N., I. C. Hallett, F. R. Harker, and J. De Baerdemaeker. 2000. Influence of ripening and turgor on the tensile properties of pears: A microscopic study of cellular and tissue changes. *J. Amer. Soc. Hort. Sci.* 125: 350-356.
- 25. Morimoto, T., T. Torii, and Y. Hashimoto. 1995. Optimal control of physiological processes of plants in a green plant factory. *Control Eng. Practice* 3: 505-511.
- 26. Morimoto, T., J. Suzuki, and Y. Hashimoto. 1997. Optimization of a fuzzy controller for fruit storage using neural networks and genetic algorithms. *Eng. Appl. Artif. Intell.* 10: 453-461.
- 27. Hashimoto, Y. 1997. Applications of artificial neural networks and genetic algorithms to agricultural systems. *Comput. Electron. Agric.* 18: 71-72.
- 28. Hashimoto, Y., H. Murase, T. Morimoto, and T. Torii. 2001. Intelligent systems for agriculture in Japan. *IEEE Control Systems Magazine* 21: 71-85.
- 29. Chen, S., S. A. Billings, and P. M. Grant. 1990. Non-linear system identification using neural network. *Int. J. Control* 51: 1191-1214.
- 30. Hirafuji, M. 1991. A plant growth model by neural networks and L-system. *Preprints of 9<sup>th</sup> IFAC Symp. on Identification and System Parameter Estimation*, 1: 605-600.
- 31. Morimoto, T., I. Cho, and Y. Hashimoto. 1991. Identification of hydroponics in an advanced control system of the greenhouse. *Proc. of 9<sup>th</sup> IFAC Symposium on Identification and System Parameter Estimation*, (1): 610-615.
- 32. Seginer, I., and R. W. McClendon. 1992. Methods for optimal control of the greenhouse environment. *Trans. ASAE* 35: 1299-1307.
- 33. Goldberg, D. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley.
- 34. Luo, H., and T. Kato. 1987. Studies on the characteristics of seedlings raised in pot under various conditions and their productivity in eggplant and sweet pepper. *Environ. Control in Biol.* 25: 91-96 (in Japanese).
- 35. Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control Signals and Systems* 2: 303-314.
- 36. Isermann R., S. Ernst, and O. Nelles. 1997. Identification with dynamic neural networks. *Preprints of 11<sup>th</sup> IFAC Symp. on System Identification*, 3: 997-1022.
- 37. Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. Learning representation by back-propagation error. *Nature* 323: 533-536.
- 38. Purwanto, W., T. Morimoto, and Y. Hashimoto. 1996. Simulative estimation for the identification of the cumulative response of a plant using neural networks. *Journal of SHITA* 8: 112-118.

- 39. Mizrahi, Y., E. Taleisnik, V. Kagan-Zur, Y. Zohar, R. Offenbach, E. Matan, and R. Golan. 1986. A saline irrigation regime for improving tomato fruit quality without reducing yield. *J. Amer. Soc. Hort. Sci.* 113: 202-205.
- 40. Morimoto, T., J. De Baerdemaeker, and Y. Hashimoto. 1997. An intelligent approach for optimal control of fruit-storage process using neural networks and genetic algorithms. *Comput. Electron. Agric.* 18: 205-224.
- 41. Paull, R. E., and N. J. Chen. 1990. Heat shock response in field-grown, ripening papaya fruit. *J. Amer. Soc. Hort. Sci.* 115: 623-631.
- 42. Collins, G. G., X. Nie, and M. E. Saltveit. 1995. Heat shock proteins and chilling sensitivity of mung bean hypocotyls. *J. Exp. Bot.* 46: 795-802.
- 43. Thai, C. N., R. L. Shewfelt, and J. C. Garner. 1990. Tomato color changes under constant and variable storage temperatures: Empirical models. *Trans. ASAE* 33: 607-614.

# 5.4 Machine Vision in the Agricultural Context

J. A. Marchant

Abstract. This section is merely an introduction to machine vision, an exciting technology that has only recently become feasible in the context of agricultural engineering. Some textbooks, which the author has found particularly useful, are cited to help the interested reader. The section concentrates on a small number of topics that the author considers to be particularly relevant for the type of scenes found in agriculture, i.e. those of natural, biological objects, sometimes found in uncontrollable outdoor lighting situations. Some research papers have been cited to illustrate particular points—leaving work out of the list does not imply any adverse criticism. Also, the author has tried to include generic work from outside the field of agriculture where this is relevant. This is partly to spread the message that agricultural engineers must look outside their particular applications to reap the full benefit of machine vision (and other) research.

**Keywords.** Machine vision, Image analysis, Pattern recognition, Computer vision, Sensing, Agriculture, IT.

#### 5.4.1 Introduction

#### What is Machine Vision?

There are a number of terms that are common when discussing this topic. Among these are *image analysis*, *pattern recognition*, *computer vision*, and *machine vision*. They are all concerned with analyzing two-dimensional data arranged on some sort of spatial grid which, in nearly all practical situations, is rectangular and most likely square. Thus an image is taken to be an arrangement of data that forms a picture when displayed (on a suitable device such as a computer screen or via a printer) to a human observer. There is not complete agreement on terminology but *image analysis* is probably the most general and would include the mathematical manipulation of image

data for such things as contrast enhancement and de-blurring for subsequent human interpretation. *Pattern recognition* usually indicates techniques heavily based on statistical treatment [1] and classification techniques. Often *computer vision* and *machine vision* are synonymous. We will take them both to mean image analysis to make decisions automatically (i.e., without human interpretation) and usually to control some downstream process or device. Note that Davies [2] (whose textbook is highly recommended) has a rather more general view. Thus, this section is concerned with using the technology as a sensing modality for automatic machinery or systems.

# Machine Vision Capability

Machine vision systems have the capability to acquire a very large amount of data in a rapid way. A typical system will be able to capture images of size  $768 \times 576$  pixels in color with 3 bytes for each pixel (one for each color channel). Until recently, image capture and display was linked to standard TV systems which meant that image rates were 25 (if working to European TV standards) or 30 (US standards) times a second. This equates to a data rate of over 3.3 million bytes per second. Because machine vision data is digital, it is ideally suited to analysis by computer with subsequent usage as control information. The data capture device, a camera, is non-contact in nature, which leads to simple engineering for installation, no interference with the subject being sensed (useful, for example, for damage-free inspection), and little intrusion (for example, for observing animals without interfering with natural behavior).

Until recently the large amount of data presented storage problems both on acquisition and during processing. These difficulties are now largely overcome with the availability of very dense memory devices. Even now, processing the data at a rate to match the input is often not possible. One solution is to capture at a lower rate, say 5 images per second, or at a reduced resolution. However, it is quite often the case that the intense processing required to solve real problems makes even these lower rates difficult to handle. Also, a machine control system may demand significantly higher rates than this in order to work well, and high image resolution may be required for effective image analysis. Fortunately the capability of computer systems continues to increase and what was impossible, say, five years ago, is now routine.

The really difficult aspect of machine vision is to be able to approach the ability of the human reasoning system. This is compounded by the fact that, although we can do certain things very easily, we cannot put into words how we do them. One example is how we recognize the face of someone, even if we have not seen the person for some time. We can do this from an extreme range of angles, in most lighting conditions, when the subject has a contorted face, and so on. To design an automatic system we not only have to put this competence into words but into algorithms to operate on the numerical data from the image.

When the subjects to be imaged are fairly simple and are presented in a controlled way to an imaging system the task can be easy. This is why factory inspection systems work well. What makes the agricultural environment especially challenging is that few, if any, of the conditions for "easy" machine vision are met. Among the problems that occur with the biological objects that we deal with are:

- Even if the objects in an image are genetically similar, in engineering terms they will all be different. This characteristic variability means that we cannot have any fixed expectation (or model) of what may appear.
- Sometimes (especially when working outside) the lighting cannot be controlled.
- Sometimes (for example when working with live animals) the presentation of the objects to the camera is variable.
- Objects of interest often touch or overlap.
- Backgrounds can contain much irrelevant and confusing data. It is often difficult even to fit statistical models to this data, which is described as *clutter*.

## **Technological Drivers**

Fortunately, there have been a number of technological and social drivers in recent years that have made the application of machine vision to agriculture feasible. In the mid 1960s G.E. Moore observed an exponential growth in the number of transistors per integrated circuit and predicted that this trend would continue. The often-quoted "Moore's Law" says that the number of transistors on a single chip will double every two years or so. This means that the power of computers has increased dramatically. In a speech in 1997, Moore himself said that this trend could run out in about 2017. However, even with this limitation, capability should continue to increase for some way yet. Until about the mid 1980s the drivers for technology came from the aerospace and defense industries. Thus the costs were high as this market was able to support them. Now the most significant driver (in this author's view) is from the market for consumer items. Initially this was from high-value items such as automobiles, down in the cost range through home computing, and now from very low-cost and high-volume areas such as games and toys. The technology needed for machine vision has been a particular beneficiary of the combination of high capability and low cost. Video interfaces for automobile instruments, games, photography, and home Internet use have all helped to bring machine vision costs down to levels that are acceptable in agriculture.

#### 5.4.2 The Machine Vision Process

#### Image Formation and Sensing

As explained in the previous section, machine vision in the agricultural context is often difficult. It therefore makes sense to understand and use as much information about the process as possible. One source of information is contained in the physical principles behind image formation and sensing. This has been exploited previously in the general context [3,4], and also in our domain of agriculture [5,6]. A useful model for image formation is shown in Figure 1.

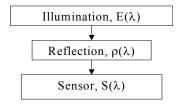


Figure 1. The image formation process.

The illumination, E, is a function of wavelength,  $\lambda$ , and passes through two operations which are also wavelength dependent, i.e. reflection at the object described by the reflectivity,  $\rho$ , and interception at the sensor (after passing through color filters) described by the spectral response, S. The output of a camera can therefore be described by

$$C_I = a_I \int S_I(\lambda) \rho(\lambda) E(\lambda) \, d\lambda \tag{1}$$

where  $C_I$  is the output of camera channel I (i.e., red, green, or blue for a conventional color camera) and  $a_I$  is a constant that describes the gain of the channel. More complete models that take account of highlights (sometimes called *specular reflection*) such as the Dichromatic Reflection Model [7] are available and have been used in the agricultural context [5].

A family of curves for an important illuminant, daylight, is shown in Figure 2 along with a typical reflectivity and a typical set of camera channel sensitivities.

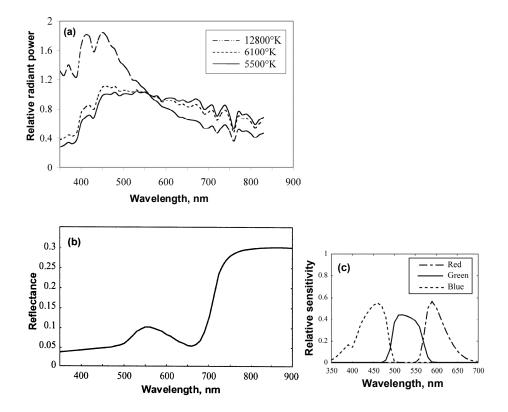


Figure 2. (a) Daylight spectra at three color temperatures. (b) Reflectance spectrum of typical vegetation. (c) Sensitivities of color channels of a typical camera.

The spectrum of daylight is not constant but varies with sky conditions. Work by Judd et al. [8] showed that, although variable, daylight can be characterized by a family of curves with a single variable parameter, the correlated color temperature (CCT). Figure 2 shows three spectra drawn from the Commision Internationale de l'Eclairage (CIE) standard [9] which is based on Judd's work. These CCTs are typical of the variation that could occur over a single day and were, in fact, recorded at Silsoe, UK. The lowest was recorded in full sun—note the relatively high proportion of red (longer wavelengths) in the spectrum. The highest was measured with a clear blue sky in shadow, where the illumination is from the skylight, which is much bluer. The intermediate was measured when an appreciable amount of cloud appeared and the sun was hidden. Equation 1 shows that the camera outputs (i.e., the sensed colors) change as the illuminant spectrum changes. The human perception system accommodates this via a competence called *color constancy* and we do not really notice it. However, artificial systems do record varying colors as the illumination spectrum changes and so any recognition system based on color needs to take this into account. This can be done by using a standard reflecting surface in the image or a separate sensor viewing the sky but, in practice, it is difficult to ensure the illumination conditions at the reference are the same as those at all points in the image. In recent work, by taking advantage of the known characteristics of illuminant families, Marchant and Onyango [10-12] have shown how to compensate for daylight changes in cameras and also in multispectral sensors without making separate reference measurements.

#### Image Capture

The most common capture device for machine vision is the CCD camera. A lens focuses the incoming image onto a silicon array which is divided into a number of independently sensitive areas (pixels), most commonly arranged on a square grid. Silicon sensors are sensitive within the approximate range 350-1000 nm which includes part of the near-infrared (NIR) waveband. This has been exploited within the agricultural context as vegetation is highly reflective in the NIR [13]. However, the user must be careful when using this facility as some manufacturers fit an infrared blocking filter to their cameras. Color cameras have three filters to derive information in (typically) the red, green, and blue areas of the visible spectrum. These filters can be part of the sensor and arranged over each pixel (single-CCD cameras) or the incoming light can be split, e.g. with a prism, and each component directed to a separate sensor (3-CCD camera). These latter are usually more expensive.

As mentioned above, until recently most cameras used TV standard outputs. TV standards were drawn up in the days of analogue cameras whereas modern devices have discrete pixels. Adhering to TV standards requires the output from the image sensor to be converted to an analogue signal (to be compatible with conventional TV displays) and then re-sampled for input to a computer. As well as being complex electronically, this leads to a certain loss of image quality. TV standards have been largely abandoned in recent times and so modern systems can have higher transfer rates. With certain types of image sensors (e.g., CMOS) where the array can be read out randomly, image rate can be traded off against area of view and so smaller areas of inter-

est can be transferred very rapidly. Technology and standards for interfacing cameras to computers is changing rapidly and so no attempt to cover this is made here. The reader is referred to the literature and web sites of the numerous manufacturers for information

#### 5.4.3 Image Analysis

#### **Point Operations**

Figure 3(a) shows an image taken with a monochrome camera using the visible waveband. The gray level histogram (b) is a representation of the so-called first order statistics of the image gray levels. What we hope for is that pixels in one component in the image (say the vegetation) have different statistics from the other (the soil). Then it should be possible to differentiate the two components on the basis of gray level by *thresholding*, i.e., regarding all gray levels below the threshold as one component and all those equal to or above it as the other. The histogram shows that separation cannot be achieved by gray level in this case (using color would have been more productive).

Point operations are ways of transforming pixels without regard to neighboring pixels. They are mappings of the pixels in one image to form another using a single mapping function. A common point operation is histogram equalization where the gray levels are transformed with a function such that all gray values are equally represented in the histogram. In general the function is non-linear. The image of Figure 3(c) has been so transformed. It is important to realize that whatever point operation is used, the separation will still not be possible. Thus, although the image in (c) appears easier to separate into components, a machine vision system will still not be able to do it on the basis of point operations only. The human observer is using much more information than this without realizing it.

Figure 4(a) shows a similar image but extending the waveband into the NIR. As expected the contrast is much better (see Section 5.4.2) and the histogram shows two clear modes which correspond to vegetation and soil. Thresholding at the trough between the modes (gray level value 128) and setting high values to white and low values to black gives the separation shown in (c).

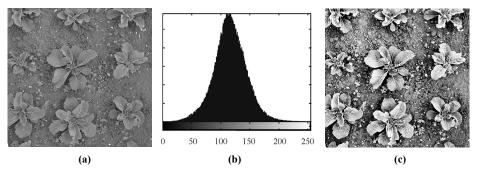


Figure 3. (a) Monochrome image of vegetation and soil. (b) Gray-level histogram. (c) Scene in (a) with histogram equalization.

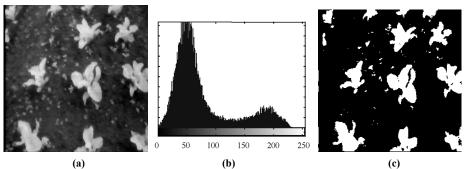


Figure 4. (a) Monochrome image of vegetation and soil extending into the NIR. (b) Gray-level histogram. (c) Thresholded image.

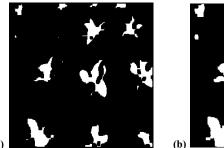
# Neighborhood Operations

Point operations take no account of the spatial organization of the image, whereas it is this very organization that forms the fundamental character of images as opposed to other types of data. It is not surprising therefore that operations that do take spatial relationships into account are much more powerful. These work on groups of pixels that are usually contiguous, hence the title *neighborhood operations*.

One simple type of neighborhood operation is the linear filter and the simplest example is the averaging or low-pass filter. Each pixel in the original image is considered to be at the center of a square region with a fairly small odd dimension, e.g.  $3 \times 3$  pixels. (The neighborhoods do not have to be odd-sized, square, or even rectangular but they generally are for ease of calculation.) With the averaging filter, the mean value of the pixels in the neighborhood is used to form a pixel in a new image at the same location as the center pixel in the original image. Like all averaging operations this can be used to reduce some types of noise at the expense of a loss of sharpness. The converse of the low-pass filter is the high-pass filter, which operates by differencing adjacent pixels. These are used to detect edges, which are characterized by sharp changes in gray level. As in other areas of signal processing, high-pass filtering amplifies noise. This will become apparent if a naive attempt is made to find object boundaries in realistic scenes with a simple edge detector.

Sometimes, non-linear operations on neighborhoods yield better results. An example is the use of a median filter rather than an averaging filter to remove noise. Consider an area of a scene of reasonably uniform gray-level; the resulting image may have a single pixel with a very different gray-level due to some noise process. The output of an averaging filter would contain a proportion of this outlier pixel value. However, the median filter would set the output pixel to the median value (the  $5^{th}$  largest gray level in a  $3\times3$  neighborhood) and so would be unaffected by the outlier.

A large and powerful class of non-linear neighborhood operations are *morphological* methods [14]. In general, morphological methods extend naturally to gray-level (and multiband) images [14]. Only a brief illustration is given here, where small objects are removed from a binary (2-value) image. Figure 5(a) shows the image in



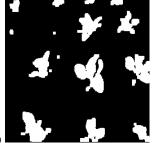


Figure 5. (a) Image of Figure 4(c) with erosion operator applied. (b) Dilation operator applied to (a).

Figure 4(c) after a 3 × 3 binary *erosion operator* has been applied twice. This operator sets the result pixel to logical '0' if any of the pixels in the source neighborhood (including the source pixel) is '0' and to logical '1' otherwise. The effect on a large blob is to remove a skin from the surface. If the blob is small enough it will be removed completely. The opposite of erosion is *dilation*: If any pixels in the neighborhood are logical '1,' the result is set to '1.' This operator has been applied twice to 5(a) to give 5(b). As can be seen, following erosion by dilation (an operation known as *opening*) changes the original image—one does not undo the other. In particular, once removed small blobs do not grow back whereas large ones regain approximately their original size. However, there may be some undesirable effects, for example parts of blobs that were connected can become disconnected. It would no doubt be possible to tune an algorithm to retain the desirable effects without the undesirable ones. However, a general lesson in machine vision is that algorithms can become tuned to one image or a small set and do not work on others. This is a failing of much work in machine vision and the importance of proper algorithm testing cannot be overstressed [15].

# Regions

An important procedure in image analysis is grouping individual pixels into regions. A *region* is an image area whose pixel properties have some commonality. The implication is that the regions will also be meaningful physical entities like a single plant or the body of an animal. The blobs in Figure 4(c) are regions and each one shares the property that the gray levels of its pixels are over the threshold value. In this image the individual plants are surrounded by darker soil and so each blob is often a single plant.

It is possible to find the extent of each blob in a number of ways. One fast method that leads to a compact data representation is the *Freeman chain code* [16]. The direction from a pixel to any of its neighbors in a  $3 \times 3$  grid is given a label between 0 and 7. After finding the first pixel on a blob boundary the direction to the next one is found and the label stored. The algorithm travels around the boundary until the start pixel is re-visited thus storing the whole direction sequence. In early work on vegetable grading, this algorithm was developed in hardware [17] but it may well be possible now to achieve sufficient speed with a software version, given the large increase in computing power since.

Often a simple grouping strategy like thresholding is not sufficient. Region growing is the process whereby a "seed" pixel is used along with a statistical model of the attributes of pixels that should form part of a given region. The model might consist, for example, of a specification of the mean and standard deviation of some aspect of the pixel's color, such as (G-R)/(G+R) where G and R are the green (higher for vegetation) and red (often higher for soil) channel outputs. The specification could be derived from prior knowledge or from training images. The general technique is to grow the region from the seed by accepting into the region neighboring pixels that probably belong to the statistical population represented by the model. The relevant statistical technique for inclusion might be a t-test [18].

Snakes [19] form an interesting class of region-finding techniques. Their operation is rather like placing an elastic string near the boundary of a region where forces act on the string to pull it on to edges. The mechanics of the string allows the snake to bridge gaps in the boundary and even small occlusions. The tension in the string stops the region growing indefinitely. They can thus overcome some of the problems that obtain in our particular area (see Section 5.4.1) and have been used in the agricultural context [20]. However, choosing the mechanical properties of the snake and finding an initial position can be difficult. The advantages of region growing and snakes in outdoor scenes have been combined by Alexander and Buxton [21].

#### Features and Classification

A large part of machine vision is concerned with extracting features from images and classifying components. For example, an agricultural problem might consist of classifying a color image into vegetation and soil, and then further classifying the vegetation into crop plants and weeds. An obvious starting point would be to exploit the color differences between vegetation and soil [22] and then use a shape analysis on the individual blobs to discriminate between crop plants and weeds [22-24]. As an incidental point, our reasoning here assumes that blobs equate to characteristic plant parts like whole leaves. In practice this will not be the case as leaves or parts of leaves merge into single blobs or single leaves split into many blobs. Finding individual parts of images that are meaningful entities (segmentation) is a very difficult problem with realistic scenes.

Classification can be done at different levels. For example, in the vegetation/soil problem above the classification could be done on individual pixels based on the ratio (G-R)/(G+R). A histogram of this ratio should be bi-modal (e.g., [22]), the valley between the two peaks could be identified automatically, and the two components classified by thresholding. This would be an example of *unsupervised classification*. Alternatively, *supervised classification* requires examples of the classes to be presented to the system. Statistical models can then be derived and an unknown pixel can be assigned to a category depending on how close the pixel's features are to the statistics of each class. Methods for supervised classification include Bayesian methods, some neural networks, and nearest-neighbor techniques [25]. In the example above, using the color ratio, each pixel has a single feature or measurement associated with it. Most techniques for classification extend naturally to more than one feature whereby better

discrimination can often be achieved. A higher level of classification can be done on blob features. For example, various size and shape measures could be derived for each blob, then used with a multi-feature classifier to classify the blobs into types of object.

As is mentioned frequently in this section, machine vision in the agricultural context is far from easy. Many researchers have resorted to artificial intelligence techniques in the hope that they may be able to emulate the effectiveness of human reasoning. Examples are in sorting and grading [26-29], plant identification [23,30,31], plant structure identification [32], and the quality of snack foods [33]. Although offering promise, an undesirable aspect of many AI techniques is that a number of parameters have to be chosen with few, if any, rules governing the choice. An example is with neural networks, where the structure of the network cannot be designed with firm rules. Yet another approach is to assume that no single technique or measurement will provide the best answer whereupon information from various sensors and methods are combined in an intelligent way [34,35].

#### Model-Based Techniques

When searching for patterns in data, knowledge of what to expect is often useful. This has been mentioned briefly above in the context of statistical models for region growing. Here we consider using prior knowledge of the direct spatial arrangement of pixels in the image. This includes models for the shape of objects and the patterns of gray levels within them.

A naive approach to machine vision for natural objects may include the formation of a template. This would be a specified arrangement of gray levels (or colors) that could be derived from a sample section of an image. For example, the gray levels in a single weed leaf could be stored. In order to find another example, the image could be searched for sections that match (to some tolerance) the template. In a realistic situation this would almost certainly fail due to the natural variability mentioned in Section 5.4.1. The target area could have a completely different arrangement of gray levels and still be a weed leaf. It would seem that there is no place for fixed templates in our domain. However, in at least two areas they can be useful. These are motion estimation and 3-D triangulation. Motion can be measured from differences between image features in successive frames of a sequence. If the time difference between frames is short so that the apparent inter-frame movement is small, then the appearance of individual areas will not be too different. A template from a small area of one image can be matched to a corresponding area in the next. Often, knowledge of the expected motion can be used to define a (hopefully) small area in which to search for a match. The degree of match can be calculated using the normalized cross correlation [36] although it may be worth considering a Fourier transform approach [37] which can give additional information (via the phase) on the position of the match to sub-pixel resolution. In a very similar way, templates can be used to match features in two images for stereo triangulation.

A template that could change its shape and appearance could potentially bring the advantages of using prior knowledge to the analysis of images of biological objects. Among these advantages would be the ability to bridge gaps in relevant image data,

such as may occur with overlaps, or in indistinct regions, or where background clutter is present. The idea of a flexible template was first introduced by Widrow [38] and the general method was applied to biological objects (human faces) by Yuille et al. [39]. In this latter work, templates were hand-crafted to suit particular features such as eyes. A similar idea for plant identification has been used by Manh et al. [40]. An interesting approach is to include not only the expected shape but the major ways in which the shape can change. This was termed a *point distribution model (PDM)* by Cootes et al. [41]. The positions of landmarks around the object boundary are identified by hand in several training images. A principal component analysis is then done to identify the modes of variation of the shape. The method can often represent most of the variability in a few modes. More recently, the surface appearance has been included in a method that can very rapidly fit models to image data [42]. The basic idea has been used for tracking pigs [43] and for estimating the mass of fish [44]. However, just like the snake, fitting these models poses problems with initialization.

A different type of model-based method is the *Hough transform* [2], which can be used when searching for abstractions such as lines or circles in images. As an example, consider the general equation for a straight line

$$y = mx + c \tag{2}$$

Now consider that the position of an image feature has been found. This could be, say, the centroid  $(x_f, y_f)$  of one of the blobs in Figure 4(c). The hypothesis is that this feature lies on a line. If this is true, then there must be a relationship between the two parameters, m and c, of the line, which is obtained from Equation 2 as

$$m = \frac{y_f - c}{x_f} \tag{3}$$

The parameter space is discretized such that any pair of (m, c) values is represented by a cell in the space. A two-dimensional accumulator is now set up and all the cells represented by Equation 3 are incremented. The accumulator is a representation of all the possible lines that could pass through  $(x_f, y_f)$ . If another feature is treated in the same way then a second set of cells will be incremented. One of these cells will be incremented by both operations and will, in fact, represent the parameters of the line that joins the features. As more feature values are loaded into the accumulator, cells that accumulate most will represent lines along which many features lie. After all features have been loaded, the accumulator can be searched for high values, which will represent significant lines in the image. The method is resistant to missing data (missing parts of lines in the original image) and tolerant to spurious data (features not lying on lines). It is thus a useful method with the difficult images found in our applications. The method can be extended to other types of abstractions that can be represented parametrically, e.g. circles and ellipses [2]. There is also much work on techniques for loading and efficiently searching the accumulator. In the agricultural context the technique has been used for the location of crop rows in images like that of Figure 4(c) [45]. A model for the distortion of three parallel rows of crop as viewed through the camera of an autonomous vehicle was developed where the parameters were the heading angle and offset of the vehicle. Using three rows at once made the method robust to missing plants or whole missing rows.

#### **5.4.4 Conclusions**

Machine vision is a technology with many advantages for agricultural engineering. Fortunately there are technological and social factors that are driving the costs of system components down. At the same time the capability of computing systems continues to increase. These factors will tend to remove barriers to uptake. However, it must be recognized that human capability in visual sensing and interpretation far outstrips that of artificial systems. What may appear easy for us can be extremely difficult to emulate in machine vision. Thus there is great scope for engineers to use their skills and creativity to develop systems that solve practical problems in a reliable way.

#### References

- Devijver, P. A., and J. Kittler. 1982. Pattern Recognition. London, UK: Prentice-Hall.
- 2. Davies, E. R. 1997. *Machine Vision*, 2<sup>nd</sup> ed. London, UK: Academic Press.
- 3. Klinker, G. J. 1993. *A Physical Approach to Color Image Understanding*. Wellesley, MA: A. K. Peters Ltd.
- 4. Healey, G. E., S. A. Shafer, and L. B. Wolff, eds. 1992. *Physics-Based Vision: Principles and Practice*. Boston, MA: Jones and Bartlett.
- 5. Pla, F., F. Juste, F. Ferri, and M. Vicens. 1993. Colour segmentation based on a light reflection model to locate citrus fruits for robotic harvesting. *Computers and Electronics in Agriculture* 9: 53-70.
- 6. Andersen, H. J., C. M. Onyango, and J. A. Marchant. 2000. Design and operation of an imaging sensor for detecting vegetation. *Int. J. Imaging Systems and Tech.* 11: 144-151.
- 7. Shafer, S. A. 1985. Using color to separate reflection components. *Color Research and Application* 10: 210-218.
- 8. Judd, D. B., D. L. MacAdam, and G. W. Wyszecki. 1964. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* 54: 1031-1040.
- 9. Commission Internationale de l'Eclairage (CIE). 1986. *Colorimetry*. Tech. Rep. 2<sup>nd</sup> ed. Paris, France: CIE.
- 10. Marchant, J. A., and C. M. Onyango. 2000. Shadow-invariant classification for scenes illuminated by daylight. *J. Opt. Soc. Am. A.* 17: 1952-1961.
- 11. Marchant, J. A., and C. M. Onyango. 2001. Color invariant for daylight changes: Relaxing the constraints on illuminants. *J. Opt. Soc. Am. A.* 18: 2704-2706.
- 12. Marchant, J. A., and C. M. Onyango. 2002. Spectral invariance under daylight illumination changes. *J. Opt. Soc. Am. A.* 19: 840-848.
- 13. Walter-Shea, E., and J. Norman. 1991. Leaf optical properties. *Photon-Vegetation Interactions*, eds. R. B. Myneni, and J. Ross, 230-250. New York, NY: Springer-Verlag.

- 14. Soille, P. 1999. *Morphological Image Analysis*. Berlin, Germany: Springer-Verlag.
- 15. Haralick, R. M. 1992. Performance characterization in computer vision. *BMVC92*, eds. D. Hogg, and R. Boyle, 1-8. London, UK: Springer-Verlag.
- 16. Freeman, H. 1961. On the encoding of arbitrary geometric configurations. *IEEE Trans. Electronics and Computing* 10: 260-268.
- 17. Onyango, C. M., and J. A. Marchant. 1994. Transputers for high speed grading. *Food Control* 5: 29-32.
- 18. Snedecor, G. W., and W. G. Cochran. 1989. *Statistical Methods*, 8<sup>th</sup> ed. Ames, IA: State University Press.
- 19. Kass, M., A. Witkin, and D. Terzopoulos. 1987. Snakes: Active contour models. *Int. J. Computer Vision* 1: 321-331.
- 20. Marchant, J. A., and C. P. Schofield. 1993. Extending the snake image processing algorithm for outlining pigs in scenes. *Computers and Electronics in Agriculture* 8: 261-275.
- 21. Alexander, D. C., and B. F. Buxton. 1996. An evaluation of physically based statistical colour models for image region characterisation. *Proc. of the International Conference on Image Processing*, ed. P. Delogne, 1023-1026. New York, NY: IEEE.
- 22. Perez, A. J., F. Lopez, J. V. Benlloch, and S. Christensen. 2000. Colour and shape analysis techniques for weed detection in cereal fields. *Computers and Electronics in Agriculture* 25: 197-212.
- 23. Oide, M., and S. Ninomiya. 2000. Discrimination of soybean leaflet shape by neural networks with image input. *Computers and Electronics in Agriculture* 29: 59-72.
- 24. Molto, E., N. Aleixos, L. A. Ruiz, J. Vazquez, F. Fabado, and F. Juste. 1998. Determination of weeds and artichoke plants position in colour images for local herbicide action. *Acta Horticulturae* 421: 279-283.
- 25. Ripley, B. D. 1996. *Pattern Recognition and Neural Networks*. Cambridge, UK: Cambridge University Press.
- 26. Shahin, M. A., E. W. Tollner, and R. W. McClendon. 2001. Artificial intelligence classifiers for sorting apples based on watercore. *J. Agricultural Engineering Research* 79: 265-274.
- 27. Jahns, G., H. M. Nielsen, and W. Paul. 2001. Measuring image analysis attributes and modelling fuzzy consumer aspects for tomato quality grading. *Computers and Electronics in Agriculture* 31: 17-29.
- 28. Kondo, N., U. Ahmad, M. Monta, and H. Murase. 2000. Machine vision based quality evaluation of Iyokan orange fruit using neural networks. *Computers and Electronics in Agriculture* 29: 135-147.
- 29. Morimoto, T., T. Takeuchi, H. Miyata, and Y. Hashimoto. 2000. Pattern recognition of fruit shape based on the concept of chaos and neural networks. *Computers and Electronics in Agriculture* 26: 171-186.
- 30. Tang, L., L. Tian, and B. L. Steward. 2000. Color image segmentation with genetic algorithm for in-field weed sensing. *Trans. ASAE* 43: 1019-1027.

- 31. Burks, T. F., S. A. Shearer, R. S. Gates, and K. D. Donohue. 2000. Backpropagation neural network design and evaluation for classifying weed species using color image analysis. *Trans. ASAE* 43: 1029-1037.
- 32. McFarlane, N. J. B., B. Tisseyre, C. Sinfort, R. D. Tillett, and F. Sevila. 1997. Image analysis for pruning of long wood grape vines. *J. Agricultural Engineering Research* 66: 111-119.
- 33. Sayeed, M. S., A. D. Whittaker, and N. D. Kehtarnavaz. 1995. Snack quality evaluation method based on image features and neural-network prediction. *Trans. ASAE* 38: 1239-1245.
- 34. Ozer, N., B. A. Engel, and J. E. Simon. 1995. Fusion classification techniques for fruit quality. *Trans. ASAE* 38: 1927-1934.
- 35. Ros, F., S. Guillaume, G. Rabatel, F. Sevila, and D. Bertrand.. 1997. Combining global and individual image features to characterize granular product populations. *J. Chemometrics* 11: 483-500.
- 36. Rosenfeld, A., and A. C. Kak. 1982. *Digital Picture Processing*, 2<sup>nd</sup> ed., Vol. II. Orlando, FL: Academic Press.
- 37. Gonzalez, R. C., and R. E. Woods. 1992. *Digital Image Processing*. Reading, MA: Addison Wesley.
- 38. Widrow, B. 1973. The "rubber mask" technique. *Pattern Recognition* 5: 175-211.
- 39. Yuille, A. L., P. W. Hallinan, and D. S. Cohen. 1992. Feature extraction from faces using deformable templates. *IEEE International Journal of Computer Vision* 8: 99-111.
- 40. Manh, A. G., G. Rabatel, L. Assemat, and M. J. Aldon. 2001. Weed leaf segmentation by deformable templates. *J. Agricultural Engineering Research* 80: 989-996.
- 41. Cootes, T. F., C. J. Taylor, D. H. Cooper, and J. Graham. 1992. Training models of shape from sets of examples. *BMVC92*, eds. D. Hogg, and R. Boyle, 9-18. London, UK: Springer-Verlag.
- 42. Cootes, T. F., G. J. Edwards, and C. J. Taylor. 2001. Active appearance models. *IEEE Trans. Pattern Analysis and Machine Intelligence* 23: 681-685.
- 43. Tillett, R. D., C. M. Onyango, and J. A. Marchant. 1997. Using model based image processing to track animal movements. *Computers and Electronics in Agriculture* 17: 249-261.
- 44. Lines, J. A., R. D. Tillett, L. G. Ross, D. Chan, S. Hockaday, and N. J. B. McFarlane. 2001. An automatic image-based system for estimating the mass of free-swimming fish. *Computers and Electronics in Agriculture* 32: 151-168.
- 45. Marchant, J. A., and R. Brivot. 1995. Real time tracking of plant rows using a Hough transform. *Real Time Imaging* 1: 363-371.

# 5.5 Fertilizer Application Control

P. Van Liedekerke, J. De Baerdemaeker, and H. Ramon

Abstract. This section highlights uses of computing and modeling in fertilizer application. The goal is a better control of spread pattern and applied dose for precise fertilizer management. After a short description of different fertilizer types, on-line measurement and adaptation of dose control is discussed. These techniques are still developing, but they will allow specific spreading doses that conform to environmental regulations. Next, a new technique using numerical models is discussed. This allows calculation of spreading patterns as a function of both particle and machine properties. In this way, expensive tests can be avoided. Finally, both methods are linked to GIS, such that an optimal fertilizer dose can be matched to specific soil needs.

Keywords. Fertilizer control, Spread pattern, Weed detection.

# 5.5.1 Fertilizer Spreader Types

In agriculture, four types of granular fertilizer spreaders are used:

Centrifugal or spinning disc spreader—In a centrifugal type spreader, which are the most used type, the fertilizer particles are dropped on a rotating (conical) disc and accelerated by the vanes. With this system, the spread pattern is affected by the disc rotation speed (usually 500-600 rpm), the length and shape of the vanes, the inclination of the disc, and the initial impact point of the fertilizer on the disc. Types with two rotating discs may have a longitudinal pattern up to 36 m. The shapes of the patterns can be trapezium-like or triangle-like, depending on the rotational direction of the discs.

Pendulum spreader—This system works with a horizontally oscillating pendulum by which the fertilizer is accelerated. In general, these types have a smaller transversal and longitudinal spread pattern and thus the risk of irregularities may be lower. The shape of the pattern is triangle-like.

Pneumatic spreader—This spreader is equipped with a boom (like a spray boom) in which the fertilizer particles are driven to the soil by compressed air through a piping system. The advantages of this system are its high accuracy, the evenness of the spread pattern, and the low influence of wind and fertilizer condition. In addition, there is almost no transversal offset, such that the farmer is able to spread very accurately at the borders of the area of application. On the other hand, the width of the pattern is limited by the length of the boom.

Band fertilizer—A machine much like a seeding machine or pellet applicator can be used to apply fertilizer particles in rows. This equipment can be combined with other field actions like seeding or drilling, and thus provides a way to save time. This kind of application close to the crop row is suitable for granular fertilizer types that are effective over a longer time interval.

## 5.5.2 Fertilizer Dose Control in Centrifugal Spreaders

## On-Line Measurement of Spread Pattern

With a growing concern about the environment, fertilizer inputs should be reduced as much as possible without yield loss. This can be achieved by improving the quality of spread patterns. The calibration of spreaders in test halls is, however, time consuming and costly. Therefore, Hofstee [1] proposed a "predict" rather than a "collect" method.

An on-line spread pattern is predicted through an optical sensor that measures the particle velocities and diameters immediately after they leave the impeller. The sensor will complete several cycles around the edge of the disc, sampling for a well-defined amount of angular positions of the particles. The particle velocities and diameters are thus monitored for each of these angles and introduced into a ballistic model, which results in the on-line predicted distribution.

In combination with GPS, the system could be used to develop smart systems, where the spreader operates at the maximum rate and the overlap distances are varied to obtain the locally desired application rate. This indeed would result in a minimum total work time. Another advantage is that it can be used for discrete element method (DEM) validations.

## On-Line Measurement and Adaptation of Dose

When the desired fertilizer dose is known, the correct fertilizer application rate can be set. The mass flow of a centrifugal spreader is usually controlled by the aperture of a valve or the rotational speed of roller feeders. However, the mass flow is largely affected by the friction between the particles [2], which depends on the air moisture content and hygroscopic properties of the fertilizer. As a consequence, with each new application (e.g., each different fertilizer type or even weather type), a calibration of the mass flow control system should be carried out. This calibration is usually performed by weighing the instantaneous amount of fertilizer in the hopper, and comparing it with the predicted mass distributed by the spreader. The weighing device can be a load cell placed between the tractor hitch and the implement. The load cell measures the weight of, for instance, the spreader frame.

Contrary to this static measurement, automatic calibration may be preferred during application. This can be accomplished by using a second reference load cell (this is basically an accelerometer) that measures a known small weight, so that all disturbances caused by varying field conditions can be compensated for. In this case, it is possible to enter the desired quantity (kg/ha) in an on-board computer and the spreading rate is automatically implemented without executing a preceding calibration test.

Finally, adaptation of the dose can also be performed directly by controlling and monitoring the impulse on the disc. In this case, the discs are driven by a hydrostatic transmission. By measuring the pressure differences between inlet and outlet (impulse measurement), not only can the dose be regulated, but also variations in the mass flow, due to changes in the fertilizer properties [3].

Information on actual settings and flow rates can be recorded together with position data to obtain treatment maps. This information is important for field management and

also for documentation of production methods for sale of the crop and possible regulations on environment and fertilizer use. Precision agriculture has led to the development of software to store and use data per field, including handling of spatial data in a GIS.

#### Position and Time Lag

In modern agriculture, either a tractor equipped with a GPS antenna and a fertilizer map, or an on-line N-sensor, can provide direct information about the fertilizer needs of soil and crop. In this case a time lag exists between the information received by the sensor/antenna and fertilizer dropped on the field. This time lag can cause a positional offset, D.

The offset can be split in the distance *X* between sensor/antenna and the center of the spreading cone (or the boom position in case of a pneumatic spreader) and an adjustment time lag *Y*. The total positional offset is then:

$$D = X-Y$$

$$= X-v_{tractor}(t_0 + t_1)$$

where  $v_{tractor}$  = tractor speed

 $t_0$  = reaction time of the dosage and control system

 $t_I$  = reaction time of spreader (from hopper orifice to field)

As a consequence, every tractor-spreader combination with a GPS control system should account for this lag in order to acquire satisfactory results with variable rate application.

# 5.5.3 Information Technology for Design: Spread Pattern Optimization for a Centrifugal Spreader

In a centrifugal spreader, the fertilizer behavior on the disc affects the fertilizer distribution pattern. For the simulation of the fertilizer behavior on the disc, the parameters that have a significant influence on the spreading pattern must be well understood. Until now, simulations have been performed by assuming that the fertilizer particles are spheres and show no interaction with each other during their acceleration on the disc. With these assumptions, the equations of motion and the trajectories of the fertilizer particles can be obtained quite easily [1,2,4-6]. Experimental research in test halls revealed that the most important parameters influencing the spread pattern are velocity of the disc, shape and amount of mass flow through the orifice, friction between fertilizer and vanes, and the fertilizer type. Comparison between simulations and experiments shows a reasonable agreement although it seems that the errors are much larger when real fertilizers instead of particles such as steel ball bearings are used. In general, the measured spread distribution is smoother than the shape of the simulated particle distributions, probably due to model simplifications.

In the analytical models, interactions between the particles is not considered, the simulated trajectories are for spherical particles, and in addition, the initial position and speed of the fertilizer particles on the disc is not known [2,6]. Possibly, these problems can be solved with simulation tools that account for all interactions between

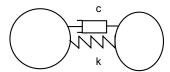


Figure 1. Contact conditions between particles.

particles and vanes. The simulation starts in the bin with a huge amount of particles stored together, goes through the motion of the particles on the disc, and ends up with the particle flow through the air.

A discrete element model (DEM) simulation is based on the following principles:

- 1. At each time *t*, all possible contacts between the objects (particles, vanes,...) are considered.
- 2. For the objects in contact, all contact forces are calculated between all objects (particles, vanes, ...); these forces are modelled as damped springs F = -Kx c(dx/dt) (Figure 1). No centrifugal or Coriolis forces are directly involved as is the case in the previous approach.
- 3. The equations of motion are integrated, resulting in the position and speed of all particles on time *t*.
- 4. The next time step is considered at  $t+\Delta t$  (time steps are typically  $10^{-6}$  s).

Particles in DEM can be modelled as spheres but also as clusters of spheres (irregular particles). With DEM, all kinds of situations can be simulated, although the restriction here is the simulation time. By considering all possible contacts, simulation time will rise when a large number of particles is involved [7].

Existing simulations with DEM with one particle already show a very good agreement with simple experiments, so that the real multi-particle systems are worthwhile to consider for the near future. Finally, if the DEM simulations agree well with experiments for multiple particle applications, it may help in optimizing spreading patterns and decrease loss of fertilizer on the disc, either by adjusting spreader characteristics, or by improving the spreadibility properties of the fertilizer (shape, stiffness, friction, etc.).

Together with a continuous increase of computational power, DEM can become a powerful simulation tool, able to model and predict the dynamic behavior of all kinds of granular material.

#### 5.5.4 Determining the Optimal Local Treatment

The application rate of fertilizer is very important since every soil and crop needs its specific amount of nitrogen, phosphate, and other nutrients. To determine the optimal dose, soil sample analysis is traditionally used. For nitrogen, the fertilizer dose can be determined based on the deficit between the nitrogen need of the crop and the nitrogen provided by the soil (including both nitrogen in the soil and estimated nitrification in the soil).

However, the traditional analysis of mixed soil samples yields an average optimal dose for areas of (tens of) hectares of soil. Since growing conditions are rarely uniform, this will lead to over- or under-fertilization for some areas. In precision agricul-

ture, the variation within fields is studied in order to apply the locally optimal dose, thus increasing the efficiency of inputs (e.g., no over-fertilization). Optical tools were developed to have an on-line adjustment of the nitrogen dose based on the optical crop response at certain growth stages. Crop plants with a shortage of nitrogen will have slower growth and lower chlorophyll content, which can be detected in the crop reflection spectrum [8,9]. This system is commercially available as Hydro N-Sensor by YARA (Brussels, Belgium; www.yara.com) for corn and wheat crops, providing realtime sensing and application of an adjusted nitrogen dose to the crop. This system works under the assumption that only nitrogen limits crop growth. If nitrogen is not the only factor affecting crop growth, an off-line approach is preferred to determine the optimal local nitrogen dose. This gives the opportunity to construct a fertilizer application map that takes local field information into account. If certain areas in the field are expected to produce less, the fertilizer dose can be adjusted accordingly, to avoid over-fertilization. This can be done in a geographic information system (GIS), combining existing field information such as yield maps, soil maps, and other observations linked to certain places in the field.

Both concepts of on-line and off-line adjustment of local treatments are studied in precision agriculture (see Figure 2) and can be applied to other crop production measures as well, like crop protection. In the case of weed treatment, the off-line approach leads to mapping of weeds, which can be used to derive treatment maps or herbicide application maps with the adjusted herbicide formulation and dose. The latter are used

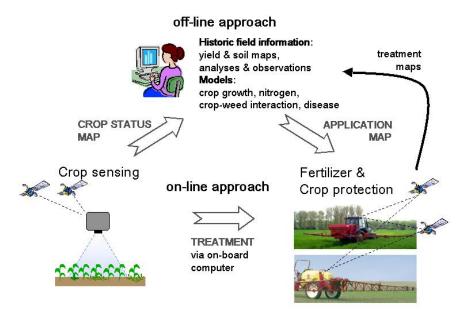


Figure 2. Principle of on-line and off-line approach to precision fertilizing and precision crop protection.

during spraying operations to activate the spraying system through the onboard computer for the field sprayers. Since weed treatments can be determined in a fairly straightforward manner (for example, based on weed threshold value), the real-time approach can be useful here, providing the right herbicide treatment at the right time. This is sometimes referred to as *weed-activated spraying*.

Compared to the real-time concept, the off-line mapping concept provides better opportunity to choose the appropriate treatment (e.g., fertilizer dose, herbicide or pesticide formulation). It is, however, more labor intensive, because several automated or manual field inspections are necessary and a specialist must analyse available information to make or check the application map.

To apply an adapted dose in the field, localization of the equipment is necessary. This can be done with a Global Positioning System (GPS). GPS uses receivers for satellite signals and triangulation to pinpoint a location, yielding XY or other coordinates with an accuracy of about 1 m in the XY-plane. Commercial systems exist whereby an on-board computer with the data of the fertilizer application map, linked to a GPS receiver, is able to automatically set the correct dose for each location in the field. Similar systems are devised for controlling the local spray dose for crop protection.

#### References

- 1. Hofstee, J. W. 1993. Physical properties of fertilizer in relation to handling and spreading. PhD thesis. Wageningen, The Netherlands: Agricultural University.
- 2. Hofstee, J. W. 1995. Handling and spreading of fertilizers Part 5: The spinning disc type fertilizer spreader. *J. Agricultural Engineering Research* (62): 143-162.
- 3. http://www.kuhnsa.com/internet/actualit.nsf/0/2EF3FA8B0397DDC4C12569B2 0026EA98?OpenDocument)
- 4. Patterson, D. E., and A. R. Reece. 1962. The theory of the centrifugal distributor I: Motion on the disc, near-centre feed. *J. Agricultural Engineering Research* 7(3): 232-240.
- 5. Inns, F. M., and A. R. Reece. 1962. The theory of the centrifugal distributor II: Motion on the disc, off-centre feed. *J. Agricultural Engineering Research* 7(4): 345-353.
- Olieslagers, R., H. Ramon, and J. De Baerdemaeker. 1996. Calculation of fertilizer distribution patterns from a spinning disc spreader by means of a simulation model. J. Agricultural Engineering Research 63: 37-152.
- 7. Tijskens, E., H. Ramon, and J. De Baerdemaeker. 2003. Discrete element modeling for process simulation in agriculture. *J. Sound and Vibration* 266(3): 493-514.
- 8. Filella, I., L. Serrano, J. Serra, and J. Peñuelas. 1995. Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. *Crop Sci.* 35: 1400-1405.
- 9. Reusch, S. 1997. Entwicklung eines reflexionsoptischen Sensors zur Erfassung der Stickstoffversorgung landwirtschaftlicher Kulturpflanzen. Dissertation, Universität Kiel.

# 5.6 Sensing and Information Handling for Crop Protection

H. Ramon, D. Moshou, C. Bravo, E. Vrindts, and J. De Baerdemaeker

Abstract. Precision farming techniques have given new impetus to the development of efficient crop protection, enabling locally optimized crop treatments after evaluation of the crop status. During recent years different sensing techniques have been developed for this evaluation. In this section we look at weed sensing and crop disease detection. The wide variety of automatic evaluation possibilities rely on the radiation characteristics of the crop. Solar light reflection as well as thermal radiation can be used. Chlorophyll fluorescence provides deeper insight in the physiological changes during plant growth. These techniques, however, need to be used with great care. Upto-date computational methods synthesize and provide reliable information concerning the crop status in real field circumstances. Neural networks and sensor fusion architectures are discussed.

**Keywords**. Weed detection, Reflection, Fluorescence, Disease detection, Sensor fusion, Neural network, Spot spraying.

#### 5.6.1 Introduction

Traditionally, crop protection measures are decided after an inspection of the field to assess the presence of weeds, the weed species and density, and the presence and spread of diseases and pests. This task is performed by the farmer, who later decides on the necessary treatments. In recent years, optical sensing techniques have been tested for automation of weed and disease detection. Developments in precision farming have made it possible to apply variable treatments in the field, with appropriate treatment for every place in the field. Figure 2 of Section 5.5 of this handbook illustrates the working principles of such a system. The following sections give an overview of techniques that can be used for more automated detection of weeds and of crop health status.

# 5.6.2 Optical Techniques for Weed Detection

A few commercially available systems (Detectspray by Concord Inc., Australia; Spot Shot by Progressive Farm Products Inc., USA; and Weedseeker by Patchen Inc., USA) use green plant detection based on red and near-infrared reflection to focus weed treatments on-line. These systems do not make the distinction between crop and weeds, but they are useful between row crops, on fallow land, in orchards or vine-yards, and in other non-agricultural areas [1,2].

GPS and computers can be used to pinpoint weed patches in the field during field inspection, but this is labor-intensive. Through remote sensing (aerial photography, high resolution satellite imagery), high-density weed patches in the field can be detected. However, low densities and mixed species cannot be correctly identified with remote sensing [3,4]. Field inspection is necessary to be certain of the species in the weed population.

Machine vision has been tested for automatic weed detection. Shape, texture, and color can be used for crop and weed identification [5-8]. Segmentation, or making the distinction between plant and background, is the first important step. Plants are mostly recognized based on color, but this is dependent on the ambient light. Tian and Slaughter [9] designed a new segmentation method that takes variable light conditions into account, resulting in better segmentation. Besides color and shape, the row pattern of crops can be used to distinguish crop and weeds [5]. More sophisticated shape analysis includes active shape modelling [10], where types of leaf shapes are defined and used for classification. Preliminary tests show 80% correct classification with this method [10]. Overlapping leaves are still a problem for shape analysis, since the shapes of overlapping leaves cannot be correctly defined in the image.

Chlorophyll fluorescence also exhibits species-specific patterns that can be used to discriminate between species and thereby to detect weeds. Chapelle et al. [11] demonstrated the potential for species recognition with laser-induced fluorescence. Fluorescence signals of green vegetation have low intensity in comparison to normal reflection and are thus difficult to measure. This makes it difficult to develop a robust system for outdoor conditions. However, Kebabian et al. [12] described a passive system that can measure fluorescence in the oxygen absorption band around 760 nm, where there is no ambient light and no reflection. The performance of this system for weed detection is not yet documented.

Spectral methods for weed detection have been investigated, for different numbers of spectral wavebands and different spectral band widths (spectral resolution) [13-15]. Wang et al. [16] used reflection in 5 bands with constant illumination and achieved 80% correct weed classification. Line imaging spectrography with higher spectral resolution was tested by Borregaard et al. [14], for detection of weeds, potato, and sugar beet in the range of 600 nm to 1060 nm with constant illumination. The imaging spectrograph performed well, with above 90% correct classification. Vrindts et al. [13] tested similar equipment in the 420 to 830 nm range in field conditions, with natural light. Classification results were similar (above 90% correct classification), provided the classification model is adapted to the light condition. Moshou et al. [17] found that neural networks classification yielded better results than classical multivariate discriminant techniques on crop and weed spectral data.

Currently, the changing light conditions are still a challenge in the development of a robust weed detection system. Measuring incident light and using classification models that are adapted to specific light conditions could help solve this problem. Advanced processing like sensor fusion and neural networks can improve the results of weed detection. The necessary high quality of the images and high computation time is a drawback of using computer vision for weed detection, but continuing development of cameras and hardware will enable applications like on-line weed detection.

#### 5.6.3 Disease and Stress Detection

Since nutrient stress, diseases, and pests cause a visible deterioration of the crop, optical techniques can be used to determine their spread in the field. The general response of reflectance spectra to different kinds of stresses have been investigated by

[18] and [19]. Spectral reflectance characteristics of leaves were shown to be highly correlated to their chemical composition. Lorenzen and Jensen [20], Polischuk et al. [21], and Sasaki et al. [22] succeeded in using spectral disease detection on (respectively) barley, tomato, and cucumber leaves. Masoni et al. [23] pointed out the effect of different mineral stresses on separate leaf spectral properties. In-situ plant nitrogen status could be estimated successfully by line-imaging spectrography [24], with a prediction error around 10% for chlorophyll content. Pre-mapping of diseases and stresses could also be achieved using air-borne systems. Spatial resolutions down to a few meters are possible from satellites and to below 1 m from aircraft [25]. Current commercial satellite sensing is probably not suitable for early disease detection (even if the wavelengths at which data are collected were suitable) because of limitations in spatial resolution. At best, satellite images can be useful by highlighting relatively large areas of disease or other stresses in a crop, which can then be checked by the farmer. In addition, revisit time and variability in cloud cover could mean that even this simple information may not be available when required. Aircraft-mounted systems do not have these constraints and could be used when required. However, data acquisition equipment would likely have to be faster, more sophisticated, and more expensive than for terrestrial vehicle-mounted systems.

Three types of measurement are useful for optically based disease detection systems: reflectance, fluorescence, and thermal sensing. There are several techniques available for each type of measurement: spectrophotometry, spectral line imaging, and multispectral imaging for reflection measurements; fluorescence kinetics, spectrometry and imaging for fluorescence; and, thermoradiometry and thermography for thermal sensing.

# Sensing Based on Reflectance

Methods based on reflectance rely on making measurements simultaneously in one or more wavebands. Detail of the measured spectra (spectral resolution) can vary from narrow (0.5-5 nm) to broad (20-100 nm) wavebands depending on the type of instrument used. If the field of view is large, the sensitivity of spectrophotometers may be compromised as diseased areas may represent only a small part of the reflection combined with reflection from healthy leaves and soil. It then may be difficult to discriminate between diseased and healthy areas of the crop.

To make the best use of reflectance spectral measurements, they need to be normalized to account for variations in illumination [26]. Thus, practical reflectance measurement systems should include a method of monitoring the spectral content of incoming illumination [27].

Discrimination between diseased and healthy areas of the crop depends on identifying differences in the spectra reflected by diseased and healthy tissue. To avoid large amounts of data handling, it is necessary to identify wavebands or combinations of wavebands that can discriminate between diseased and healthy plants. An algorithm to identify a suitable set of wavebands from a large range was used to interpret images acquired in a wide range of field conditions. The results compared well with visual assessment of disease severity in the same plots [28,29]. If a relatively small number

of wavebands can be used to detect the presence of disease, detection equipment could be simplified (e.g., using filters instead of spectrographs), thus reducing costs.

Reflectance spectra not only depend on the spectral quality of the ambient illumination, but also on the angle between the illumination and the direction of view of the sensor. Interpretation algorithms may need to account for solar angle and proportions of direct to diffuse radiation. Additionally, the viewing angle of the optical sensors must consider the impact of reflections from background soil, which are transmitted through crop leaves and mix with reflections from the canopy. Another consideration is that due to pathogen incubation periods, the top leaf often has no disease symptoms when it has only recently unfolded.

#### Sensing Based on Fluorescence

Disease detection based on fluorescence requires an excitation source as well as equipment to detect the fluorescence signal. Excitation sources can be UV lamps or lasers. Fluorescence images can be recorded using digital cameras fitted with an appropriate filter (e.g., a single band at 690 nm) or multispectral cameras if more than one fluorescence waveband were used. Fluorescence signals can easily be swamped by background ambient illumination. This can be overcome using pulsed sources coupled with synchronized gated detectors. Pulsed gated systems have been used for field monitoring in daylight, by integrating the fluorescence signal over some hundred excitation exposures [30,31]. Faster measurement systems, more suitable for moving platforms (e.g., tractors, airplanes), have been designed to detect single-point canopy fluorescence features both in terms of spectral content [32,33] and time-resolved intensity [34,35]. Although techniques have been developed that can detect chlorophyll fluorescence in the field, with potential to identify pre-visual disease symptoms, the complexity of equipment limits their use in practice. Developments in remote-sensed kinetic fluorescence techniques, particularly in very rapid fluorescence measurement, will be of particular interest for field applications, as they are insensitive to atmospheric and ambient light variations.

#### Thermal Sensing

Thermal imaging offers potential for disease detection [36,37], but equipment is very expensive, limiting its use in practical and low-cost vehicle-mounted systems.

## 5.6.4 Information Handling for Crop Protection

#### Multisensor Fusion

As noted above, each type of disease detection system has its own technical problems and interpretation difficulties. An alternative approach is to use more than one sensor technology and to integrate measurements in order to obtain a more sensitive and discriminating system than could be obtained using a single sensor.

Multisensor data fusion systems combine data from multiple sensors to perform inferences that may not be possible from a single sensor alone [38,39]. Data fusion is analogous to the ongoing cognitive process used by humans to integrate data continually from their senses to make inferences about the external world. Development issues for data fusion systems depend on the phenomena that are observed, the type of

sensors utilized, and the inferences sought. Generally, applications aimed at higher-level inferences require techniques from the artificial intelligence domain such as expert systems, template matching, neural networks, and fuzzy logic.

A fundamental issue is the architecture selection [39]. The issue revolves about the question of where to fuse the data in the processing flow of two or more sensors. Three basic multisensor fusion architectures exist:

- Centralized fusion architecture—This uses raw data from multiple sensors to determine the identity of the observed object. It is useful in the case that the sensors observe the same physical manifestation of an object (e.g., infrared and visual images).
- Centralized fusion with feature vectors—Pre-processing is applied to each sensor reading in order to extract a feature vector. These feature vectors are concatenated in a single vector and are used as input to a classification algorithm.
- Autonomous fusion—In this approach the output from each sensor is a decision (i.e., declaration of identity). Smart sensors can be integrated in this kind of fusion.

The optimal architecture for disease and stress detection would be one that results in a minimum information loss and is fast enough to be implemented in real time. Centralized fusion with feature vectors seems the most appropriate because it satisfies both requirements of accuracy and speed.

# Artificial Neural Networks

Artificial neural networks are general mapping devices that resemble the function of their biological counterparts [40]. Through the use of optimization-based learning algorithms a neural network can be trained from examples by adjusting an internal set of weights. After training, the neural network is able to generalize over unexpected situations (i.e., data that it did not encounter before) [38]. The main advantages of neural networks are that they are model-free and universal approximators.

# Multilayer Perceptrons (MLP)

Feed-forward neural networks [41] provide a general framework for representing non-linear functional mappings between a set of input and output variables. This is achieved by representing the non-linear function of many variables in terms of composition of non-linear functions of a single variable, which are called activation functions. Some commonly used activation functions include the logistic sigmoid and the hyperbolic tangent. The training of an MLP consists of finding the minimum of an energy function with respect to the weights of the network. The optimal weights produce the globally minimal value of the energy function. A variety of optimization procedures can be used to find the weights that minimize the energy function.

# Self-Organizing Map

The *self-organizing map (SOM)* [42] is a neural network that maps signals from a high-dimensional space to a one- or two-dimensional discrete lattice of neuron units. The map preserves topological relationships between inputs in a way that neighboring inputs in the input space are mapped to neighboring neurons in the map space.

A way of using the SOM to find correlations between the data is to label the neurons of the SOM using a different set than the training set and finding the best-matching units (BMUs) for every example in the testing set or labelling set. Some of the neurons that are selected most frequently by examples of one class are labelled based on a voting procedure. These neurons are then able to estimate the class of a new example presented to the SOM by calculating the euclidean distance of the example vector to the codebook vector of each neuron and finding the BMU. The label of the BMU is then the estimated class of the new example vector. The neurons that are not labelled constitute the borders of the classes and show the degree at which example observations from one class can be misclassified as observations belonging to another class.

In a case study, Moshou et al. [43] attempted to discriminate disease stress from nutrient-deficiency stress in field conditions using spectral reflectance information. Yellow rust-infected winter wheat plants were compared to nutrient-stressed and healthy plants. In-field hyperspectral reflectance images were taken with an imaging spectrograph. Leaf recognition was possible using a Normalized Difference Vegetation Index (NDVI) threshold on the high spatial resolution image. Diseased and nutrient-stressed canopies show lower NDVI values. This can be partly explained by the lower chlorophyll activity in the stressed plants. A normalization method based on reflectance and light intensity adjustments was applied. For achieving high performance stress identification, SOMs were introduced. Winter wheat infected with yellow rust was successfully recognized from nutrient-stressed and healthy plants. Overall performance using 5 wavebands was more than 99% for detection and identification of (disease or nutrient) stress. Figure 1 shows typical spectral responses for healthy, diseased, and nitrogen-stressed winter wheat. Figure 2 represents the classification of healthy, nutrient-stressed, and diseased samples by the neural network of the case study.

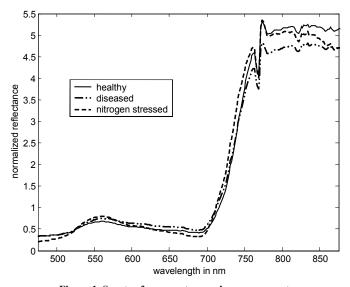


Figure 1. Spectra from spectrograph measurements.

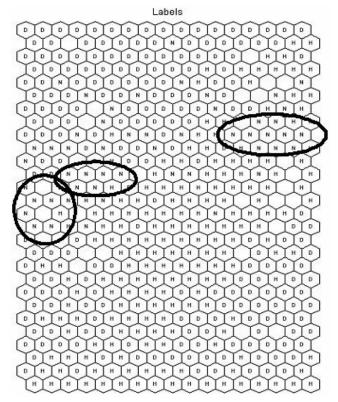


Figure 2. SOM-based identification of spectra from diseased, nitrogen-deficient, and healthy plants. A new spectrum is assigned to one of the shown units and is identified using the label of the unit it has been assigned to (D = diseased, H = healthy, N = nitrogen-deficient).

Ellipsoids indicate identified spectra from nitrogen-deficient canopies.

#### 5.6.5 Deciding on Crop Protection Treatments

For weed treatments, the choice of herbicide or other treatment depends on the weed species and severity of the infestation. Low weed populations may be tolerated, but future weed problems caused by seed production must also be considered in the decision on the weed treatment. Optimizing necessary weed treatments can be based on historical information: weed density-yield response curves and herbicide dose response curves. Product guidelines for correct herbicide use are available on packaging or from manufacturers, denoting the necessary dose for different circumstances. Weed management can be improved by using crop-weed competition models and weed population models to determine the best long-term strategy [44].

Product guidelines, determined with extensive tests, give clear instructions on the use of agrochemicals. For controlling pests and diseases, a preventive treatment may be necessary, but the use of agrochemicals should be justified. Regional information networks can alert farmers on the presence of diseases and pests, and trigger the cor-

rect treatments on fields in the area. Again, models can be used to improve pest and disease control, for example using infection level, climatic information, and disease development models to decide on crop protection measures (see the EPIPRE system by Zadoks [45]). One important question that models can help to answer is: Is the treatment justified for the probable yield loss by the infection?

Vast practical experience and knowledge is necessary for development of decision support for the management of weeds, pests and diseases, so that treatments are adjusted to the local crops, soils, climate and infestations.

#### References

- 1. Felton, W. L. 1995. Commercial progress in spot spraying weeds. *Brighton Crop Protection Conference—Weeds—1995.* 3: 1087-1096. UK: British Crop Protection Council.
- 2. Ahrens, W. H. 1994. Relative costs of a weed-activated versus conventional sprayer in northern great plains fallow. *Weed Technology* 8: 50-57.
- 3. Brown, R. B., J.-P.G. A. Steckler, and G. W. Anderson. 1994. Remote sensing for identification of weeds in no-till corn. *Trans. ASAE* 37(1): 297-302.
- 4. Goel, P. K., S. O. Prasher, R. M. Patel, D. L. Smith, and A. DiTommaso. 2002. Use of airborne multi-spectral imagery for weed detection in field crops. *Trans. ASAE* 45(2): 443-449.
- 5. Pérez, A. J., F. López, J. V. Benlloch, and S. Christensen. 2000 Colour and shape analysis techniques for weed detection in cereals fields. *Computers and Electronics in Agriculture* 25: 197-212.
- 6. Woebbecke, D. M., G. E. Meyer, K. Von Bargen, and D. A. Mortensen. 1995. Shape features for identifying young weeds using image analysis. *Trans. ASAE* 38(1): 271-281.
- 7. Zhang, N., and C. Chaisattapagon. 1995. Effective criteria for weed identification in wheat fields using machine vision. *Trans. ASAE* 38(3): 965-974.
- 8. Hemming, J., and T. Rath. 2001. Computer-vision-based weed identification under field conditions using controlled lighting. *J. Agricultural Engineering Research* 78(3): 233-243.
- 9. Tian, L. F., and D. C. Slaughter. 1998. Environmentally adaptive segmentation algorithm for outdoor image segmentation. *Computers and Electronics in Agriculture* 21: 153-168.
- 10. Søgaard, H. T., and T. Heisel. 2002. Weed classification by active shape models. EurAgEng Budapest 2002. EurAgEng Paper Nr. 02-AE-004.
- Chappelle, E. W., F. M. Jr. Wood, J. E. McMurtry III, and W. W. Newcomb. 1984. Laser-induced fluorescence of green plants 1: A technique for the remote detection of plant stress and species identification. *Applied Optics* 23(1): 134-138.
- 12. Kebabian, P. L., A. F. Theisen, S. Kallelis, H. E. Scott, and A. Freedman. 1999. Passive two-band plant fluorescence sensor with applications in precision agriculture. *Proc. of the Precision Agriculture and Biological Quality*

- Conference, eds. G. E. Meyer, and J. A. DeShazer, 3543: 238-245. Bellingham, WA: SPIE.
- 13. Vrindts, E., J. De Baerdemaeker, and H. Ramon. 2002. Weed detection using canopy reflection. *Precision Agriculture* 3(1): 63-80.
- 14. Borregaard, T., H. Nielsen, L. Nørgaard, and H. Have. 2000. Crop-weed discrimination by line imaging spectroscopy. *J. Agricultural Eng. Research* 75: 389-400.
- 15. Noble, S. D., and T. G. Crowe. 2001. Plant discrimination based on leaf reflectance. ASAE Paper No. 01-1150. St. Joseph, MI: ASAE.
- 16. Wang, N., N. Zhang, F. Dowell, and D. Peterson. 2001. Design of an optical weed sensor using plant spectral characteristics. *Trans. ASAE* 44(2): 409-419.
- 17. Moshou, D., E. Vrindts, B. De Ketelaere, J. De Baerdemaeker, and H. Ramon. 2001. A neural network based plant classifier. *Computers and Electronics in Agriculture* 31(1): 5-16.
- 18. Carter, G. A., and A. K. Knapp. 2001. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *American J. Botany* 88(4): 677-684.
- 19. Cibula, W. G., and G. A. Carter. 1992. Identification of a far-red reflectance response to ectomycorrhizae in slash pine. *International J. Remote Sensing* 13: 925-932.
- 20. Lorenzen, B., and A. Jensen. 1989. Changes in spectral properties induced in Barley by cereal Powdery Mildew. *Remote Sensing Environment* 27: 201-209.
- Polischuk, V. P., T. M. Shadchina, T. I. Kompanetz, I. G. Budzanivskaya, and A. A. Sozinov. 1997. Changes in reflectance spectrum characteristic of Nicotiana debneyi plant under the influence of viral infection. *Archives of Phytopathology and Plant Protection* 31(1): 115-119.
- Sasaki, Y., T. Okamoto, K. Imou, and T. Torii. 1998. Automatic diagnosis of plant disease-Spectral reflectance of healthy and diseased leaves. *Proc. of 3<sup>rd</sup> IFAC/CIGR Workshop on Artificial Intelligence in Agriculture*, Makuhari, Chiba, 158-163.
- 23. Masoni, A., E. Laura, and M. Mariotti. 1996. Spectral properties of leaves deficient in iron, sulphur, magnesium and manganese. *Agronomy J.* 88(6): 937-943.
- 24. Dumont, K., and J. De Baerdemaeker. 2001. In field wheat nitrogen assessment using hyperspectral imaging techniques. *Proc. of the Third European Conference on Precision Agriculture*, 905-910, eds. G. Grenier, and S. Blackmore, Montpellier, France: Agro Montpellier (ENSAM).
- 25. Blakeman, R. H., R. J. Bryson, and P. Dampney. 2000. Assessing crop condition in real time using high resolution satellite imagery. *Aspects of Applied Biology* 60, Remote Sensing in Agriculture, 163-71. Wellesbourne, UK: The Association of Applied Biologists.
- 26. Price, J. C. 1994. How unique are spectral signatures? *Remote Sens. Environ.* 49: 181-186.

- 27. Borel, C. C., and S. A. W. Gerstl. 1994. Are leaf chemistry signatures preserved at the canopy level? *Proc. Int. Geoscience and Remote Sensing Symposium*, 996-998.
- Bravo, C., D. Moshou, R. Oberti, J. West, A. McCartney, L. Bodria, and H. Ramon. 2002. Detection of foliar disease in the field by the fusion of measurements made by optical sensors. ASAE Paper No. 023087. St. Joseph, MI: ASAE.
- Moshou, D., C. Bravo, J. S. West, S. Wahlen, H. A. McCartney, and H. Ramon. 2004. Automatic detection of yellow rust in wheat using reflectance measurements and neural networks. *Computers and Electronics in Agriculture* 44: 173-178.
- 30. Johansson, J., M. Andersson, H. Edner, J. Mattson, and S. Svanberg. 1996. Remote fluorescence measurements of vegetation spectrally resolved by multicolour fluorescence imaging. *J. Plant Physiol.* 148: 632-637.
- 31. Saito, Y., K. Hatake, E. Nomura, T. D. Kawahara, and A. Nomura. 1997. Range-resolved image detection of laser-induced fluorescence of natural trees for vegetation distribution monitoring. *Jpn. J. Appl. Phys.* 36: 7024-7027.
- 32. Corp, L. A., J. E. McMurtrey, E. W. Chappelle, C. S. T. Daughtry, and M. S. Kim. 1997. UV band fluorescence (in vivo) and its implications for the remote assessment of nitrogen supply in vegetation. *Remote Sens. Environ.* 61: 110-117.
- 33. Ludeker, W., H. G. Dahn, and K. P. Gunther. 1996. Detection of fungal infection of plants by laser-induced fluorescence: An attempt to use remote sensing. *J. Plant Physiol.* 148: 579-585.
- Flexas, J., J. M. Briantais, Z. Cerovic, H. Medrano, and I. Moya. 2000. Steadystate and maximum chlorophyll fluorescence responses to water stress in grapevine leaves: A new remote sensing system. *Remote Sens. Environ.* 73: 283-297.
- 35. Morales, F., R. Belkhodja, Y. Goulas, J. Abadia, and I. Moya. 1999. Remote and near-contact chlorophyll fluorescence during photosynthetic induction in iron-deficient sugar beet leaves. *Remote Sens. Environ.* 69: 170-8.
- 36. Chaerle, L., W. Van Caeneghem, E. Messens, H. Lambers, M. Van Montagu, and D. Van Der Straeten. 1999. Presymptomatic visualization of plant-virus interactions by thermography. *Nat. Biotechnol.* 17: 813-816.
- 37. Lili, Z., J. Duchesne, H. Nicolas, and R. Rivoal. 1991. Détection infrarouge thermique des maladies du blé d'hiver. *Bulletin OEPP/EPPO Bulletin* 21: 659-672.
- 38. van Dam, J., B. Krösse, and F. Groen. 1996. Neural network applications in sensor fusion for an autonomous mobile robot. *Reasoning with Uncertainty in Robotics*, eds. L. Dorst, M. van Lambalgen, and F. Voorbaak, 263-277. New York, NY: Springer-Verlag.
- 39. Hall, D. L. 1992. *Mathematical Techniques in Multisensor Data Fusion*. Boston, MA-London, UK: Artech House.
- 40. Haykin, S. 1994. *Neural Networks—A Comprehensive Foundation*. New York, NY: Macmillan College Publishing Company.

- 41. Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. Learning internal representations by error propagation. *Parallel Distributed Processing*, Vol. 1, eds. D. E. Rumelhart, and J. L. McClelland, 318-362. Cambridge, MA: MIT Press.
- 42. Kohonen, T. 1995. Self-Organizing Maps. Berlin, Germany: Springer-Verlag.
- 43. Moshou, D., C. Bravo, S. Wahlen, J. West, A. McCartney, J. De Baerdemaeker, and H. Ramon. 2003a. Simultaneous identification of plant stresses and diseases in arable crops based on a proximal sensing system and self-organizing neural networks. *Proc. of the Fourth European Conference on Precision Agriculture* (4ECPA Berlin), 425-432. Wageningen Academic Publishers.
- 44. Paice, M. E. R., W. Day, L. J. Rew, and A. Howard. 1998. A stochastic simulation model for evaluating the concept of patch spraying. *Weed Research* 38: 373-88.
- 45. Zadoks, J. C. 1989. EPIPRE: A computer-based decision support system for pest and disease control in wheat: its development and implementation in Europe. *Plant Disease Epidemiology*, Vol. 2: Genetics, Resistance and Management, ed. G. R. Conway Wiley, 344-351. Chichester, UK.

# 5.7 Application Techniques for Crop Protection

J. Anthonis, J. De Baerdemaeker, and H. Ramon

Abstract. In this section, different issues of spray application are discussed. The first part discusses spray boom motions in the vertical plane and means to automatically adjust the sprayer to the slope of the field. Subsequently the importance of horizontal spray boom motions is highlighted and solutions to reduce these vibrations are suggested. For precision spraying, there is a need to adjust liquid flow and liquid pressure separately; this is also discussed. Finally, information systems for spraying are presented.

**Keywords.** Boom stabilization, Automatic adjusting system, PWM nozzle, Sprayer guidance, Spray application, Spray boom.

#### 5.7.1 Introduction

Efficient distribution of chemical agents is highly correlated with the uniformity of spray coverage in the canopy [1]. Experimental work and simulation studies point out that undesired spray boom motions give rise to an uneven spray deposition pattern, resulting in under- and over-application of spray deposit ranging from 0% to 800% of the desired (100%) spray distribution [2-5].

The most important spray boom motions affecting the uniformity of the spray pattern are *boom rolling*, which results in boom motions in the vertical plane, and *yawing* and *jolting*, which give rise to vibrations in the horizontal plane. Each of these vibrations is shown in Figure 1 and their effect on the spray distribution pattern is depicted.

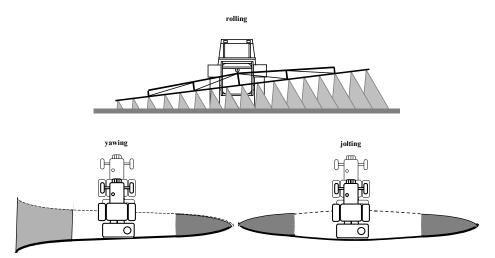


Figure 1. Most important spray boom vibrations and their effect on the spray distribution pattern.

Increasing the boom stability improves the homogeneity of the spray coverage and raises the efficiency of pesticides.

Targeted spraying is an additional important argument for suppressing vertical and horizontal boom vibrations. Nozzles approaching the soil too close generate a very small spray cone and as such become unable to reach certain patches in the field (see Figure 1). In on-line optical weed detection, integration of a defined number of the most recent horizontal strips of pixels from the optical sensor provides information for the decision algorithm that activates the controller of the spray system [6]. Excessive horizontal boom vibrations allocate wrong pixel strips (i.e., wrong information) to the decision algorithm and the spray controller.

#### 5.7.2 Vertical Stabilization

Most current agricultural field sprayers are equipped with a vertical pendulum system to attenuate boom roll. The suspension tries to keep the boom perpendicular to gravity by isolating the boom from vibrations of the tractor or trailer, induced by soil unevenness. The suspension acts as a low-pass filter and adequately filters off high-frequency roll of the tractor. Unfortunately, at low frequencies the natural frequency of the suspension amplifies tractor roll. Since spray booms are weakly damped structures, during field operation this amplification causes excessive but very slow rolling motions of the boom, which are difficult to stabilize. As the pendulum tries to keep the boom perpendicular to gravity, special locking or adjusting mechanisms are needed to keep the boom parallel to the soil on hilly fields [7].

Since the performance of the passive pendulum suspension is satisfactory at high frequencies, i.e. beyond the natural frequency of the pendulum, to save energy an automatic adjusting mechanism should only be active at low frequencies. For that reason, it is called a *slow active system*. The automatic adjusting mechanism is conceived

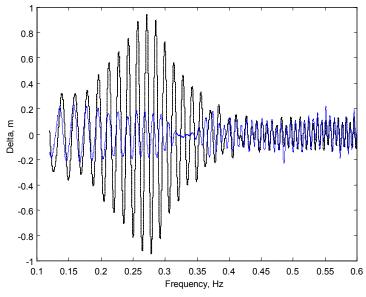


Figure 2. Movement of the boom tip by applying a swept sine disturbance signal to the suspension. The dashed line is without controller; the full line (lower amplitude) is with slow active controller.

of a hydraulic ram as actuator, two ultrasonic sensors to measure the position of the boom with respect to the field, and a controller. Figure 2 shows that the controller acts at low frequencies by adding damping to the passive system. Beyond the natural frequency of the pendulum it does not interfere with the passive suspension [8,9].

#### 5.7.3 Horizontal Stabilization

For horizontal spray boom stabilization, there is no reference, such as the soil or an external force such as gravity, to position the boom parallel to the field. Furthermore, experimental and theoretical research has shown that yawing and jolting of the boom are more critical with respect to an uneven spray pattern than rolling boom motions. Ramon and De Baerdemaeker [2] calculated that as long as the distance between the spray nozzles and the soil or crop does not exceed 50% of the desired distance, variation in spray distribution pattern due to boom roll will remain between 85% and 140% of the desired distribution. In the horizontal direction, however, even small vibrations at the boom tips with an amplitude around 30 cm can cause overdoses of three times the desired dose [2]. Current designs are restricted to the inclusion of rubber blocks between the boom and the frame fixed to the tractor. In one case, a concept of an active horizontal suspension has successfully been developed and its effect simulated [10,11].

## 5.7.4 Appropriate Spray Equipment for Targeted Spraying

In selective spraying, the quality of the spray nozzles has an especially major influence on the spraying dynamics. Their opening and closing times must be as short as possible to minimize their contribution to dead time, time delay, rise time, and peak time of the hydraulic system. It is advisable that each nozzle operates independently to keep the spatial resolution as small as possible. The nozzles must also be safe to operate, implying a long life cycle and a correct dosage. Their impact on the pressure in the hoses must be as small as possible with a view to keeping the droplet spectrum stable. The ability to change the flow rate through the nozzle without influencing the droplet spectrum would be an advantage.

In practice, with conventional spray nozzles, the flow rate of the spray is controlled by adapting the liquid pressure. However, droplet size distribution is inversely associated with fluid pressure. Solenoid and motor valves are mounted on a spray boom to lock boom sections and cannot be used for operating individual nozzles. During opening and closing, they create pressure variations in the hydraulic equipment, which are difficult to compensate. Their rise time is high and can increase to 15 seconds for motor valves, pointing to unacceptable slow dynamics. In addition, the safety of operation of these valves is questionable. From this, it may be concluded that solenoid and motor valves are better avoided in selective crop protection.

In this respect, *pulse width modulated (PWM)* nozzles offer new possibilities in selective spraying. Each PWM nozzle (see Figure 3) is supplied with an electromagnetic solenoid valve, which enable nozzles to open and close individually in a short and fixed time called the cycle time. When a voltage is applied to the solenoid, the originated magnetic force pulls the ferromagnetic core upwards and opens the nozzle. When the voltage drops, the spring pushes the core back to its original position and closes the nozzle. The ratio between the on position (i.e., duty cycle) and the off position determines the flow rate through the nozzle, which can vary with a factor of 10

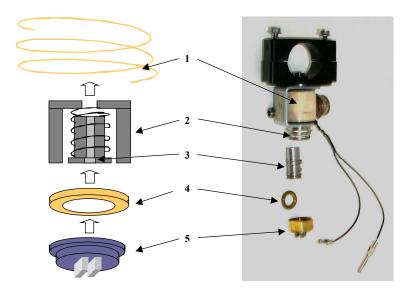


Figure 3. Construction of a PWM nozzle (1, solenoid; 2, hollow cylinder; 3, armature and spring; 4, gasket; 5, spray tip).

without changing the droplet spectrum significantly if the pressure in the conduits remains stable during a variable flow rate through the nozzle. As electrical conduits are cheap and easy to install, each nozzle can easily be operated individually. In addition, Giles and Ben Salem [12] showed that PWM nozzles have very fast dynamics so their transient behavior after a new flow rate setting is negligible. Current commercially available nozzles have a cycle time of 0.1 s or cycle frequency of 10 Hz. However, theoretical studies supported by experiments [13] proved that, with a cycle frequency of 10 Hz, spray liquid is released in stripes, especially when the duty cycle is small. Within the framework of the research, the cycle frequency of PWM nozzles could be increased to 25 Hz, which was satisfactory.

## 5.7.5 Information Technology for Sprayers

Modern high-end sprayers are equipped with a sophisticated computer system. Once the borders and obstacles of the field are known, a guidance system using a positioning system (such as GPS) guides the sprayer through the field. When the map of the field already exists, it can be directly downloaded. Updates with new or disappeared obstacles can be made. The system reports on a screen the progress of the work and warns for hazards. With these information systems overlaps are avoided. Given the actual available boom widths up to 40 m, overlaps are very likely to occur.

These modern computer systems also record the pressure and flow rate in the hydraulic system, the number of hectares sprayed, the distance travelled on the road, etc. In the future, the combination of guidance systems with information about the spray operation can provide spray dose maps that can be related to future weed maps to get an idea about the effectiveness of past treatments. From this knowledge, trends in weed development can be derived and more effective treatment schemes can be set up. Spray dose maps are also of great value for quality labelling of agricultural products.

Nowadays, booms with active vertical suspensions, which follow the slope of the field, are appearing. Information provided by sensors that measure the height above the canopy can also be recorded. Even information from vibration sensors measuring boom vibrations could be logged, which would provide very useful information for boom manufacturers to optimize their suspensions.

### References

- 1. Göhlich, H. 1985. Deposition and penetration of sprays. *BCPC monogram No 28, Symposium on Application and Biology,* 172-182.
- 2. Ramon, H., and J. De Baerdemaeker. 1997. Spray boom motions and spray distribution Part 1: Derivation of a mathematical relation. *J. Agricultural Engineering Research* 66: 23-29.
- 3. Ramon, H., B. Missotten, and J. De Baerdemaeker. 1997. Spray boom motions and spray distribution Part 2: Experimental validation of the mathematical relation and simulation results. *J. Agricultural Engineering Research* 66: 31-39.
- 4. Langenakens, J., H. Ramon, and J. De Baerdemaeker. 1995. A model for measuring the effect of tire pressure and driving speed on horizontal sprayer boom movements and spray pattern. *Trans. ASAE* 38(1): 65-72.

- 5. Nation, H. J. 1982. The dynamic behaviour of field sprayer booms. *J. Agricultural Engineering Research* 27: 61-70.
- Ramon, H., J. Anthonis, E. Vrindts, R. Delen, J. Reumers, D. Moshou, K. Deprez, J. De Baerdemaeker, F. Feyaerts, L. Van Gool, R. De Winne, and R. Van den Bulcke. 2002. Development of a weed activated spraying machine for targeted application of herbicides. *Aspects of Applied Biology* 66: 147-164.
- 7. Anthonis, J., K. Deprez, M. Lannoije, H. Van Brussel, and H. Ramon. 1999. Mathematical modelling and comparison of several passive vertical spray boom suspensions. *Third International Symposium on Mathematical Modelling and Simulation in Agricultural and Bio-Industries*, Uppsala (Sweden), 35-41.
- 8. Deprez, K., J. Anthonis, H. Ramon, and H. Van Brussel. 2002a. Development of a slow active suspension for stabilizing the roll of spray booms Part 1: Hybrid Modelling. *Biosystems Engineering* 81(2): 185-191.
- 9. Deprez, K., J. Anthonis, H. Ramon, and H. Van Brussel. 2002b. Development of a slow active suspension for stabilizing the roll of spray booms Part 2: Grey Box Modelling. *Biosystems Engineering* 81(3): 273-279.
- Anthonis, J. 2000. Design and development of an active horizontal suspension for agricultural spray booms. PhD thesis. Leuven, Belgium: Katholieke Universiteit Leuven.
- 11. Anthonis, J., H. Ramon, and J. De Baerdemaeker. 2000. Implementation of an active horizontal suspension on a spray boom. *Trans. ASAE* 43(2): 213-220.
- 12. Giles, D. K., and E. Ben-Salem. 1992. Spray droplet velocity and energy in intermittent flow from hydraulic nozzles. *J. Agricultural Engineering Research* 51: 101-112.
- 13. Delen, R, L. Clijmans, J. Anthonis, and H. Ramon. 2004. A non-linear model to approximate the dynamics of a pulse width-modulated spray nozzle. *Mathematics and Computers in Simulation* 65(1-2): 39-48.

# 5.8 Special Aspects of IT for Greenhouse Cultivation

K. P. Ferentinos, K. G. Arvanitis, H. J. Tantau, and N. Sigrimis

Abstract. This section presents the recent and near-future uses of information technology (IT) for greenhouse cultivation. After some introduction on the basic physical mechanisms that govern greenhouse cultivation systems, two different approaches to greenhouse management and control are analyzed, the horizontal aspect and the vertical aspect. Then, low-level control, medium-level control and management, and high-level management is discussed. An additional section categorizes and presents some of the latest tools and complete products for greenhouse control and integrated management.

**Keywords.** Greenhouse cultivation, Information technology, Control, Integrated production management.

#### 5.8.1 Introduction

Advances in information technology (IT) during the last decades have been applied to greenhouse cultivation, meeting the need for uniform year-round plant production. Plant cultivation in a controlled environment, such as that of greenhouses, is a very complicated process with numerous parameters that can directly or indirectly affect productivity. For these parameters to be controlled, all physical phenomena of the greenhouse environment have to be analyzed to calculate energy and mass balances. Feedback control relies only on real-time measurements, but for optimal control and better management, complete models of the physical [1,2] and biological [3-7] systems are sought. Physical systems are well-defined and have long been elaborated while biological systems are more complex and uncertain. Efforts in biophysical modeling have only recently reached a practical utilization stage [8] and have a long way to go to become a mature coupling of bioscience and technology.

However, the societal requirements for environmental respect and the consumer demands for quality, under global pricing competition, adds new dimensions and constraints in optimal management of a viable business. The driving force of integrated production management provides both the reason and the means for advances in this field. Bio-models (models concerning insects, disease, production, etc.) and IT implementations will need to reach new levels of achievement to become reliable and to be considered as necessary inputs to the production process. Cultivation technologies (hydroponics, robotic harvesters, plant factories, etc.) become mature and less costly as they gain widespread acceptance and this drives the needs for a knowledge-rich IT as we move from the information age to a knowledge-driven society. Efforts have started based on modern communication technologies to provide the missing bridge from the expert teams or knowledge bases to the low-level controllers of the production side [10].

The understanding of transport mechanisms leads to the estimation of energy and mass balances of the greenhouse system, where three main transport mechanisms can be distinguished:

- *conduction*, which takes place through the construction and the cover and in a large degree through the soil (heat only);
- convection, which takes place between the greenhouse air and the internal surfaces, like heating pipes, cover, plants, soil surface, etc., and between the outside air and the outer surfaces of the greenhouse as well as the inside depending on infiltration or ventilation (heat and mass); and
- *radiation*, which includes the transport of energy between the surfaces of all components inside the greenhouse by electromagnetic waves (light and heat).

Several components determine the energy inflow and outflow. For the greenhouse cover and structural part, energy inflow and outflow are composed of solar radiation, radiation exchange with the sky and the interior of the greenhouse, exchange by convection between the structural parts and the inside and outside air, and finally latent heat produced by condensation of water vapor inside the greenhouse. In the greenhouse air, energy is exchanged by convection with the structural parts and cover, the

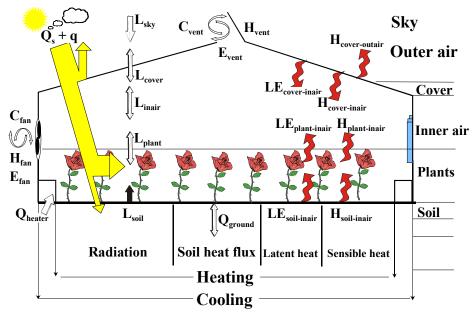


Figure 1. A schematic representation of physical processes of energy exchange in greenhouse crop production systems.

heating system, the soil and the plants, and of course with the outside air during ventilation. The soil exchanges energy by the absorption of solar radiation, radiation exchange with the greenhouse cover and its structural parts and the plants, convective exchange with the greenhouse air, and conductive exchange with the underlying soil layers. Finally, the plants absorb solar radiation; they exchange radiation with the cover and the structural parts of the greenhouse, the soil and the heating system; and they exchange energy by convection with the inside air and latent heat via evapotranspiration. The processes of thermal energy exchange among the greenhouse, the surroundings, and the greenhouse components are illustrated in Figure 1.

The thermal status of all greenhouse compartments is represented by the following differential equations:

$$\begin{split} Z_{c}C_{c}\frac{dT_{c}}{dt} &= \\ \left(Q_{s}a_{c} + qa_{q}\right) + \Delta L_{\text{cover}} - H_{\text{cover-outair}} - H_{\text{cover-inair}} - LE_{\text{cover-inair}} - LE_{\text{cover-outair}} \\ Z_{i}C_{i}\frac{dT_{i}}{dt} &= \\ \Delta L_{\text{inair}} + H_{\text{cover-inair}} + H_{\text{plant-inair}} + H_{\text{soil-inair}} - H_{\text{vent}} - H_{\text{fan}} + Q_{\text{heater}} \end{split}$$

$$(1a)$$

(1d)

$$Z_{p}C_{p}\frac{dT_{p}}{dt} =$$

$$(Q_{s}+q)\tau_{c}a_{p} + \Delta L_{plant} - H_{plant-inair} - LE_{plant-inair}$$

$$Z_{s_{o}}C_{s_{o}}\frac{dT_{s_{o}}}{dt} =$$

$$(Q_{s}+q)\tau_{c}\tau_{p}a_{s} + \Delta L_{soil} - H_{soil-inair} - LE_{soil-inair} - Q_{ground}$$

$$(14)$$

where  $T_c$ ,  $T_b$ ,  $T_p$  and  $T_{s_o}$  are temperatures of the cover, inner air, plant canopy and soil-surface, respectively.  $Z_x$  and  $C_x$  are the average height and thermal capacity of the compartment x.  $H_{x-y}$  and  $LE_{x-y}$  are the sensible heat and latent heat exchanges between compartment x and y. Because of the sign convention,  $H_{y-x} = -H_{x-y}$  and  $LE_{y-x} = -LE_{x-y}$ .  $Q_{heater}$  is the thermal energy input from heating, and  $Q_{ground}$  is the ground heat flux density between the top soil and subsequent soil layers.  $a_x$  and  $\tau_x$  are the absorption coefficient and transmittance coefficient of compartment x to the short wave radiation.  $a_c$  and  $a_q$  represent the absorption coefficient of glass to the direct and diffuse radiation respectively. For the soil compartment, only top soil temperature is described in Equation 1d. The soil heat flux density at the surface ( $Q_{ground}$ ) is:

$$Q_{ground} = 2k_{so} \left( \frac{T_{so} - T_{s_1}}{Z_{so} + Z_{s_1}} \right)$$
 (2)

which is determined by soil heat conductivity and temperature gradient at the top layer of soil.

The thermal status of other soil layers needs to be treated separately. Soil temperatures at various depths can be simulated with the following differential equation:

$$Z_{s_{j}}C_{s_{j}}\frac{dT_{s_{j}}}{dt} = 2k_{s_{j}}\left(\frac{T_{s_{j-1}} - T_{s_{j}}}{Z_{s_{j}} + Z_{s_{j-1}}} + \frac{T_{s_{j+1}} - T_{s_{j}}}{Z_{s_{j}} + Z_{s_{j+1}}}\right)$$
(3)

where *j* represents the *j*th sub-layer of soil

 $T_s$  = soil temperature

 $C_s$ = volumetric specific heat

 $k_s$  = thermal conductivity

 $Z_s$  = thickness of soil layer

The main measurable variables in a greenhouse environment are, for the aerial environment, temperature, relative humidity, light intensity, and carbon dioxide. For the root microenvironment they are pH, electrical conductivity (EC), soil temperature and moisture, salinity, and nutrient concentrations. The goal of greenhouse cultivation is the achievement of specific set-points for all these parameters, according to the appropriate desired values of the cultivated plants. This is carried out through some control and management operations in the greenhouse environment. These operations include, for the aerial environment, heating systems, ventilation and cooling systems, shading

screens, supplemental lighting systems, and techniques for CO<sub>2</sub> enrichment. For the root microenvironment, according to the type of the cultivation system (soil cultivation or hydroponics), heaters, pH and EC control systems, and more general hydroponics management systems. Recently, more advanced decision-support systems that take the salinity tolerance of the cultivated plants into account have been developed [9].

All these systems and operations are bounded in complicated ways, thus the accurate control of the greenhouse environment is a challenging task that requires sophisticated methodologies. This is the so-called *vertical aspect* of greenhouse control and management (Figure 2), which makes the required control methodologies very demanding. The other major aspect of greenhouse control and management is the *horizontal aspect* (Figure 3), which refers to the different time scales of the involved processes in a greenhouse cultivation system [10,11].

Greenhouse cultivation systems consist of two quite different parts: the physical part and the biological part. The physical part is formed by the environmental parameters both inside and outside of the greenhouse, while the biological part is basically the cultivated plants, as well as any biochemical reactions taking place between the crop and the environment (such as soil or substrates, insects, diseases). The physical part has many effects on the biological part and at the same time the biological system has numerous influences on the enclosing environment. Generally, the physical systems of plant production respond quickly, while the biological systems respond relatively slowly [12]. This makes the control and management of the greenhouse environment even more difficult and complex.

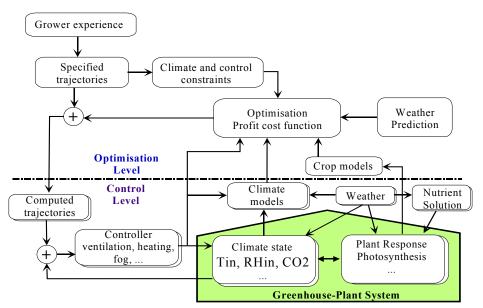


Figure 2. Vertical aspect of the greenhouse management and control problem.

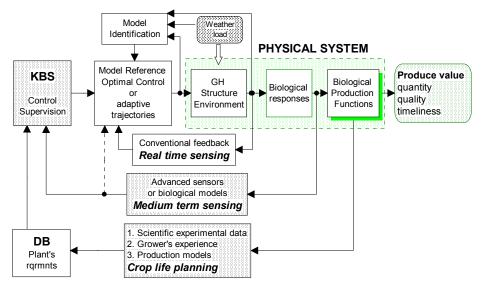


Figure 3. Multi-time-scale management and control of the greenhouse environment (horizontal aspect).

## 5.8.2 Low-Level Control Loops

The basis of greenhouse environment control consists of methods that use several aspects of IT to form low-level control loops [13-17]. They can be classified as classical control techniques or intelligent methodologies for real-time control.

#### Classical Control

In classical control, the systems to be controlled are considered as input-output systems. Inputs are usually control inputs and disturbances, while outputs are usually the variables to be controlled. In the greenhouse environment, control inputs can be the heating amount, the ventilation rate (window opening, speed of fans), the amount of supplemental lighting, the position of the shading screen, and the CO<sub>2</sub> enrichment rate. The outside temperature and humidity, the wind speed and direction, the solar radiation, and the outside CO<sub>2</sub> concentration are considered as disturbances. The outputs are the inside temperature, relative humidity, CO<sub>2</sub> concentration, and light intensity at plant level, i.e. the controlled variables.

The conventional control technique most widely used in greenhouse cultivation systems is feedback control. The controller is often of the simple ON/OFF type or the *proportional-integral-derivative (PID)* type. A PID controller has the ability to handle set-point changes, to compensate step-load disturbances, and to face wide model uncertainty [18]. To improve the management and control of a greenhouse process, an adaptive PID control strategy (Figure 4) may be applied to compute the optimal control signals used for a defined cost-performance function. Simpler versions of the PID controller have also been used in greenhouse environment control [13,19-21], which

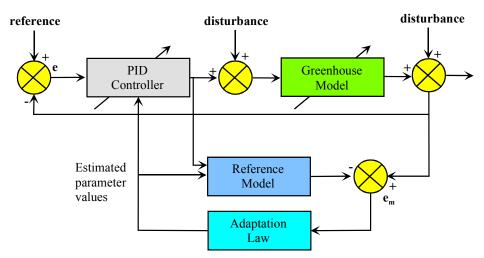


Figure 4. Structure of the adaptive PID controller.

has served greenhouse facilities for many years (at the start of feedback implementations) as switches (thermostats, hygrostats, pressostats). Because most greenhouse equipment is the binary-switch type, the application of such dynamic control methodologies is further complicated. In order for this type of equipment to be included in the dynamic control scheme, each of the dynamic equations for each possible state and control of the switching rate must be linearized, similar to pulse-width modulation [22]. A variation of the PID controller, the pseudo-derivative feedback algorithm (PDF) [23], has also been used successfully in temperature and humidity control of greenhouses [12,24]. Another control methodology, which originated by research in greenhouse environment control, is the proportional-integral-plus (PIP) controller [16], which has shown several advantages over the conventional PID or PI control, including robustness to pure-time transport delays, power and flexibility due to its state variable feedback, and a structure that avoids common control problems such as integral wind-up [25,26]. An improvement of the PID controller, the Smith predictor [27], compensates dead times that lower closed-loop stability margins and its performance. It has been used in greenhouse environment control with positive results [3]. Finally, an approach that leads to better temperature distribution and minimizes heat losses is that based on the nested control loop configuration and the load divider concept [28], which divides the input needs to the corresponding actuators, achieving better performance of the control system.

## Intelligent Real-Time Control

Over the last few years, IT has been playing a growing role in the development and materialization of greenhouse cultivation control systems. In particular, methodologies of IT in the area of artificial intelligence (AI) have been widely used to develop highly sophisticated intelligent systems for real-time control and management of greenhouse

facilities, where conventional mathematical control approaches do not easily apply [29]. Artificial neural networks (NNs) have been the most-used tool in intelligent control of both greenhouse environment and hydroponics. Their main advantage is that they do not require explicit evaluation of transfer coefficients or any model formulation. They are based on inherent learning capabilities of training data from the process to be modeled. Initially NNs were used in modeling the aerial environment of greenhouses, generally using as inputs the outside environmental parameters (temperature, humidity, solar radiation, wind velocity, etc.), the control variables, and the state variables, i.e., the conditions of the cultivated plants [30-32]. Simpler models that do not take into account the conditions of the plants have also been applied successfully in temperature modeling [33,34]. It should be noted here that NNs are usually bad extrapolators, meaning that they do not perform satisfactorily in conditions considerably different than those of the training data. In hydroponics systems, NNs have been used to model with great accuracy the pH and the electrical conductivity of the nutrient solution in deep-trough cultivation systems [35], as well as the photosynthetic rate of the cultivated plants [36]. Furthermore, NNs have been successfully used in control applications of the greenhouse environment [37]. Very recently, their combination with genetic algorithms (GAs) in hydroponics modeling has been proven even more successful than conventional neural network modeling [38].

GAs are another AI technique that has been applied to greenhouse cultivation management and control. Their ability to find optimal solutions in large and complex search spaces, together with their innovative design capabilities inspired by the simulation of natural evolution, make them very powerful tools for design and optimization in several engineering applications. They have been used as an optimization tool for controller tuning of the greenhouse environment [39], as training methodologies of neural network agricultural models [40], as optimizers that determine optimal set-point values [41-43], and as optimizers of other soft computing-based controllers like fuzzy controllers [44]. Another technique similar to GAs, the *photosynthetic algorithm* [45], is a biologically inspired optimization algorithm that simulates the optimization processes involved in photosynthesis by plants. It was successfully applied to the training of neural network agricultural models [40].

Fuzzy logic is a quite commonly used intelligent technique in advanced control and management of greenhouse cultivation systems. The complex processes and interactions of the greenhouse environment make the kind of soft control that fuzzy logic incorporates very powerful and successful in accurate control and management of greenhouse systems, either as fuzzy logic on its own [46,47], or in combination with GAs and NNs [41,44]. It has been used to provide superior scaling among different production system sizes and loads in ventilation control [48] and in staged heating and ventilating systems in greenhouses [49]. It has also been used to provide management decisions in intelligent real-time control of greenhouse environment and hydroponics [3,50].

### 5.8.3 Medium-Level Control and Management

The control schemes described so far can be used as the basic elements of mediumlevel control methodologies which form sophisticated systems for greenhouse management with two main focus areas: longer than instantaneous time-horizon control management and conflicting-resolving control management.

## Medium Time-Scale Processes

As mentioned above, there are several different time scales in the processes involved in greenhouse cultivation. The first step in the development of sophisticated control schemes for medium-level control methodologies that deal with medium timescale processes, with an aim for energy saving, involves the method of averaging some of the parameters of interest (e.g., temperature or light). This is possible because biological properties of plants indicate that they comprise integrating capacities, which means that short-term fluctuations of temperature or light intensity do not affect plants' growth as long as an average value of each parameter is maintained over a certain period [51-54]. Energy saving is achieved because, for example in the case of temperature integration, the developed technique requires a desired average temperature during some specific time period and not specific temperature set-points for specific moments. In this way, a methodology can be developed that adapts the low-level temperature set-point according to the outside temperature conditions, so that minimum heat losses are achieved, keeping in mind of course that some specific average temperature has to be achieved by the end of the integration period [55]. If a good algorithm can be found, considerable energy savings can be achieved.

Some early investigation on the development of such integrating control strategies, with a focus on temperature integration, was performed by de Koning, who developed an algorithm based temperature averaging for 24-hour periods [56], and later for periods of several days [57]. In [58] an algorithm is presented that compensates for periods of temperature deviations in a greenhouse by slowly modifying the heating setpoint. The authors report good results in compensating for deviations either above or below blueprint temperatures. Timmons and Gates [59] developed a time-integrated approach to relative humidity control, and extended it to heat stress conditions for livestock [60] and optimal time-integrated variable set-points [61]. Marsh and Albright [62,63] presented a strategy for minimizing heating costs using an algorithm for calculating the economically optimum temperatures for greenhouse lettuce production. An alternative method based on formal optimization methods to achieve optimal setpoints for greenhouse lettuce production was proposed in [64]. Recently, several more advanced control strategies that exploit the temperature integrating capability of plants to achieve considerable energy savings have been developed, using the technological developments of IT [65-68]. In the case of light integration, Albright et al. [69] developed a rule-based algorithm to maintain the accumulated light intensity to a consistent daily integral, while Ferentinos et al. [70] optimized that control policy in relation to the CO<sub>2</sub> concentration of the greenhouse environment, according to the rules developed in [54].

The integration control policy for some environmental parameter can influence and make problematic some other control parameters. For example, in the case of temperature integration, relative humidity can often fluctuate and reach dangerously high values [71]. In long-term control strategies, short-term dynamics of the plants should be taken into consideration [72]. Thus, integrated control of the greenhouse environment requires the development of greenhouse management systems with capabilities of conflict-resolving schemes [73].

Additional information that can refine and improve the control strategies applied in a greenhouse cultivation system can be provided by forecasts, which lead to predictive control schemes. Weather forecasts have been used to optimize control of temperature integrating strategies, with considerable energy savings [74], while, in other cases, neural network models have been used to simulate greenhouse behavior in response to anticipated meteorological conditions, so that the predictive control strategy materializes [75].

Hydroponics is the other part of a greenhouse cultivation system that includes medium time-scale processes and needs to be controlled and managed, together with the aerial environment. Irrigation scheduling and nutrient supply in hydroponic systems are crucial and require precise control to optimize quality and quantity of crop production and to minimize cost and pollution due to effluents. A design for a water supply controller using system identification was proposed in [76], but the proposed controller performs well only when a feedforward element is added in the control loop, in order to estimate water uptake as a function of global radiation. On the other hand, progress has been achieved in model prediction of crop irrigation needs. Usually, a transpiration model [77] that predicts plant transpiration based on ambient conditions of temperature, solar radiation, CO<sub>2</sub> concentration, and vapor saturation deficit is used, while hybrid approaches using simplified transpiration models to predict the necessary water supply have also been proposed [78].

## Multi-Process Coupled Systems and Conflicts

Although biological systems consist of complex, inexactly defined, interacting processes, they have been until now treated with optimal control strategies that are applied separately to each process and not to the entire system as a whole [79,80]. The treatment of individual processes of a complex system, especially when those processes often conflict with each other (as in the case of greenhouse climate and hydroponics control), does not necessarily lead to the optimal solution of the entire system. The definition of set-points and constraints for the control of such systems is difficult and problematic. In addition, each process can contribute in a different and maybe changing degree to the final output of the system. Thus, conventional control methodologies encounter some major difficulties in situations where control variables are coupled; for example, temperature and humidity, which are highly coupled through nonlinear thermodynamic laws. In these cases, the control actuators are usually subject to changing characteristics, as the gain is largely perturbed by cross-product terms with disturbances. Another example of this kind of problem is addressed in hydroponics systems, where, when acid is added to the solution in order to reduce the pH value,

then electrical conductivity is strongly affected. The same holds true during electrical conductivity control, in which case some fertilizers affect the pH value.

Recently, these kinds of control problems have been addressed for the case of temperature and humidity coupling [81] and the case of simultaneous temperature and CO<sub>2</sub> concentration control [82]. In [81], conflicts due to temperature-humidity coupling are faced through an approach that consists of a powerful combination of linearizing and non-interacting feedback/feed-forward controllers, outer-loop conventional dynamic controllers (e.g., PID or PDF controllers) as well as a pre-compensator and command generator module, which computes set-point trade-offs based on psychrometric properties and actuator limits and costs to provide optimized set-points that will allow the feedback/feed-forward controller to operate without hunting or chattering.

In a more general context, in cases of process coupling and conflicts each process can have its own local dynamic control system, optimally tuned to the local goal, which is defined in a general environment, taking into account the final output of the entire system. This context leads to the development of *multi-agent systems* that can lead to the resolution of conflicting control decisions in complicated situations during plant growth [83].

## 5.8.4 High-Level Management

Recent demands of product quality and production performance have made advanced crop management techniques and intelligent control systems absolutely necessary for the operation of modern greenhouse cultivation facilities. Traditionally, information systems have consisted of databases, application programs, and user interfaces. This practice is changing because the new demand is for open integrated architectures with a more global scope through cooperative action [84]. Knowledge-based information systems, database management systems, and intelligent control are increasingly being integrated into IT. Databases offer information sharing while new computational intelligence techniques allow data mining, multi-agent systems, planning, scheduling, and negotiation. Greenhouse cultivation management systems are becoming increasingly sophisticated and are using many of the advanced methods and tools of industrial automation, modern control theory, and IT. Computer and communications technologies are closely linked to these developments. All these make feasible the development of integrated management systems incorporating a high degree of intelligence and flexibility in the manipulation of long-term effects in the involved processes in the greenhouse cultivation system.

## Long-Term Effects and Crop Lifetime Horizons

Several models have been developed to predict growth rate as a function of environmental parameters, with a final goal to optimize environmental control by providing optimal set-point strategies for each environmental parameter, taking into account mainly the long-term effects on the crop [85,3].

Physiological models together with actual real-time monitoring of the physiological status of the plants can lead to real-time control strategies. In practice, the measurement and identification of plant responses and the optimal control of the environment

based on plant responses are necessary. This leads to the *speaking plant approach* (SPA) [86 and Section 5.3 above]. This approach can benefit from a large number of plant-oriented measurements. Over the years, SPA has been integrated with modern techniques from the field of artificial intelligence, with an aim towards integrated intelligent control. Knowledge-based systems play an important role towards this effort [87,88].

Important issues in all these highly computerized and automated systems is the quality of information provided by the sensors and the quality of decisions passed to the actuators. IT can provide the capability of developing intelligent control systems capable of self-examination. The combined information from different sensors can lead to quality classification of isolated information derived from specific sensors or actuators [89]. In this way, several fault-detection and diagnosis methodologies have been developed in greenhouse cultivation systems [90-93].

## IGM: Integrated Greenhouse Management

Current major consumer concerns are direct exposure to traces of pesticides and other chemicals in food, and the indirect consequences to the environment, from the use of synthetic chemicals in the agricultural production process. Current research on consumer and environmental protection issues, and also on grower protection, has led to two basic solutions: biological products and integrated management. The grower has to comply with certain quality features in order to obtain certification for his products from specialized certification organizations. The actions that would lead to such compliance are the subject of continuous research and development of methods relevant to using natural and/or biological (not synthetic chemical) methods of plant protection, by the development of integrated pest management (IPM) techniques [94], and minimizing water and fertilizer use.

This integrated management of greenhouses has to be supervised and managed by an *integrated management system*, which is a specially designed program that leads to the development of certified products and provides the following:

- correct agricultural practices,
- employees' safety and hygiene,
- safety of products,
- · traceability, and
- environment-friendly actions.

The main goal of integrated management of greenhouses is profitable production in an economically viable and environmentally conscious agricultural facility, which incorporates beneficial natural processes into modern cultivation practices. The benefits from the application of the system are:

- performance assurance for the entire cultivation and grower's income,
- decrease of environmental impacts of agricultural actions, and
- protection of the environment and agricultural products with fewer amounts of synthetic chemical compounds.

## **5.8.5 ICT Tools and Complete Products**

The level of the technology from various aspects are:

- *System architectures*—Stand-alone controllers, PC-based, networked PC-based, and SCADA systems.
- Languages and tools—
  - 1. DOS-based; some systems still exist in 2004 due to reliability of this operating system and the wide availability of the PC platform.
  - 2. LabView or Matlab research setups, coupled to I/O command/control cards.
  - 3. Visual OOPS and Windows complete applications, usually used for the presentation and configuration functions of the controller part. Proprietary controllers are programmed in assembly, PLC types in ladder or similar, and advanced high-functionality systems in C.
  - 4. Decision support systems are developed using various programming platforms and technologies that reach higher levels of decision making (e.g., agents technologies for conflict management and optimization) and used as add-on systems to guide other lower-level but reliable systems, using specific DDE structures for data exchange.
- Communications—Wired with optical or magnetic isolation or RF at the system
  interconnection level from the sensor up to the controller and onward to the PC
  server. Internet connectivity has become a common feature for remote monitoring and control. Mobile connectivity is becoming an embedded technology to
  provide regular and alarm reporting to the "pocket of the user."
- *Greenhouse (GH) controllers*—Hardware aspects, autonomy aspects, advanced processing aspects, customizability and configurability issues, fixed-customizable-open IT systems for greenhouses.

## 5.8.6 Future IT Technologies for Greenhouse Production Management

The level we have reached in real field applications is advanced owing to efforts by researchers and companies involved in agricultural automation. These advances are becoming the mainstream technology; the cost is becoming lower and we see greenhouse automation including advanced features such as wireless sensors and actuators, distributed microcontrollers, main controllers supporting web cameras, remote supervisory systems with Internet connection, and, finally, remote support and troubleshooting networks. Knowledge networks are under development; these will remotely monitor the operation, assist the system or the grower in making critical decisions in risky situations (e.g., infections, nutrition problems in complex hydroponic recirculating systems), or to change the environment for a crop with unusual requirements. Special greenhouse units have been developed to become food-producing factories (e.g., leafy vegetables produced continuously in sealed chambers), fully equipped with tight environment control equipment and management systems with embedded knowledge.

## References

- 1. Bot, G. P. A. 1991. Physical modelling of greenhouse climate. *Mathematical and Control Applications in Agriculture and Horticulture*, eds. Y. Hashimoto, and W. Day, 7-12. Oxford, UK: Pergamon Press.
- 2. Bakker, J. C., G. P. A. Bot, H. Challa, and N. J. van de Braak. 1995. *Greenhouse Climate Control: An Integrated Approach*. Wageningen, The Netherlands: Wageningen.
- 3. Jones, J. W., E. Dayan, L. H. Allen, H. Vankeulen, and H. Challa. 1991. A dynamic tomato growth and yield model (TOMGRO). *Trans. ASAE* 34: 663-672.
- 4. Challa, H., and G. van Straten.1993. Diurnal control in greenhouses. *The Computerized Greenhouse*. eds. Y. Hashimoto, G. P. A. Bot, W. Day, H.-J. Tantau, and H. Nonami, 119-137.San Diego, CA: Academic.
- 5. Aaslyng, J. M., N. Ehler, P. Karlsen, and E. Rosenqvist. 1999. Intelligrow: A component-based climate control system for decreasing greenhouse energy consumption. *Acta Horticulturae* 507: 35-42.
- Bates, M. E., and D. L. Bubenheim. 1999. Applications of process control to plant-based life support systems. SAE Technical Paper No. 972359. Warrendale, PA: Society of Automotive Engineers.
- 7. Volk, T., B. Bugbee, and R. M. Wheeler. 1995. An approach to crop modeling with the energy cascade. *Life Support and Biosphere Sci.* 1: 119-127.
- 8. Marcelis, L. F. M., E. Heuvelink, and J. Goudriaan. 1997. Modelling biomass production and yield of horticultural crops: A review. *Sc. Horticulturae*, 83-104.
- 9. Ferentinos, K. P., A. Anastasiou, G. D. Pasgianos, K. G. Arvanitis, and N. Sigrimis. 2003. A decision support system as a tool to optimal water management in soilless cultures under saline conditions. *International ISHS Symposium on Managing Greenhouse Crops in Saline Environment*.
- 10. Sigrimis, N., K. G. Arvanitis, and G. D. Pasgianos. 2000. Synergism of high and low level systems for the efficient management of greenhouses. *Computers and Electronics in Agriculture* 29: 21-39.
- 11. Seginer, I. 1993. Crop models in greenhouse climate control. *Acta Horticulturae* 328: 79-97.
- 12. Albright, L. D., R. S. Gates, K. G. Arvanitis, and A. Drysdale. 2001. Environmental control for plants on Earth and in space. *IEEE Control Systems Magazine* 21(5): 28-47.
- 13. Gates, R. S., L. W. Turner, and D. G. Overhults. 1992. A survey of environmental controllers. *Trans. ASAE* 35: 993-998.
- Albright, L. D. 1990. Environment Control for Animals and Plants. St. Joseph, MI: ASAE.
- 15. Young, P. C., W. Tych, and A. Chotai. 1991. Identification, estimation and control of glasshouse systems. *Mathematical and Control Applications in Agriculture and Horticulture*, eds. Y. Hashimoto, and W. Day, 33-40. Oxford, UK: Pergamon Press.

- Chotai, A., P. C. Young, P. Davis, and Z. Chalabi. 1991. True digital control of glasshouse systems. *Mathematical and Control Applications in Agriculture and Horticulture*. eds. Y. Hashimoto, and W. Day, 41-46. Oxford, UK: Pergamon Press.
- 17. Young, P. C., A. Chotai, and W. Tych. 1993. Identification, estimation and true digital control of glasshouse systems. *The Computerized Greenhouse*, eds. Y. Hashimoto, G. P. A. Bot, W. Day, H.-J. Tantau, and H. Nonami, 3-50. Academic Press.
- 18. Åström, K., and T. Hägglund. 1998. *PID controllers: Theory, Design and Tuning*. Research Triangle Park, NC: ISA Publications.
- 19. Jones, P., J. W. Jones, L. H. Allen, and J. W. Mishoe. 1984. Dynamic computer control of closed environmental plant chambers: Design and verification. *Trans. ASAE* 27: 879-888.
- 20. Stanghellini, C., and W. T. M. van Meurs. 1992. Environmental control of greenhouse crop transpiration. *J. Agricultural Engineering Research* 51: 297-311.
- 21. Arvanitis, K. G., P. N. Paraskevopoulos, and A. A. Vernardos. 2000. Multirate adaptive temperature control of greenhouses. *Computers and Electronics in Agriculture* 26: 303-320.
- 22. Chao, K., and R. S. Gates. 1996. Design of switching control systems for ventilated greenhouses. *Trans. ASAE* 39: 1513-1523.
- 23. Phelan, R. M. 1977. *Automatic Control Systems*. Ithaca, NY: Cornell University Press.
- 24. Setiawan, A., L. D. Albright, and R. M. Phelan. 2000. Application of pseudo-derivative-feedback algorithm in greenhouse air temperature control. *Computers and Electronics in Agriculture* 26: 283-302.
- 25. Young, P. C., M. J. Lees, A. Chotai, W. Tych, and Z. S. Chalabi. 1994. Modelling and PIP control of a glasshouse microclimate. *Control Engineering Practice* 2: 591-604.
- 26. Lees, M. J., M. J. Taylor, A. Chotai, P. C. Young, and Z. S. Chalabi. 1996. Design and implementation of a Proportional-Integral-Plus (PIP) control system for temperature, humidity and carbon dioxide in a glasshouse. *Acta Horticulturae* 406: 115-123.
- 27. Smith, O. J. M. 1957. Closer control of loops with dead time. *Chemical Engineering Progress* 53: 217.
- 28. Sigrimis, N., K. G. Arvanitis, G. D. Pasgianos, and K. P. Ferentinos. 2002. Computer integrated management and intelligent control of greenhouses. *Environment Control in Biology* 40(1): 39-53.
- Hashimoto, Y. 1997. Applications of artificial neural networks and genetic algorithms to agricultural systems. *Computers and Electronics in Agriculture* 18: 71-72.
- 30. Seginer, I., T. Boulard, and B. J. Bailey. 1994. Neural network models of the greenhouse climate. *J. Agricultural Engineering Research* 59: 203-216.
- 31. Seginer, I., Y. Hwang, T. Boulard, and J. Jones. 1996. Mimicking an expert greenhouse grower with a neural network policy. *Trans. ASAE* 39: 299-306.

- 32. Kok, R., R. Lacroix, G. Clark, and E. Taillefer. 1994. Imitation of a procedural greenhouse model with an artificial neural-network. *Canadian Agricultural Engineering* 36: 117-126.
- 33. Ferreira, P. M., E. A. Faria, and A. E. Ruano. 2002. Neural network models in greenhouse air temperature prediction. *Neurocomputing* 43: 51-75.
- 34. Frausto, H. U., and J. G. Pieters. 2004. Modelling greenhouse temperature using system identification by means of neural networks. *Neurocomputing* 56: 423-428.
- 35. Ferentinos, K. P., and L. D. Albright. 2002. Predictive neural network modeling of pH and electrical conductivity in deep-trough hydroponics. *Trans. ASAE* 45: 2007-2015.
- 36. Frick, J., C. Precetti, and C. A. Mitchell. 1998. Predicting lettuce canopy photosynthesis with statistical and neural network models. *J. American Society for Horticultural Science* 123: 1076-1080.
- 37. Seginer, I. 1997. Some artificial neural network applications to greenhouse environmental control. *Computers and Electronics in Agriculture* 18: 167-186.
- 38. Ferentinos, K. P. 2005. Biological engineering applications of feedforward neural networks designed and parameterized by genetic algorithms. *Neural Networks* 18: 934-950.
- 39. Goggos, V., and R. E. King. 2000. Qualitative-evolutionary design of greenhouse environment control agents. *Computers and Electronics in Agriculture* 26: 271-282.
- 40. Hashimoto, Y., H. Murase, T. Morimoto, and T. Torii. 2001. Intelligent systems for agriculture in Japan. *IEEE Control Systems Magazine* 21(5): 71-85.
- 41. Morimoto, T., K. Hatou, and Y. Hashimoto. 1996. Intelligent control for plant production system. *Control Engineering Practice* 4: 773-784.
- 42. Morimoto, T., T. Torii, and Y. Hashimoto. 1995. Optimal-control of physiological processes of plants in a green plant factory. *Control Engineering Practice* 3: 505-511.
- 43. Husmann, H. J., and H. J. Tantau. Integrated optimization of energy supply systems in horticulture using genetic algorithms. *Computers and Electronics in Agriculture* 31: 47-59.
- 44. Caponetto, R., L. Fortuna, G. Nunnari, L. Occhipinti, and M. G. Xibilia. 2000. Soft computing for greenhouse climate control. *IEEE Transactions on Fuzzy Systems* 8: 753-760.
- 45. Murase, H. 2000. Finite element inverse analysis using a photosynthetic algorithm. *Computers and Electronics in Agriculture* 29: 115-123.
- 46. Kurata, K., and N. Eguchi. 1990. Machine learning of fuzzy rules for crop management in protected cultivation. *Trans. ASAE* 33: 1360-1368.
- 47. Salgado, P., J. B. Cunha, and C. Couto. 1998. A fuzzy identification and controller for the agricultural greenhouse. 7<sup>th</sup> *International Conference on Computers in Agriculture*. St. Joseph, MI: ASAE.
- 48. Gates, R. S., K. Chao, and N. Sigrimis. 2001. Identifying design parameters for fuzzy control of staged ventilation control systems. *Computers and Electronics in Agriculture* 31: 61-74.

- 49. Chao, K., R. S. Gates, and N. Sigrimis. 2000. Fuzzy logic controller design for staged heating and ventilating systems. *Trans. ASAE* 43: 1885-1894.
- 50. Sigrimis, N. A., K. G. Arvanitis, and R. S. Gates. 2000. A learning technique for a general purpose optimizer. *Computers and Electronics in Agriculture* 26: 83-103.
- 51. Krug, H., and H.-P. Liebig. 1980. Diurnal thermoperiodism of the cucumber. *Acta Horticulturae* 118: 83-94.
- 52. Langhans, R. W., M. Wolfe, and L. D. Albright. 1981. Use of average night temperatures for plant growth for potential energy savings. *Acta Horticulturae* 115: 31-37.
- 53. Miller, W. B., R. W. Langhans, and L. D. Albright. 1985. Plant growth under average day/night temperatures. *Acta Horticulturae* 174: 313-320.
- 54. Both, A. J., L. D. Albright, and R. W. Langhans. 1997. Coordinated management of daily PAR integral and carbon dioxide for hydroponic lettuce production. *Acta Horticulturae* 456: 45-51.
- 55. Saffell, R. A., and B. Marshall. 1983. Computer control of air temperature in a glasshouse. *J. Agricultural Engineering Research* 28: 469-478.
- 56. De Koning, A. N. M. 1988. An algorithm for controlling the average 24-hour temperature in glasshouses. *J. Horticultural Science* 63: 473-477.
- 57. De Koning, A. N. M. 1990. Long term temperature integration of tomato: Growth and development under alternating temperature regimes. *Scientia Horticulturae* 45: 117-127.
- 58. Hooper, A. W., and P. F. Davis. 1988. An algorithm for temperature compensation in a heated greenhouse. *Computers and Electronics in Agriculture* 3: 251-262.
- 59. Timmons, M. B., and R. S. Gates. 1987. Relative humidity as a ventilation control parameter in broiler housing. *Trans. ASAE* 30: 1111-1115.
- 60. Timmons, M. B., R. S. Gates, R. W. Bottcher, T. A. Carter, J. T. Brake, and M. J. Wineland. 1995. Simulation analysis of a new temperature control method for poultry housing. *J. Agricultural Engineering Research* 62: 237-245.
- 61. Gates, R. S., M. B. Timmons, D. G. Overhults, and R. W. Bottcher. 1997. Fuzzy reasoning for environment control during periods of heat stress. *Proc. of the Fifth International Symposium, Livestock Environment*, 553-562. St Joseph, MI: ASAE.
- 62. Marsh, L. S., and L. D. Albright. 1991. Economically optimum day temperature for greenhouse hydroponic lettuce production Part 1: A computer model. *Trans. ASAE* 34: 550-556.
- 63. Marsh, L. S., and L. D. Albright. 1991. Economically optimum day temperature for greenhouse hydroponic lettuce production Part 2: Results and simulations. *Trans. ASAE* 34: 557-562.
- 64. Seginer, I., G. Shina, L. D. Albright, and L. S. Marsh. 1991. Optimal temperature setpoints for greenhouse lettuce. *J. Agricultural Engineering Research* 49: 209-226.

- 65. Tantau, H.-J. 1998. Energy saving potential of greenhouse climate control. *Mathematics and Computers in Simulation* 48: 93-101.
- 66. Sigrimis, N., A. Anastasiou, and N. Rerras. 2000. Energy saving in greenhouse using temperature integration: A simulation survey. *Computers and Electronics in Agriculture* 26: 321-341.
- 67. Bot, G. P. A. 2001. Developments in indoor sustainable plant production with emphasis on energy saving. *Computers and Electronics in Agriculture* 30: 151-165
- 68. Korner, O., and H. Challa. 2003. Design for an improved temperature integration concept in greenhouse cultivation. *Computers and Electronics in Agriculture* 39: 39-59.
- 69. Albright, L. D., A.-J. Both, and A. J. Chiu. 2000. Controlling greenhouse light to a consistent daily integral. *Trans. ASAE* 43: 421-431.
- 70. Ferentinos, K. P., L. D. Albright, and D. V. Ramani. 2000. Optimal light integral and carbon dioxide concentration combinations for lettuce in ventilated greenhouses. *J. Agricultural Engineering Research* 77: 309-315.
- 71. Korner, O., and H. Challa. 2003. Process-based humidity control regime for greenhouse crops. *Computers and Electronics in Agriculture* 39: 173-192.
- 72. van Straten, G., H. Challa, and F. Buwalda. 2000. Towards user accepted optimal control of greenhouse climate. *Computers and Electronics in Agriculture* 26: 221-238.
- 73. van Straten, G., L. G. van Willigenburg, and R. F. Tap 2002. The significance of crop co-states for receding horizon optimal control of greenhouse climate. *Control Engineering Practice* 10: 625-632.
- 74. Chalabi, Z. S., B. J. Bailey, and D. J. Wilkinson. 1996. A real-time optimal control algorithm for greenhouse heating. *Computers and Electronics in Agriculture* 15: 1-13.
- 75. Lacroix R, and R. Kok. 1999. Simulation-based control of enclosed ecosystems—A case study: Determination of greenhouse heating setpoints. *Canadian Agricultural Engineering* 41: 175-183.
- 76. Gieling, T., H. J. J. Janssen, G. van Straten, and M. Suurmond. 2000. Identification and simulated control of greenhouse closed water supply systems. *Computers and Electronics in Agriculture* 26: 361-374.
- 77. Stanghellini, C. 1987. Transpiration of greenhouse crops, an aid to climate management. PhD diss. Wageningen, The Netherlands: Wageningen Agricultural University.
- 78. Sigrimis, N., K. G. Arvanitis, G. D. Pasgianos, and K. Ferentinos. 2001. Hydroponics water management using adaptive scheduling with an on-line optimiser. *Computers and Electronics in Agriculture* 31: 31-46.
- 79. Gal, S., A. Angel, and I. Seginer. 1984. Optimal control of greenhouse climate: methodology. *European J. Operational Research* 17: 45-56.
- 80. Seginer, I., and R. W. McClendon. 1992. Methods for optimal control of the greenhouse environment. *Trans. ASAE* 35: 1299-1307.

- 81. Pasgianos, G. D., K. G. Arvanitis, P. Polycarpou, and N. Sigrimis. 2003. A nonlinear feedback technique for greenhouse environmental control. *Computers and Electronics in Agriculture* 40: 153-177.
- 82. Linker, R., P. O. Gutman, and I. Seginer. 1999. Robust controllers for simultaneous control of temperature and CO<sub>2</sub> concentration in greenhouses. *Control Engineering Practice* 7: 851-862.
- 83. Ferentinos, K. P., K. G. Arvanitis, S. Hontas, D. Lambrou, and N. Sigrimis. 2004. A multi-agent system for integrated production in greenhouse hydroponics. *AgEng 2004, Leuven, Belgium,* accepted for publication.
- 84. King, R. E., and N. Sigrimis. 2001. Computational intelligence in crop production. *Computers and Electronics in Agriculture* 31: 1-3.
- 85. Jolliet, O. 1994. HORTITRANS, a model for predicting and optimizing humidity and transpiration in greenhouses. *J. Agricultural Engineering Research* 57: 23-37.
- 86. Morimoto, T., and Y. Hashimoto. 2000. An intelligent control for greenhouse automation, oriented by the concepts of SPA and SFA—An application to a post-harvest process. *Computers and Electronics in Agriculture* 29: 3-20.
- 87. Chao, K., R. S. Gates, and R. G. Anderson. 1998. Knowledge-based control systems for single stem rose production Part I: Systems analysis and design. *Trans. ASAE* 41: 1153-1161.
- 88. Anastasiou, A., N. Rerras, and N. Sigrimis. 1998. A knowledge based SCADA system for agricultural process control. *Proc. of International Federation of Automatic Control—CAEA'98*, eds. N. Sigrimis, and P. Groumpos, 163-168. Oxford, UK: Pergamon.
- 89. Ferentinos, K. P., and L. D. Albright. 2003. Fault detection and diagnosis in deep-trough hydroponics using intelligent computational tools. *Biosystems Engineering* 84: 13-30.
- 90. Beaulah, S. A., and Z. S. Chalabi. 1997. Intelligent real-time fault diagnosis of greenhouse sensors. *Control Engineering Practice* 5: 1573-1580.
- 91. Linker, R., P. O. Gutman, and I. Seginer. 2002. Observer-based robust failure detection and isolation in greenhouses. *Control Engineering Practice* 10: 519-531.
- 92. Ferentinos, K. P., L. D. Albright, and B. Selman. 2003. Neural network-based detection of mechanical, sensor and biological faults in deep-trough hydroponics. *Computers and Electronics in Agriculture* 40: 65-85.
- 93. Linker, R., and I. Seginer. 2003. Water stress detection in a greenhouse by a step change of ventilation. *Biosystems Engineering* 84: 79-89.
- 94. Antignus, Y., 2001. Manipulation of wavelength-dependent behaviour of insects: An IPM tool to impede insects and restrict epidemics of insect-borne viruses. *Virus Research* 71: 213-220.

## 5.9 Precision Livestock Production

I. A. Nääs, V. C. Carvalho, D. J. Moura, and M. Mollo

Abstract. Modern animal production has changed in recent years due to the use of precision tools. Results of recent research have been used as inputs to preventive diagnostics and development of decision-making software in several areas, as well as to predict events. Evaluation of animal welfare can also be determined by telemetry; image and sound analysis can be valuable tools for understanding the animal's response and enable the producer to make the right decision based on real-time management. In this section, examples of developing technology in the fields of animal monitoring, traceability, and preventive diagnostics are presented.

**Keywords.** Livestock production, Real-time management decisions, Image analysis, Preventive veterinary diagnostics.

#### 5.9.1 Introduction

The future of animal commerce depends mainly on an industry reacting to the following concepts: honesty, openness, detailed information available, traceability, assurance of quality, and flexibility for changes [1]. For the retailer or fast food buyer, it is only possible to build up a business when quality is always renewed and always available in the right place at the right time.

The need to identify animals in a herd is known through history, for property ownership, and more recently for purposes of genetic studies. The first known way of definitive identification of livestock was by tattooing with a hot iron. This is still used for beef cattle in some countries. Later, a piercing code was introduced to mark swine.

Health-status certification is also a requirement for trade of livestock, mainly at the international level. Successful trade relationships require trust to warranty quality of the product [2]. In the livestock business, one element of the quality characteristics is the health or safety status of animals or their products. The seller usually certifies this status. However, the buyer usually cannot check the accuracy of this status before the animals or their products arrive at the destination; trust is required. Trust can be based on evidence provided, as well as on a history of honesty, transparency, and competence [3]. When animals are traded locally or regionally between farms of different health status, the buyer might request certification based on veterinary inspection and/or diagnostic tests. This information might be provided by the producer or by animal-health services. Identification is then needed in order to assure the accuracy and precision of the data, which is the basis for the traceability system.

## 5.9.2 Monitoring Animal Functions and Conditions

Monitoring animal functions and conditions can be thought of as a multipurpose tool in which the health status of an animal can be closely traced in several ways. Currently, the main purposes of animal monitoring are (1) to assure quality of the final bioproducts of agricultural industries, (2) to correlate animal behavior to health and

welfare, and (3) to evaluate pathologies related to faulty locomotion and its impact on animal welfare. These have been accomplished by the use of technologies recently developed in the fields of force and pressure sensors, identification transmitters, and image processing. The next sections will discuss current technologies being developed in these fields.

## Radio Frequency Identification (RFID)

Electronic identification systems are a key technology for the automation of processes. Their implementation is targeted to help improve the quality, economy, and environmental impact of animal production.

The readability of bolus and injected transponders, as well as ear tags, for ruminants was compared by [4]. The authors found that the designed rumenal bolus was successful as a unique carrier of transponders for the electronic identification of different livestock ruminant species (sheep, goat, and cattle). It is possible to use the bolus in combination with management practices on farm conditions, such as weighing with electronic scales and dynamic reading. The use of ear tags carries a relatively high loss risk and a high possibility of undesired exchange of tags.

In practice, RFID implementations can solve several problems in intensive animal production management. Reading speed and distance must be optimized for specific applications. The International Committee for Animal Recording (ICAR) developed in 1995 a set of requirements regarding (among others) the reading distance and reading speed. Other issues include biocompatibility of encapsulation, as well as the injection site in connection with migration problems, recovery in slaughterhouses, standardization for open trade, and proper effective management of issued unique life-numbers.

The location of the transponder may not change after the application (i.e., no migration). Controlling migrations (movement) of transponders is a critical aspect for their use. The main problem is that moving transponders could be a risk for some essential organs. Moreover, migrated transponders may cause difficulties in the abattoir as they cannot be recovered at the expected site [5].

## Preventive Veterinary Medicine

The modern dairy industry is one of the sectors that has greatly benefited from research on livestock housing. Much of this research deals with the effects of freestall concrete surfaces on weight-bearing biomechanics. It started in the past decade using force plates and was greatly improved with the development of plantar pressure sensitive mats (MatScan<sup>TM</sup>, Footscan<sup>TM</sup>) used in modern housing research [6,7].

Image analysis (high-speed videography) is an older technology used to detect faulty locomotion and gait deviations affecting posture and ergonomics of humans and other animals. Locomotion can play an important role in health because its restriction will result in the animal's prostration and eventual death. This same video analysis has also been used for behavioral studies because it removes bias resulting from human fatigue and the consequent misperceptions over long periods of time.

## Kinematics and Preventive Diagnostics

Lameness is among the most prevalent and costly of clinical disease conditions in dairy cattle. Flooring is of particular importance, because of pressure distribution and redistribution on claws. Uneven weight-bearing of hoof walls of cows managed on hard floors (i.e., concrete) leads to pressure redistribution on claws thus causing greater pressure concentration and stress on claws. Therefore, weight bearing and plantar pressure distribution is an important measurement and especially useful for the appropriate understanding of the biomechanical abnormalities usually encountered within the agricultural industry's modern confinement housing and how to prevent the costly locomotory disorders incurred as its consequence.

Force measurement equipment usually consist of force plates or platform scales; however, the Massachusetts Institute of Technology (MIT) has developed a new and more accurate form of force/pressure measurement. It consists of ultra-thin films containing several arrays of piezoelectric crystal sensors developed for human gait analysis. The MatScan (Tekscan Inc.) pressure measurement film, based on this technology, was used to evaluate pressure distribution under a cow's stride [7]. The system was able to yield reliable pressure data from 32 cows allowing the comparison of two populations of interest (trimmed and untrimmed cows). Results showed that the highest pressures on the rear feet of both trimmed and untrimmed cows occurred on region 1 with 30.97% for untrimmed vs. 29.10% for trimmed, but were not different between groups, followed by regions 4, 5, and 2. The main differences on the rear feet caused by trimming, although small, occurred on regions 5 and 3 and to a lesser extend on region 2 (Figure 1).

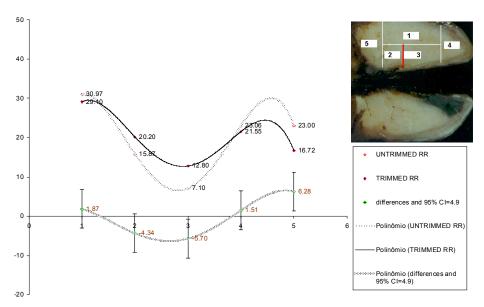


Figure 1. Rear right feet LSMenas for Group × Leg × Region interaction (modified from [7]), mean differences and their 95% CI (intervals including zero are not statistically significant;  $\alpha = 0.05$ ).

These changes accounted for a small improvement towards the anterior part of the claw, that is, the higher pressure concentrations at the heel (region 5) decreased from 22.99% to 16.72% ( $\sim$  6% difference, p < 0.05) increasing mostly at the anterior portion of the sole on trimmed claws from 7.09% to 12.8% ( $\sim$  6% difference, p < 0.05), for untrimmed vs. trimmed, respectively.

Another tool for measuring and modeling animal locomotion that can produce a great impact on the diagnosis of an animal's health and welfare is kinematic analysis. The use of dynamic video images can help evaluate abnormal gait and small deviations that are not perceived by human eyes. These images can be aided by biomechanical software that has the ability of modeling gait to the point of performing mathematical calculations of position in space and time. The resulting data can be used to compare populations and help further studies of load impacts along the body brought by abnormal loading caused by faulty locomotion.

Data of linear and angular kinematics were obtained using a motion analysis system and video recordings of the walking strides of two groups of cows [7]. A digital video camera (JVC GDR-120U, 30 Hz, 520 lines vertical resolution) was used for acquisition of 2-D (two-dimensional) video kinematics data. The video data collected were captured into a PC using video editing software (Adobe Premiere 6.5<sup>TM</sup>)[8]. Linear (spatial and temporal) and angular (feetlock joint range of motion) kinematics were obtained and modeled using biomechanical software developed for human gait analysis, the Human Movement Analysis Software [9] developed by the HMA Technology Inc. (Ontario, Canada). Unfortunately, when dealing with lameness in cattle, the earliest pathological gait signs are typically characteristic of mild to severe degrees of lameness. Usually by this point veterinary intervention is required, incurring economic losses to the dairy industry and animal welfare.

With the objective of developing an expert system based on a fuzzy logic algorithm for the preventive diagnostic and decision-making on dairy cattle lameness, a preliminary knowledge base was created by gathering information linking pressure distribution on claws of dairy cattle [7] and nutritional components data. The fuzzy set controller was designed using the software [10] based on 162 rules organized through the Karnaugh mapping method. The system links four input variables: toe length (mm) [7,11], neutral digestive fiber (NDF, %), non-structural carbohydrates (NSC, %), and non-fiber carbohydrate (NFC, %) [12]. It outputs an unitless prognostic value concerning increasing (qualitative) degrees of risks of developing lesions of the sole ulcer type, according to the information entered by the user into the software interface as shown in Figure 2.

The decision support intended by the system lies in either controlling levels of the essential nutritional components and/or trimming the excessive horn tissue from claws into acceptable lengths.

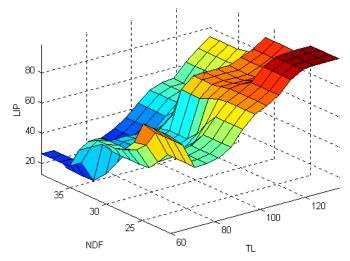


Figure 2. Surface chart of toe length (TL) and neutral digestive fiber (NDF) versus lesion incidence possibilities (LIP).

## 5.9.3 Modeling Animal Response

In animal production, specific improvements in the production system may bring certain benefits. To get more significant results it is necessary to examine the production system in an integrated way. It is also important to direct research to areas where the knowledge is limited or where the new knowledge will have greater impact. Through the development of simulation models it is possible to identify knowledge gaps, where research becomes necessary.

Simulation models can be used in the elaboration of strategies to optimize growth, to reduce mortality and production cost, and to improve the quality of the carcass, among others. They can also be used to simulate the potential of alternative systems of production before their implementation. To develop a good model, is necessary to know the physiological mechanisms of the animals involved. Then the model will only be able to predict the behavior of a system with a reasonable degree of precision. A good model allows the estimation of the results of an experiment before it is carried out

Mathematical modeling of the events in animal production makes possible the maximization of the efficiency in operations, through the maximization of operational schedules, events, automation, notification of problems, and transference of data and information. The complete system acts inside of a segment of intelligence within the software that formulates scenarios using auto-proofing and net topology methods, evaluating its performance, managing and monitoring all the electronic devices. It provides the automation as a function of the registered behavioral answers. Graphics of the environment inside the housing allows the visualization of the productive processes and the intervals of the production cycles, through a geographic visualization of the interior of the housings, distribution of the electronic devices and their respective

localizations. In short, the modeling carries through the monitoring of all the animals and makes possible the accomplishment of analysis of behavior as a function of its welfare.

#### Behavior

New technologies for animal behavior monitoring have been developed that allow the estimation of a series of pertinent information related to health and productivity of the animals. Some systems were developed for monitoring animal behavior. It was demonstrated by [13] that the Global Positioning System (GPS) can be used for monitoring sheep on pasture. In confinement housings it could be demonstrated that the analysis of images is a good tool for monitoring the behavior of the animals[14]. Passive infrared detectors (PIDs) were used by [15] to measure the activity of swine. According to [14] the only commercially available equipment for measuring certain aspects of animal activity is the pedometer, which can be used in the detection of estrus in milking cows.

A method developed for the evaluation of tools and strategies for the measurement of animal behaviors was described by [16]. The author shows the power of new technologies and available tools, such as cameras, computers, software, and the considerable increase of the efficiency of the experimental work in analyses of animal behaviors. Therefore, the study of behaviors can be measured with an accuracy that, previously, could not be reached through the traditional methods of observation, and that is essential for the study of the internal structures of the animal behaviors.

## Real-Time Management

The benefits of using transponders for monitoring animal bioenergetics are shown by [17]. The authors established an intensive monitoring of feed consumption, heat production, and behavior, through the use of electronic identification, automatic feeding systems, calorimeters, and image processing, all connected to microprocessors. This technique introduced a higher degree of accuracy when compared to traditional observational methods of studying behavior.

The behavior of poultry breeders was recorded by [18]. The authors related the behavior to the environment characteristics using RFID and telemetry in small-scale model housing in two different solar orientations. During the experiment, the female breeders' path was registered using electronic identification technology [19]. Furthermore, a model relating the environmental temperature and the breeders' movement inside the housing was developed. The real-time thermoneutral zone for female broiler breeders was determined by analyzing their behavior through monitoring the birds individually [20]. It was possible to estimate the thermoneutral zone using real-time values of specific behavior for the female broilers breeders studied.

### Image Analysis

Like precision agriculture in crop production, animal production currently requires the use of technology that involves intensive use of image-processing supported equipment to monitor and detect animal responses, promoting better economic efficiency. In early research, the main reason to use image analysis was to automate quality control [21-23]. Today's increase in the use of remote sensing technologies and image interpretation is due to the fact that it is faster and less expensive than conducting a ground survey [24].

Images are produced by a variety of physical devices, including still and video cameras, X-ray devices, electron microscopes, radar, and ultrasound, and used for a variety of purposes, including entertainment, medical, business, industrial, military, civil, security, scientific, and now for new applications focused on agriculture. The goal in each case is for an observer (human or machine) to extract useful information about the scene being imaged. For instance, [25] presented an automated inspection system to classify wet blue leather, using image processing and under a quality control system guiding rules.

Video analysis has been shown as a potential tool for the evaluation of the movement of domestic animals, permitting the investigation of relationships between animal behavior and the environment they are provided, as well as more accurate investigation of the effects of climate on the animals' physiological responses and superficial temperature monitoring (thermography) for the animals themselves or the housing in which they are confined. For instance, in an experiment conducted in free-stall barns in southeastern Brazil, data referent to behavioral patterns of the cows to be monitored was collected [26]. Software [10] was chosen for developing an algorithm in order to process and identify the animal by image segmentation as suggested by [27,28].

## Automated Techniques for Evaluating the Behavior of Animals

Image analysis is a promising tool to evaluate the animal housing environment, minimizing the inherent problems of conventional methods [29]. According to [14], the analysis of the movement of animals in groups, performed through images, can use the animals' responses as a feedback for the environmental control. The observation of the behavior using video cameras is an inexpensive and efficient alternative, since the data can be analyzed at any time without the errors committed by the direct and subjective observation of an individual, and without the interference in the behavior of the animal caused by the presence of human being, as cited by [30].

Previous research used the behavior of the cows in confinement housing as indicative of their comfort level. Videotape images and sequential photographs had been used to monitor the different activities of housed animals [31-33]. In addition, the use of video cameras also allows the study of behaviors that occur suddenly, followed by a long period of inactivity [34]. Also, it allows the monitoring of behaviors that repeat over time, as well as nocturnal/diurnal variability of behaviors [35,36].

## Applications of Neural Analysis

Neural analysis has developed full programs since the 1980s. Its main characteristic is its intelligent potential. It also has characteristics of auto-organization, autolearning, dynamics of linear processing, and the capacity of decision making and adaptations, among others. These techniques can identify animals and have the potential to detect, in a non-invasive remote way, the occurrence of various situations related to stress behaviors, reproduction, health, etc. This information can be used in research as well as in production management.

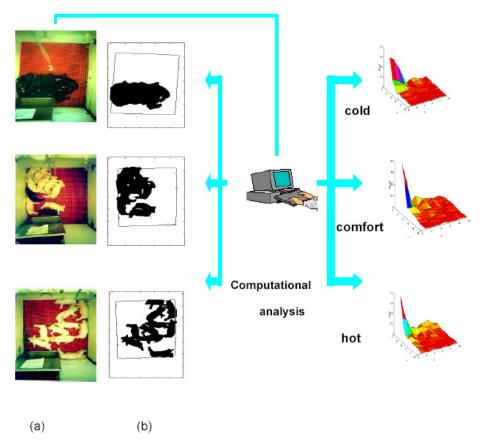


Figure 3. Illustration of environmental control through the analyses of welfare images of piglets: (a) real image; (b) segmented image (adapted from [37]).

The use of image analysis (Figure 3) to interpret the observable responses of animals regarding thermal characteristics of the environment is currently investigated and used. Studies have demonstrated the effectiveness of using image analysis to classify thermal comfort of piglets using a neural network. Image analysis is also used in livestock traceability (see below).

## Algorithm Application

The *algorithm* can be used as a generic tool to represent the solution for tasks independent of the desire to automate them, but in general it is associated with the electronic processing of data, where the algorithm represents the rough draft for software. It serves as model for these programs, therefore its language is intermediate between the language of human beings and the programming languages, being thus a good tool to validate the logic of tasks to be automated.

There are several image processing techniques for the detection of movements, but the most currently used is the method of transference of Fourier and the method of variations modeling [38,39]. The animal and its background (floor, feed bunk, water bunk, etc.) must be segmented before the behavior is classified. In this case, there is sufficient contrast between the animal and the majority of background objects. To reduce the memory requirement and to improve the processing of the images, the segmentation is made in binary format (with piglets in white = intensity 1; the background in black = intensity 0). Small objects that remain in the floor, such as wastes, are eliminated from the images by openings filters and filters of recognition of small long-distance points. The operators of openings can be visualized with a morphologic filter that generally alleviates the contour of objects, excludes indefinite objects, and eliminates small objects.

### 5.9.4 Traceability

The "farm to fork" strategic approach in integrated animal production systems is designed to cover the entire food chain. It contains all elements of the food production chain including the health, management, and welfare of animals. Traceability can be done either manually or electronically, or using both depending on the event to be registered. However, the decision is complex, related to the nature of the specific management task as well as to the economical and technological feasibility.

The first step of traceability is identifying animals. The technology of the process is not new. Animals have long been identified to proof of ownership; only lately has identification become an essential need, with the urge to document origin and implement the traceability process. The traceability process in animal production depends on accuracy for reliability. Electronic identification of cattle using RFID, for instance, has many advantages for farm management [40]. First, it can be regarded as a considerable improvement in relation to visual identification of numbers. The main advantages are the elimination of labor costs and the decrease of incorrect readings from 6% to 0.1% [41]. Allowing the automation of, for example, feed monitoring and rationing, weighing, and drafting, can implement sophisticated livestock management schemes.

Application of RFID cattle management can be carried out on the basis of the individual animal performance recording, with dispensing of feed and geographic routing dependent on the animal status. Examples are robot milking and the implementation of geographic information systems to assess the potential transmission of infectious diseases between herds [42].

Petersen et al. [43] describe a model for using this technology in swine production where a computerized health management system is used in the entire production chain from breeding to slaughter. The model is structured according to the data recording, processing, and exchange of information between farms, abattoir, and the consulting service. It was shown that the expert feedback is essential in the decision-making process.

Other important applications enabled by injected electronic transponders are improvement of disease control and eradication, as well as fraud control. The latter application is important mainly within the European Union (EU), where premiums are

being paid to stimulate extensive sheep and beef production. Also within the EU, where it is not longer allowed to eradicate some contagious diseases by means of vaccination, the individual ID plays an important role. In case of an outbreak, it is important to trace back the origin, movements, and contacts between animals in order to be able to stop the further dissemination of contagious diseases.

## 5.9.5 Conclusions

The use of information technology in animal production will help farmers decrease losses during the animal production cycle by the use of precision principles and more accuracy, improving the overall management. On the other hand, biosensor advancement in the commercial world could also be accelerated by the use of intelligent instrumentation, electronics, and multivariate signal-processing methods such as chemometrics and artificial neural networks. Increasing attention will have to be paid to the engineering of both the basic components and the entire devices.

The role of traceability in the animal protein production process, to meet consumer demands, remains a challenge, while practical solutions in the complete food chain are still missing. There is room for transfer of technology as well as the development of new devices and applications of new techniques and systems. It is in this area where agricultural engineers will play a key role in applying their knowledge of systems to improve sampling, calibration, and data analysis to provide instructions for a farmer or processor rather than raw data.

With the use of miniaturized electronic mechanisms it will be possible to record and control, at each time and in a more accurate way, events or diseases in order to optimize animal protein production. A biosensor array strategy, adaptable to multiple detections and analyses, will allow spreading development costs over several products. These improvements will produce devices that will be more competitive compared to the presently available instruments and will be able to operate under field conditions.

#### References

- 1. Holroyd, P. 2000. Tendency of the meat market for the new millennium. *Proc. APINCO*, 93-109 (in Portuguese).
- Stärk, K. D. C., M. Salman, Y. Tempelman, and U. Kihm. 2002. A review of approaches to qualify assurance of veterinary systems for health-status certification. *Preventive Veterinary Medicine* 56(2): 129-140.
- 3. Velasco-Garcia, M. N., and T. Mottram. 2003. Biosensor technology addressing agricultural problems. *Biosystems Engineering* 84(1): 1-12.
- 4. Caja, G., C. Conill, R. Nehring, and O. Ribo. 1999. Development of a ceramic bolus for the permanent electronic identification of sheep, goat and cattle. *Computers and Electronics in Agriculture* 24(1-2): 45-63.
- 5. Klindtworth, M., G. Wendl, and K. Pirkelmann. 1999. Electronic identification of cattle with injectable transponders. *Computers and Electronics in Agriculture* 24(1-2): 65-79.
- 6. Raven, E. T. 1989. Cattle Foot Care and Claw Trimming, 19-33. UK: Farming

- Press
- Carvalho, V. R. C. 2004. Effects of trimming on dairy cattle hoof weight bearing surfaces and pressure distributions during the stance phase. PhD diss. Gainesville, FL: University of Florida.
- 8. Adobe Premiere 6.5. 2003. Real-time editing for professional video production. Accessed on February 12, 2003, from http://www.adobe.com/.
- 9. Hu-m-an<sup>TM</sup>. 2003. The Human Movement Analysis Software. HMA Technology Inc., Box 42, King City, Ontario, Canada L7B 1A4. Accessed on 10 December 2002 from http://www.hma-tech.com/.
- 10. Matlab® 6.1. The Mathworks Inc. 2004. Accessed on 18 April 2004 from http://www.mathworks.com/.
- 11. Manske, T., J. Hultgren, and C. Bergsten. 2002. The effect of claw trimming on the hoof health of swedish dairy cattle. *Preventive Veterinary Medicine* 54: 113-129.
- 12. Stone, W. C. 2004. Nutritional approaches to minimize subacute ruminal acidoses and laminitis in dairy cattle. *J. Dairy Sci.* 87:(E.Suppl.): E13-E26.
- 13. Rutter, S. M., N. A. Beresford, and G. Roberts. 1997. Identification of the grazing areas of hill sheep. *Computers and Eletronics in Agriculture*. 17: 139-159.
- 14. Frost, A. R., C. P. Schofield, S. A. Beaulah, T. T. Mottran, J. A.Lines, and C. M. Wathes. 1997. A review of livestock monitoring and the need for integrated systems. *Computers and Eletronics in Agriculture* 17: 139-159.
- 15. Pedersen, S., and C. B. Pedersen. 1995. Animal activity measured by infrared detectors. *J. Agricultural Research* 61: 239-246.
- 16. Donát, P. 1991. Measuring behaviour: the tools and the strategies. *Neuroscience and Biobehavioral Reviews* 15: 447-454.
- 17. Korthals, R., L. MacDonald, and R. A. Eigenber. 1992. Experiences with transponders for monitoring bioenergetic responses. ASAE Paper No. 923010. St. Joseph, MI: ASAE.
- Nääs, I. A., D. F. Pereira, F. P. F. Curto, F. H. Behrens, J. C. C. Carvalho, M. Amendola, and E. C. Mantovani. 2000. Determining broiler female breeder's behavior using telemetry. *Proc. of the XIV Memorial CIGR World Congress*, (1): 1014-1018.
- 19. Curto, F. P. F. 2002. Behavior of female broiler breeder in distinct environment using RFID technology. PhD diss. Campinas, SP, Brazil: State University of Campinas, Agricultural Engineering College (in Portuguese).
- 20. Pereira, D. F. 2003. Determining the upper critical temperature and the thermo neutral zone for female broiler breeder using RFID. MS diss. Campinas, SP, Brazil: State University of Campinas, Agricultural Engineering College (in Portuguese).
- 21. Stapley, B. E. 1965. Automatic inspection of metal visible defects. *Iron and Steel Engineer* November: 1-4.
- 22. Aalderink, B. J., and M. W. C. DeJonge. 1976. Automated surface inspection of cold-rolled sheet. *Proc. International Meeting of Iron and Steel Making*, 11-19.

- 23. Saridis, G. N., and D. M. Brandin. 1979. An automatic surface inspection system for flat rolled steel. *Automatica* 15: 505-520.
- 24. Floyd, F. S. 1996. *Remote Sensing: Principles and Interpretation*. Los Angeles, CA: W.H. Freeman and Co.
- 25. Mollo Neto, M., O. Vendrameto, and J. P. Fusco. 2004. Automated systems for leather inspection: the machine vision–BASIS 2004. 6<sup>th</sup> IFIP International Conference on Information Technology for Balanced Automation Systems in Manufacturing and Services.
- 26. Souza, S. R. L., D. Moura, and I. A. Nääs. 2004. The use of image vision to study milking cow's behavior. *Smart Sensors 2004, Leuven.* 3<sup>rd</sup> *International Workshop on Smart Sensors in Livestock Monitoring,* (1)40.
- 27. Kearfott, K. J., R. J. Juang, and M. W. Marzke. 1993. Implementation of digital stereo imaging for analysis of metaphysis and joints in skeletal collections. *Med. Biol. Eng. Comput.* 31: 149-156.
- 28. Shao, J., H. Xin, and J. D. Harmon. 1998. Comparison of image feature extraction for classification of swine thermal comfort behavior. *Computers and Electronics in Agriculture* 19: 223-232.
- Xin, J., and B. Shao. 2002. Real-Time assessment of swine thermal comfort by computer vision. *Proc. of the World Congress of Computers in Agriculture and Natural Resources*, 362-369. ASAE. Paper No. 701PO301. St. Joseph, MI: ASAE.
- 30. Sergeant, D. M., R. D. Boyle, and J. M. Forbes. 1988. Computer visual tracking of poultry. *Computers and Electronics in Agriculture* 21: 1-18.
- 31. Overton, M. W., W. M. Sischo, G. Dechant, and D. A. Moore. 2000. Observations of dairy cattle behavior using time-lapse photography in a California free-stall barn. *Proc.* 33<sup>rd</sup> Annual. Conf. Am. Assoc. Bovine Pract., 138-139.
- 32. Haley, D. B., J. Rushen, and A. M. de Passillé. 1999. Behavioural indicators of cow comfort: Activity and resting behaviour of dairy cows in two types of housing. *Can. J. Anim. Sci.* 80: 257-263.
- 33. Herlin, A. H. 1997. Comparison of lying area surfaces for dairy cows by preference, hygiene and lying down behaviour. *Swedish J. Agric. Res.* 27: 189-196.
- 34. Martin, B. R., W. R. Prescott, and M. Zhu. 1992. Quantification of rodent catalepsy by a computer-imaging technique. *Pharmacology, Biochemistry and Behavior* 43: 381-386.
- 35. Olivo, R. F., and M. C. Thompson. 1988. Monitoring animals' movements using digitized video images. *Behavior Research Methods Instruments and Computers* 20: 485-490.
- 36. Spruijt, B. M., and W. H. Gispen. 1983. Prolonged animal observations by use of digitized video displays. *Pharmacology, Biochemistry and Behavior* 19: 765-769.
- 37. Shao, J., H. Xin, and J. D. Harmon. 1997. Neural network analysis of postural behavior of young swine to determine the thermal comfort state of the animals. *Trans. ASAE* 40(3): 755-760.

- 38. Hu, J., and H. Xin. 2000. Image-processing algorithms for swine postural behavior analysis. *Behavior Research Methods, Instruments & Computers* 32(1): 72-85.
- 39. Shao, B., and H. Xin. 2000. Real time assessment and control of pig thermal comfort behavior by computer imaging. ASAE. Paper No. 004015. St. Joseph, MI: ASAE.
- 40. Rossing, W. 1999. Animal identification. 1999. *Computers and Electronics in Agriculture* 24(1-2): 1-4.
- 41. Artmann, R. 1999. Electronic identification systems: State of art and their further development. *Computers and Electronics in Agriculture* 24: 5-26.
- 42. Geers, R., B. Puers, V. Gedseels, and V. P. Wouters. 1997. Electronic identification. *Monitoring and Tracking of Animals*. Wallingford, UK: CAB International.
- 43. Petersen, B., S. Knura-Deszczka, E. Pönsgen-Schmidt, and S. Gymnich. 2002. Computerised food safety monitoring in animal production. *Livestock Production Science* 76 (3): 207-213.

# 5.10 IT in Fish Farming

O. I. Lekang and B. F. Eriksen

Abstract. Information technology (IT) is becoming more and more important in performing intensive aquaculture. The areas in which IT can be and are being used are also increasing continuously. This section provides information on a selection of important fields in fish farming where IT is being used, including monitoring systems, production planning tools, advanced feeding systems, fish counting (including size and biomass estimation), and site monitoring.

**Keywords.** Fish farming, IT tools, Monitoring, Feeding systems, Fish counting, Biomass estimation, Aquaculture.

#### 5.10.1 Introduction

Although three-quarters of the world's surface is covered with water, only a few percent of all food produced comes from the sea. The world's population is growing rapidly and has now passed six billion. If we want to cover the world's food demand in the future, we will probably have to utilize more of this great production potential in the future.

The environment of aquatic organisms is not easy to understand without measuring chemical parameters. Fish farms are often located off the coastline or in deserted areas. Important information about water conditions, weather and wave development, unwanted visitors on the farm, etc., are often transferred with wireless technology or via telephone lines to a guard or to the head office. Modern sensor technology has given us the possibility of measuring important water quality parameters, and modern IT is an important tool to quantify, store, and communicate information in the aquaculture business.

Over the last 20 years, fish farming has become a highly industrialized form of biological production and plays an important role in the economy and settlement in many regions. Earlier, the demand was greater than the supply, but today it is quite balanced. This results in low economic margins, which promote intensively driven aquaculture facilities in which the production capacity is pushed to the limits with reduced water quality margins, shorter reaction time at the water stop, etc. Against this background, it is important to catch signals about bad water quality as early as possible. It is often too late to make arrangements when odd fish behavior or fish mortality is observed. By monitoring important water quality parameters and continuously evaluating the values, precautions can be taken at an early stage and great losses may be avoided. Modern sensor techniques and IT are the basic tools for solving this task.

Many farmers today are members of large multinational financial networks (Marine Harvest, Pan Fish, Main Stream, Stolt SeaFarm, etc.) These constellations have their own expert groups within the firm who organize exchange of know-how within the group. New knowledge is transferred between the farms via the networks, and the testing of new know-how and new production methods is organized through the same network. The network participants benefit from rapid information concerning markets and price development, new production methods, environmental threats, etc., to all farmers within a group or an area through an IT network.

In intensively driven land-based farms, many factors affect the water quality. The waterways are often used to such an extent that extra oxygen must be added, or water is recirculated. Intensively run farms must pay special attention to oxygen, carbon dioxide, pH, temperature, total gas pressure, and ammonia levels. In the case of an emergency, the changes in water quality are rapid due to high fish densities. Regulation systems must be thoroughly tuned in. These complex water quality processes are often supervised with IT-based monitoring and regulation systems. Systems like this often use field BUS technology to collect information in a control room on the farm. From this monitoring center, monitored values and alarm signals are transferred by phone and Internet to personnel on guard or in the head office. This makes modern information technology a suitable production tool to monitor, regulate, and send alarm warnings about environmental parameters of great importance to fish welfare and yields.

Input factors in production, such as liquid oxygen and feed, must be continuously brought to the farm. With modern IT, suppliers can connect directly to the farm's data system in order to supervise farmers in the use of their products and supply additional materials. Production results are often given to the suppliers who evaluate the production and report back to the farmer. Suppliers communicating online with the farm can supervise the customer in the use of products such as liquid oxygen, feed, etc.

Monitoring equipment for water quality is often located in a rough and damp environment. The reaction time after communication between the sensor and the monitoring center must be short, and the ability to solve a problem must be high. As the intensity of fish farming has increased, the monitoring and regulation systems in smolt farms often feature more than 100 measuring points, and the importance of each measuring point increases as the size of the units and the fish densities increase. Against

Against this background external help from outside the farm is often necessary to service and supervise the IT systems. An online support direct from the supplier is often established to reduce error search time and expenses for the farmer. The monitoring and alarm system will always require a quick, professional backup from an external supervisor. In many cases this supervisor can be the IT supplier, and an important part of this work is often done with direct online supervision from the supplier to the farmer

A sea cage system is not as tied to one locality as a smolt farm. It is often possible to move the production equipment from one locality to another. While a smolt farm is locked to buildings and pipe installation, a sea cage farm is not, and companies are continuously hunting for new and better locations for production. IT technology plays an important role in this work. When a possible site is located, documentation about temperature, oxygen levels, salinity, waves, and water velocity must be collected. This work is simplified by placing logging systems on the actual site to collect and store the actual environmental data over a period. These data are necessary in the evaluation of the production capacity of the potential sites. In this regard, IT is an important tool for appraising the suitability of a selected location.

The water quality in an intensively run fish farm is, as mentioned above, a complex system where the water quality through the farm changes continuously as a result of respiration, heating, pH-manipulation, oxygenation, feeding, etc. The complexity of the systems makes it difficult to foresee how a change in one of these parameters could affect the others. Over the past years, good water quality models have been developed as a tool for better understanding these unstable water systems, and this helps us to improve the water conditions in smolt farming.

Product history now plays a more important role in food marketing than ever before, and it is important to be able to prove to the consumer that the fish has lived under suitable conditions from hatching to harvesting. This includes information about important environmental parameters such as oxygen levels, temperatures, medical treatments, densities, etc. IT is the most important tool in this work. Product history is an important marketing tool, and the documentation of product history is often given through IT.

IT has grown to play an important role in modern aquaculture, and has become a very important factor in the development of this industry. All aspects can not be presented in this text. Selected important fields will therefore be presented, including monitoring systems, production planning tools, advanced feeding systems, fish counting (including size and biomass estimation), and site monitoring

More information about the topic is available in several textbooks [1-6]; [6] is mainly used as a basis for this description.

#### **5.10.2 Monitoring Systems**

On a fish farm, it is necessary to monitor different factors involved in the management of the farm; for instance, water quality factors like oxygen or pH, or water flow,

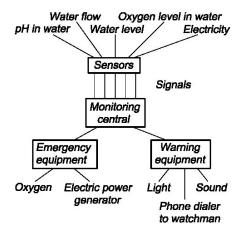


Figure 1. A monitoring system can be necessary on a fish farm, because the fish can represent a high value and failure in the equipment can be fatal (taken from [7]).

or whether the level in a head basin is correct. The need for regular monitoring is especially high in intensive fish farming with high fish densities and with strong reliance on technical installations that can fail. The time available for efforts to avoid problems is also limited in such cases and requires an automatic monitoring system (Figure 1).

To illustrate the economic consequences, an ordinary smolt farm with a yearly production of one million smolts, with a value up to one million euros, could be taken as an example. What happens when something goes wrong with the water inlet flow or water quality and nothing is done? Another example to illustrate the importance of a good monitoring system is a land-based fish farm for the farming of marine species. This requires the continuous pumping of water for a water supply. If the pumps stop, the water supply to the fish tanks stops immediately, and soon a critical situation for the fishes arises due to a lack of oxygen, which is transported with the water. The importance of continuous monitoring of the water flow running into the fish tanks and the pump is therefore obvious.

Monitoring systems used in fish farming normally consist of three major components:

- sensors and measuring equipment, which monitor the status;
- a *monitoring center*, which receives signals from the sensors and measuring equipment, interprets them, and eventually sends out alarm signals;
- warning and emergency equipment for when something fails.

Signal connections exist between the components. These are normally implemented as electrical signals through cables. Wireless connections may also be used.

# Sensors and Measuring Equipment

The task of the sensors is to control the conditions. Actual sensors in fish farming can be for monitoring water level, water flow, electricity, or water quality parameters such as oxygen level or pH. The number of sensors determines the size of the monitoring system. Electrical signals are sent from the sensors to the monitoring center, either

as digital or analog signals. Normally, these are standardized low current electric signals between 4-20 mA.

#### **Monitoring Center**

The construction of the monitoring center that receives signals, interprets them, and eventually sends out signals to the warning equipment depends on the complexity of the system. Even if the monitoring center is normally built only for fish farming purposes, it is either based on a PLC (programmable logic controller) or PC (personal computer).

A PLC has a number of input channels where electric signals are received from the sensors, and a number of output channels where electrical signals are sent to warning or regulation equipment. The output signals can be programmed based on the input signals. Both analog and digital signals may be used.

During the past years, modern BUS communication systems have become more common for the information transfer between sensors and master units like PLCs and PCs. This makes it possible to incorporate an unlimited number of measurements into one system.

To illustrate the function of a PLC monitoring system in fish farming, consider this example. A simple monitoring system is to be installed on a farm. The equipment on the farm includes a pump that delivers water to a head basin. Parallel to this pump is a backup pump in stand-by position. From this head basin, gravity causes the water to flow into the fish tank. The inlet water covers the oxygen requirement of the fish. The following measuring instruments are included in the farm: a level sensor in the head basin and oxygen sensors in each fish tank. Pure oxygen gas can be added from the gas bottles to the fish tanks directly through diffusers on the tank bottom. In a normal situation no oxygen is used. A solenoid valve (NO) connected to the emergency oxygen controls whether or not oxygen shall be added. This valve remains in a closed position if the oxygen level in the tank is acceptable. The regulating PLC can be programmed as follows: If the water level in the head basin is too low, a signal will be sent from the level sensor to the PLC. The PLC is then programmed to start the backup pump and the water level in the header tank will probably return to normal. If the water level keeps dropping, the problem must be elsewhere. If the header tank is empty for a while, the oxygen content in the tank will drop. The signal from the oxygen sensors will drop below a preset value in the PLC, and a signal from the PLC opens a solenoid valve for emergency oxygen directly into the fish tank through diffusers. A siren, light, or telephone alarm is often used in addition.

In advanced monitoring systems for fish farms, computers are commonly used as the monitoring center. The computer is normally equipped with special cards, which makes it possible to import and export analog and digital signals (I/O cards). An advantage of the use of a computer is that the input signals can easily be recorded for later use. It is also possible to achieve a hard copy of alarm times and a report on sensors that have registered the incident. It is also easy to obtain a picture of the farm, including the sensors and the status of each sensor, on the computer screen, which makes visual control easier (Figure 2).

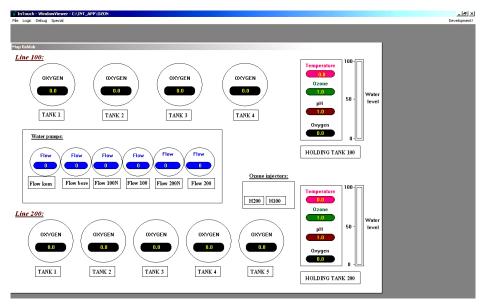


Figure 2. A visualization of an experimental plant at the Norwegian University of Life Sciences, where the status of all monitoring sensors is included.

## Warning and Emergency Equipment

Output signals are sent to warning equipment from the monitoring center if something fails. On a fish farm, this normally includes equipment that creates both light and sound as warning lights and sirens. A phone dialer may also be included if there are no permanent watchkeepers on the farm. This dials fixed programmed phone numbers to nearby watchkeepers.

Output signals may be used to start emergency equipment or to regulate the functioning of equipment. If there is, for instance, an electrical power failure, a generator that produces electric power can be turned on. It may also open a valve for the supply of emergency oxygen, which is applied directly through diffusers on the bottom of the fish tanks. Often both emergency equipment and warning equipment are started together.

#### Maintenance and Control

Maintenance and control is extremely important in monitoring systems. In fish farming, the sensors are exposed to fouling, especially those that measure water quality like oxygen and pH. The maintenance includes calibration of the sensors. It is advisable to have fixed routines for testing the monitoring system, for instance once a week. The sensors are then tested to be sure that a signal is sent out if something fails and that signals are sent from the monitoring sensor to the warning equipment. Eventually emergency systems such as the electric power generator and systems for adding of oxygen must also be tested at fixed intervals to be sure that they will function if a real emergency situation occurs.

#### **5.10.3 Production Planning Tools**

According to figures calculated by the World Food Organization, fish farming is the fastest growing food industry activity worldwide. This has created an increasing demand for documentation and traceability. IT software has been developed over the last 10 years that can cover these needs.

- Control—In a fish farm there are many biological parameters to take into consideration if the site's production potential is to be properly utilized. The farmer needs reliable datasets and recorded registrations of all key fish farming functions including origin, stocking densities, grading, feeding and growth, treatment and vaccination, fish mortalities, harvesting, and sales. These figures must be combined with environmental registrations such as water temperature, pH levels, light, oxygen levels, ammonia levels, carbon dioxide levels, etc., in an effort to optimize the site's production capacity. Many different registration and water quality calculation programs are used for these purposes.
- Plan and compare biological and economical performance—The control and registration programs must be able to communicate with each other making it possible to create software tools that take all aspects into consideration. Today, updated management systems are able to combine information from automatic feeding systems with estimated biomass development, water quality models, price development in the fish market, and economical modeling. This can be done for each specific farm and for a combination of many farms within a group.
- Integration—The management software can be used individually on-site or linked to a terminal server application for larger multi-site operations. This enables centrally based managers to access current on-site data, providing crucial information for making management decisions, planning, and comparing fish units, sites, or the whole operation.
- Traceability—As of 1 January 2005, the EU requires full traceability for all seafood products sold within the EU through the European directive "Tracefish."
  The same requirements are also a necessary part of business in North America.
  Requirements for traceability will probably become a part of all international trade with food in the future. The consumer's major concern is the need to verify medication and vaccine use on the animal. This can be shown in history reports and product certificates following each fish through the production line to the consumer.

As the aquaculture industry grows and matures, and the quality of production and production techniques increases, it is being faced with an ever-increasing demand for quality monitoring and product traceability. This traceability should cover important areas such as where the fish is grown, and which feeds, vaccines, and medicines were given throughout its life history. The challenge lies in the fact that salmon have a life cycle that includes growth at two or more localities: the hatchery, freshwater for growing, seawater to grow out, harvesting, transport vehicles and boats, and egg production in breeding stations. All participants in this production line have their own system for

environmental registration, growth registration, etc., and all these systems will have to be able to communicate with common software if the market demands for traceability are to be properly fulfilled in the future.

#### 5.10.4 Advanced Feeding Systems

In fish farming, equipment for automatic delivery of feed is commonly used. In the more simple systems, this is only a simple feed machine equipped with a hopper, a dispensing unit, and a control unit. The control unit controls the current to the motor in the dispensing system. The simplest control units include an interval relay, where the time-length of the meal and the time interval between meals are set. It may also be equipped with a light-sensitive photocell, which only permits feeding during daylight hours. In more advanced control units, it is possible to have a daily increase in the time-length of the meal. This is set in relation to the expected growth rate of the fish in the production unit.

The control unit may be individual for each feeder or it may control several feeders with the same feeding regime. There can also be one control unit with several channels, which means that it can control several feeders individually. Several relays connected in a multi-channel controller, a PLC, or a PC, can be used for this purpose.

Instead of the simple feeder, feed may be taken directly from the feed silos, transported to the fish production unit and, at the end, distributed to the fish. Such systems are called feeding systems and include three major parts: a storing unit, a transport unit, and a feed distribution unit. Two different principles are used in feeding systems: (1) systems where the feeder is centrally placed and feed is transported to the single fish production units (tanks, ponds or cages) through pipes, normally called a central feeding system, and (2) systems were the feeder is installed on a rail system that covers several units, normally called feeding robots.

#### Central Feeding System

A central feeding system consists of feed silos, a sluice valve, tubes with a flow of water or air for transporting of feed, a selector valve, and eventually a distribution unit, in addition to the control unit that controls the system (Figure 3).

Feed is delivered from the silo into an auger that brings the feed particles into a hopper above a sluice valve. The sluice valve brings the feed particles from the hopper into the pipes for further transport to the tanks or cages. Water or air is used as a medium to transport the feed particles and the sluice valve therefore also represents an air or water lock between the hopper and the transport medium. The velocity in the tubes, whether air or water, is used as a transport medium, to ensure that the feed particles always stay in suspension. A blower or a pump secures adequate velocity inside the tubes. After a short transport distance of some meters in the pipes, the feed enters the selector valve. The selector valve selects the production unit to which the feed portion shall be sent. After having passed the selector valves the feed is transported to the production unit through tubes. In sea cages the length of the tubes may be up to several hundred meters.

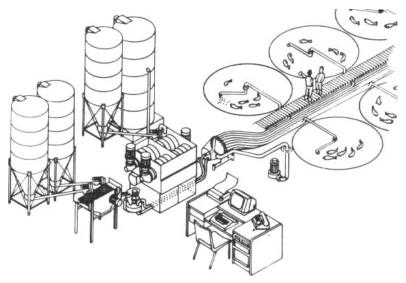


Figure 3. A central feeding system takes feed from silos and transports it to the fish [10].

In this type of feeding system a central computer controls when the components in the system shall start and to which unit the feed shall go. The amount of feed to be sent to the different units can be set on the computer. The computer may also do this automatically if we add the initial weight of the fish, the number of fish, water temperature (may also be collected automatically), expected growth and mortality. A great advantage of this system is that the computer also stores the inputs and therefore represents an important tool for production planning and production control.

In the case of a power cut or other function stoppage, the computer recalculates the feeding intensity to ensure that the correct daily amount of feed is given.

#### Feeding Robot

A feeding robot is a motorized feeder suspended in a rail system, hanging above fish tanks (Figure 4). In addition to passing through the various production units, the rail system passes the feed silos for refilling. When the robot is feeding it moves along the rail until it hits a chip that is attached to the rail over each tank. Based on information in this chip the feeder (robot) recognizes the tank. On the robot there is a computer in which the amount of feed for the actual tank is programmed. When the robot hits the chip it delivers the programmed amount of feed to the tank. Then it moves to the next tank and further on. When the robot is empty, it automatically returns to the silos for refilling and communication with the master computer. The electricity supply to the feeder and the pushing unit may be an integrated part of the rail system or it can be battery operated and recharged in a docking station. The great advantage with this system is that the same feeding mechanism can be used for the feeding of several tanks. A higher investment can be made for the feeding unit such so that it can be designed to be more accurate. A robot feeder is normally controlled by a computer.



Figure 4. A feeding robot is a motorized feeder suspended in a rail system.

### Automatic Feeding Control

The appetite of fish is affected by outside influences such as variation in water temperature, water quality, waves (in cages), light conditions, etc. With a normal control unit used on the feeder, a fixed amount of feed is distributed during a given period. There is no adjustment of feed amount to match variations in appetite. Feed loss during feeding often occurs under such conditions, particularly in sea cages where a large feed volume is used. A manually operated method is to use a submergible video camera under the cage and watch randomly for feed loss when the feeder is running.

In sea cages, a number of methods for automatic appetite control have been introduced. Such systems include the registration of the feed loss, stopping the feeder when a given level is reached. Various systems are available to detect the feed loss, including hydro-acoustic sensors, photocells, and Doppler signals. The detector can be located under the cage or inside the cage in order to measure the feed loss on a sample area.

In *dynamic feeding systems* we go even further and use the feed loss to control the amount of feed to be delivered (Figure 5). One of the systems includes a collector for feed loss placed inside the cage. When the feeding starts, the pump takes the eventually collected lost feed lying in the collector from the previous feeding and pumps this through a pipe that delivers it at the top of the cage bag. In this pipe circuit an IR detector detects eventually unconsumed feed. If there are no feed particles left, no particles are detected, and the feeder starts to add a new portion of feed. The same procedure happens with the next feeding. In this manner, the system will follow the appetite of the fish in a dynamic way: increasing appetite leads to increasing amount of feed, decreasing appetite causes a decrease of the amount of feed. Such systems ensure no feed loss, and may give an early warning of possible odd behavior. A specially designed control unit is necessary for this system.

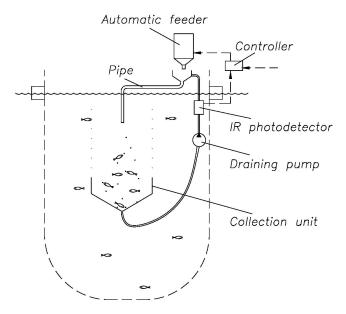


Figure 5. A dynamic feeding system [8].

#### 5.10.5 Fish Counting and Biomass Estimation

#### Counting of Live Fish

For a variety of reasons in fish farming, live fish must be counted: the sale of young fish, production control, and harvesting. Done manually this represents a major work investment on large farms were the number of individuals easily reaches several hundred thousand. The problem with automatic fish counting is to separate the individuals and to avoid two or more fish coming together and being counted as one.

Counting can be done with the fish either in or out of the water. Voluntary counting of free-swimming fish has proven to be difficult, and so far no commercial methods have been developed. This is due to the fact that it has been impossible to control the behavior of the fish. Some of the equipment used for counting may in addition be used for measuring the fish size (see Section 5.10.6).

A quite cheap and simple principle used to separate the fish when they are out of water is to let them slide in a V-shaped channel with a concave orbit. The concave orbit ensures a gradual increase in the velocity of the sliding fish, which will separate individuals. It is then possible to count the single fish, for instance with the use of light-sensitive cells. The light diodes create a beam that is broken by the sliding fish, and the fish is then counted.

Counting fish in water is normally done in connection with fish pumping or during pipe transport, where the fish is forced to participate. Traditional video cameras or linear cameras and image analysis may be used. The challenge is to get a good picture, with sufficient contrast to the background, because a contrast picture (black and white)

is normally used. Thus, a photo chamber with good light conditions is used. The next challenge is to avoid having two fishes come together and be counted as one. This is solved either by the use of two cameras or by one camera and a mirror, actually achieving two pictures. With image analysis of the catch pictures it is possible to interpret whether there are one or several fishes in the picture with quite good accuracy. The same system may also be used to calculate the fish size (see Section 5.10.6).

The difference in conductivity between water and fish may also be used in counting. The measured conductivity of a pipe will be increased when a fish is swimming inside. It is, however, difficult to separate between two fishes coming together.

#### 5.10.6 Measuring of Fish Size and Total Fish Biomass

Monitoring of fish size and the total biomass is important for production planning and production control. It is also useful for deciding when to grade and for planning the harvest. Traditional methods include taking fish samples, lifting them out of the water, and weighing them. This is time consuming and the accuracy of the sample can be rather poor, especially if samples are taken from large sea cage units. Based on these facts, automatic measuring of fish size and total biomass is of interest. Two general approaches are: (1) the fish stays in the production unit and the equipment is lowered in, and (2) the fish is taken out of the production unit; normally this involves equipment connected to a fish pump.

#### Measuring When the Fish Stays in the Production Unit

Several systems have been developed, mainly for sea cages, to monitor the fish size or the total biomass or a combination.

One principle used is a *measuring frame*, which is lowered into the cage (Figure 6). In the walls of the frame there are a number of light-emitting diodes. When the fish

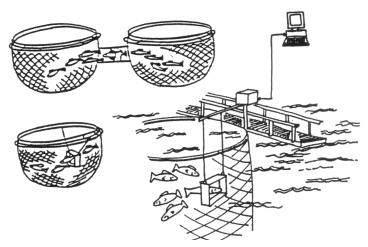


Figure 6. Measuring frames for measuring fish size [11].

swims through the frame, a shadow picture is created by the light diodes. By a computer, the approximate fish size is calculated, based on image analysis on the shadow picture of the side of the fish, which is highly correlated with its weight. During a given period of time, a casual sample of fish will voluntarily swim through the frame. This occurs in cages because the fish swim voluntarily around in the cage, but not in tanks where the fish stand still. If two fish pass the frame together this is detected during the image analysis and the computer does not use this picture. Only the passages where a single fish passes through the frame are used for calculation of the fish size. Based on the large number of individual measurements of fish, the average fish weight and its variation (standard deviation) can be calculated for the total cage. To know the variation can help us to predict when grading is necessary.

If the average fish size is known it is possible to calculate the total biomass providing that the number of fish in the cage is known. When such a system is used for calculating the total biomass, reliable mortality control is a must. Uncontrolled escape of fish must also be avoided.

Another method for finding fish size in a cage based on camera technology is by application of stereoscopy. In a unit, two video cameras are located above each other with a known distance between the cameras. By taking pictures of an object, a fish in this case, with both cameras and with the use of mathematics and geometry it is possible to calculate the distance to every point on the object. By using image analysis, the length and height of the object can be calculated. For fish there is a high correlation between these values and the fish weight.

Another method uses an echo sounder. This is the same principle that is used in traditional fishery to find out where and at what depth the fish is swimming. A source that sends sound impulses at fixed intervals is placed under the cage [9]. When the sound impulse hits a fish, a reflection signal, an echo, is achieved. A receiver fixed near the sound source beyond the cage receives this reflection, which is moving in the opposite direction of the impulse that was sent out. The functioning is as follows: first a short signal (a sound wave) is sent out, for instance for one thousandth of a second. Then the receiver listens for echoes for parts of a second and then a new sound impulse is sent out. Several echoes are achieved over the course of a minute. The size of the echoes depends on the size of the individual fish or the fish shoal, which is used in cage farming. It also possible, to some degree, to observe the behavior of the fish and the depth at which the fish remain in the cage. When the echo sounder registers an echo, the swim bladder represents a good echo because it is filled with air. Echo sounders are not suited for accurately measuring the size of a single fish. One reason for this is that the echo from the swim bladder is highly dependent on the orientation of the swim bladder in relation to the sound source. The size of the echo will vary if the fish is swimming upwards, downwards, or standing horizontally. Also, echo sounders cannot be used in tanks because sound reflections from the bottom and the walls disturb the measurements



Figure 7. A unit where a video camera installed on a well boat is used to measure fish size.

#### Measuring When the Fish Is Removed from the Production Unit

A video camera can be installed at a pump pipe or another pipe. It takes pictures from the side, and possibly from above, and calculates the size based on image analysis of the pictures (Figure 7). The system is the same as described earlier for counting. The only difference is the image analysis program used for fish size calculation.

#### **5.10.7 Site Monitoring**

Before establishing a fish farm on land or at sea, the production conditions must be investigated and documented properly. For a land-based smolt farm, environmental limitations such as rainfall during the year, pH fluctuations, temperature variations, etc., are of great importance.

Oxygen value variations at different water depths in the fresh water source must be recorded to evaluate a possible site. Proper registrations of this kind must be carried out for at least one year before one should consider establishing a farm on the site. Oxygen, pH, and temperature sensors must be put out at strategic points on the site for documentation. These sensors can be connected directly to a waterproof logger close to the registration point, or the measurement signal can be transmitted wirelessly from the logger to a computer in the area. In order to avoid too many registrations, it is recommended to reduce the number of registrations to about once a day.

In a sea cage farm, this site monitoring often becomes a continuous process since a sea cage farm is easier to move to a new site than a land-based farm. Environmental registrations in possible sites in the area around a sea-based farm are therefore often a continuous process. When checking out new sites at sea, water current, waves, and oxygen as well as temperature levels in different water layers are important parameters to consider. These parameters can be used to evaluate the production potential of the site.

#### References

- 1. Wheaton. F. 1977. Aquacultural Engineering. New York, NY: John Wiley.
- 2. Lawson, T. B. 1995. Fundamentals of Aquaculture Engineering. New York, NY: Chapman and Hall.
- 3. Hochheimer, J. 1999. Equipments and controls. *CIGR Handbook of Agricultural Engineering, Part II: Aquaculture Engineering*, ed. F. Wheaton, 281-307. St. Joseph, MI: ASAE.
- 4. Willoughby, S., ed. 1999. *Manual of Salmonid Farming*. Oxford, UK: Blackwell Science.
- 5. Huguenin, J. E., and J. Colt. 2002. *Design and Operating Guide for Aquaculture Seawater Systems*. Amsterdam, The Netherlands: Elsevier Science.
- Lekang, O. I. 200? (in preparation). Technology for Aquaculture. Blackwell Science.
- 7. Lekang, O. I., and S. O. Fjæra, 1997. *Technology in Aquaculture* (in Norwegian). Landbruksforlaget, Oslo.
- 8. Skjervold, P. O. 1993. Fish feeding station. *Fish Farming Technology*, eds. H. Reinertsen, L. A. Dahle, L. Jørgensen, and K. Tvinnereim, 269-274. Rotterdam, The Netherlands: Balkema Publishing.
- 9. Dunn, M., and K. Dallard. 1993. Observing behaviour and growth using the Simrad FCM 160 fish cage system. *Fish Farming Technology*, 269-274. Rotterdam, The Netherlands: Balkema Publishing.
- 10. Sketch from Akvasmart Asa, Bryne Norway.
- 11. Sketch from P. H. Heyerdal, Norwegian University of Life Science.

# 5.11 Advanced Life Support Systems in Space

D. H. Fleisher, L. F. Rodriguez, A. J. Both, J. Cavazzoni, and K. C. Ting

Abstract. Human life support systems for long-duration manned space missions need to be robust, reliable, self-sustaining, and efficient in terms of mass, power, and volume. The National Aeronautics and Space Administration (NASA) administers the Advanced Life Support (ALS) program, in which biological systems are combined with physical and chemical technologies to accomplish these goals. Hydroponic, controlled environment production of higher plants, such as salad crops, as well as wheat, rice, white potato, and soybean, is expected to help satisfy requirements for crew nutrition, atmospheric oxygen and carbon dioxide recycling, potable water production, and waste treatment and resource recovery. Successful ALS systems require the development and application of information and decision support tools, growth monitoring techniques, and sophisticated analysis methods, some of which will be described in this section.

**Keywords.** Advanced Life Support (ALS), Controlled environment, Crop production, Information systems, Modeling, Systems analysis.

#### 5.11.1 Introduction

The basic requirements for sustaining human life include air, water, food, and, ultimately, waste treatment. Traditional life support systems for sustaining human life during space missions, such as those used for the space shuttle and the International Space Station, use physical and chemical technologies with resources being stored and/or re-supplied from Earth as needed in order to provide for these requirements. As space missions of increasing distance and duration from Earth are planned, the ability to re-supply and store resources on-site declines due to increasing logistic and economic challenges. Establishment of a permanent human presence in space will require self-sufficient, highly reliable, Earth-independent life support systems capable of (re)generating air, water, and food on-site. These significant challenges are studied in the National Aeronautics and Space Administration's (NASA) *Advanced Life Support (ALS)* program, in which biological systems are combined with physical and chemical technologies to accomplish these goals [1].

The use of organisms such as microbes and higher plants to recover and recycle nutrients and resources from waste products in a closed system is often referred to as bioregenerative life support (BLS). This term implies that the success of an ALS system depends on the synergistic integration of biological systems with other life support components. Higher plants are grown to produce food for the crewmembers, reduce atmospheric carbon dioxide, generate oxygen, produce potable water through transpiration, and provide some wastewater treatment through hydroponic production systems. Several candidate crops have been identified for use in ALS biomass production systems as a result of their nutritional value and ease of cultivation. These crops include soybean, wheat, white potato, sweet potato, rice, peanut, tomato, carrot, cabbage, lettuce, and strawberry, among others [2]. Because the stability of an ALS system requires substantial coupling of plant processes with chemical and physical technologies, the biomass production system needs to be carefully designed, highly monitored, and controlled [3]. A substantial amount of research has been conducted for the purpose of identifying the best cultural, management, and environmental control practices that will maximize edible biomass production, while minimizing the input reguirements. Further research efforts have been focused on developing technologies to support future operation and management of such systems in real-time. These include efforts in nondestructive monitoring of plant growth and development [4,5], development of decision support algorithms [6], mathematical crop modeling [7,8], and coupling of measurement techniques with control and modeling algorithms for the purpose of real-time feedback control [9]. This demand for highly controlled and monitored plant growth production on such a small scale can be viewed as the ultimate form of precision agriculture.

Food processing is generally needed in order for the edible portions of the plants to be transformed into meals for the crew. Waste processing and resource recovery is required to recover water and nutrients from crew and plant waste, and to maintain the desired atmospheric conditions. The crewmembers form the main component of the ALS system through the consumption of food, oxygen, and water, and through the production of carbon dioxide, and solid and liquid wastes.

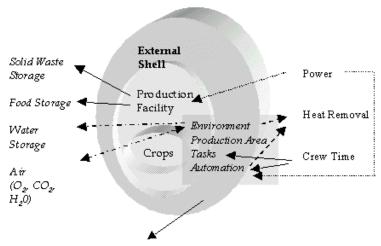
A simple systems abstraction of an ALS system reveals major subsystems for the biomass production component (BPC), waste processing and resource recovery (WPRR), food processing and nutrition (FPN), and the crew. An implied fifth subsystem is an interconnecting area (e.g., a hallway with storage capabilities) serving as the interface for the other four components. NASA has funded a significant amount of discipline-specific research within each of these ALS components over the past few decades, resulting in various knowledge bases and information sources. In order to properly design, plan, assemble, operate, and manage a fully integrated ALS facility, however, a major challenge is to identify and compile results of various research projects and knowledge bases within each subsystem so that the various technologies that make up an ALS system will function according to the overall design specifications. The techniques involved in systems analysis, integration, processing, and organization of information are critical for the design of a successful ALS system.

In fact, an ALS project objective is to "develop and apply methods of systems analysis and engineering to guide investments in technology, resolve and integrate competing needs, and guide evolution of technologies" [10]. This includes refinement and development of systems tools to perform general life support mission trade studies or simulations, guide current and future research efforts, define technology requirements for various life support missions, and develop methods to facilitate integration of concurrent science and engineering [11], so that technologies on the subsystem level may be successfully integrated into the broader systems level. This chapter describes various information flow and system studies activities within the ALS community designed to achieve this goal.

#### 5.11.2 Systems Studies and Analysis

Systems analysis is used to investigate the performance of a system as a whole by understanding all of the system components and their interrelationships. It is useful to perform systems analysis during the development and design of a new system or of an existing system. It is especially important to utilize system studies and analysis techniques in the development of novel complex systems, such as ALS systems. The objective for the analysis, for instance a specific case study or a tool-building exercise, will influence both the type of information required and the appropriate systems analysis procedures to be utilized. For example, there is a difference whether the analysis is to be based on a specific ALS mission scenario (e.g., four crewmembers, 600-day mission, 50% on-site food production, etc.), or whether the goal is to develop a tool to study various ALS mission configurations.

Systems analysis first proceeds with a systems abstraction, or decomposition of the system in question into relevant subcomponents or processes. Interactions between these subcomponents must also be identified. Figure 1 illustrates an abstraction for the BPC. Differently controlled environment compartments are used to separate plant species from the ambient environment of the main life support system, denoted by the "external shell." Plants are being cultivated in a properly designed production facility. As individual crops grow, they interact with the environment and material resources,



V olume, Mass, E quivalent System Mass

Figure 1. An example of systems analysis for a BPC of an ALS system.

The critical subsystems of a BPC are identified and potential interactions with other ALS system components are identified using descriptor variables [12].

consuming carbon dioxide, nutrients, and water, and producing biomass, oxygen, and water vapor. Equipment and machines (automation) are used to handle tasks and monitor information gathered from the plant production environment. Human and material resources are then managed to keep the production on schedule so that edible products can be delivered on time. The relationship between the BPC and the entire life support system can then be monitored through the use of descriptor variables as shown in Figure 1.

Once the objective and system to be analyzed are defined, five essential aspects of systems analysis must be considered (Figure 2). The available sources of information

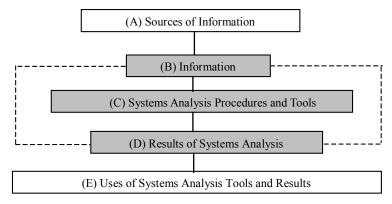


Figure 2. Essential aspects of systems analysis [13]. Items B, C, and D are the focus of the majority of systems analysis efforts.

must be identified (A), along with the type, form, and organization of this information (B). For example, is the information quantitative or qualitative? Does it exist in peer-reviewed journals or in conference proceedings, or has it been gathered into a centralized data or knowledge base?

Procedures and tools (C) refer to the development of mathematical models, algorithms, and other devices required to process the relevant information in order to understand and predict the behavior of the system under investigation. This typically consists of several steps, depending on the objective of the study. Key descriptors or variables of the system are identified that will provide a method to quantify or track the performance of the system. For ALS, a systems analysis may include the study of mass, power, volume, crew time, edible biomass, solid waste storage, and atmospheric oxygen and carbon dioxide concentration (Figure 1). Relationships among these descriptors are established using mechanistic or empirical equations, logical or relational algorithms, or conversion factors. Next, it is typical to assign system performance indicators. The type of indicators selected depend on the objective of the study, but are generally assigned to track system behavior, indicate redundancy in research or knowledge bases, or optimize the system for specific mission scenarios.

A model is then developed that represents the system, including the interactions between all descriptors and performance indicators. Mathematical modeling is one of the main methods for developing tools for ALS systems analysis and related research, and is an active research area within the ALS community. This modeling can be broadly categorized as being either a top-level (more generalized) approach or a bottom-up (more detailed) process-level approach. Modeling efforts within the ALS community for the BPC have focused on both specific crops (via process-level approaches), and on the BPC subsystem (via top-level approaches). Efforts have also been devoted to modeling other subsystems and integrating them into a top-level model for the entire ALS system.

Finally, each model is verified and validated by real-world data, interpretation of results by domain experts, physical principles or laws, etc. Simulations are then conducted using the validated models. Results of the analysis (D in Figure 2) may include evaluation of the workability and reliability of the system in accomplishing the desired goal, optimization methods, resource requirements, economic viability, and may indicate areas in which additional research is required. For example, in the case of the BPC, the interest may be in determining the optimal layout of the production facility so that requirements for labor (crew time) and power are minimized.

The last step (E) is needed to determine how to utilize and/or disseminate results of the analysis. If systems analysis tools, such as decision support systems or computerized software tools, have been delivered as a result of steps C and D, they should be utilized to increase the quality of future systems design. By the implementation of systems analysis techniques, subject areas where new knowledge is needed will be identified so that future research and development efforts may be more systematically approached. Thus, the essential aspects of systems analysis may need to be repeated several times for the same project until the optimal system design is determined. As a result, the use of analysis tools and results will lead to new information sources to be

included in new analysis procedures and tools, thereby increasing the quality of the results of systems analysis and increasing the success of the system design.

#### 5.11.3 Information and Knowledge Sources for ALS systems

One of the objectives of systems analysis and integration is to find ways to make information more useful, available, and accessible. Because the results of an analysis are only as accurate as the information source (item A in Figure 2) and the type of available information (B), effective identification and utilization of information sources and knowledge bases is critical. However, knowledge bases from different, specific discipline-oriented research areas tend to be in disassociated forms. Frequently, piecewise forms of knowledge about individual components are abundant, but information about the interrelationships among the pieces is lacking [14]. This is particularly true in research where information is generally disseminated to the ALS community through literature, personal and public communications, training programs, meetings, and other events. Because these information exchanges often take place within the confines of particular academic disciplines, it is quite common for knowledge bases to be disjointed even within the same ALS interest (e.g., engineers, microbiologists, and plant physiologists publish in different journals, even though all may be involved in BPC research).

Methods for manipulating or restructuring the data are therefore required in order to facilitate the analysis. These include the ability to define or unify representation of common parameters, identify or establish the relationships among necessary knowledge bases, modularize separate knowledge bases so that compatible building blocks for larger knowledge bases can be formed, and apply optimization techniques to the whole system under consideration [15]. In addition, complex, multidisciplinary systems such as those integrating biological with physicochemical processes present additional challenges in that information sources frequently contain variable, incomplete, imprecise, and uncertain data.

In order to address these challenges, a group of systems researchers within the ALS program called the *SIMA* group (systems integration, modeling, and analysis), have been working on organizing knowledge databases and developing systems analysis tools. This effort is intended to support the integration of various discipline-separated knowledge bases by documenting assumptions that are to be used in all ALS system analyses. In addition to centralizing systems level information, SIMA has also drafted documents to provide a list of possible mission objectives and technology requirements.

The ALS Reference Mission Document [16] provides mission details and defines realistic technology options for four different mission scenarios including the International Space Station and potential future Mars missions. This information is to be used to support meaningful systems-level analyses that address integration of all ALS research in support of mission goals defined in the document. The document addresses the required components and subsystems for each mission. Some of the available technologies and unit processes (physical, chemical, and for future Mars missions, biological) are discussed. Diagrams integrating the various technologies and subsys-

tems through identification of mass flows are provided. Thus, the document enables systems researchers to identify the specific technologies and subsystems that must be integrated into a functional ALS system for different missions and serves as the initial systems abstraction for which the essential aspects of systems analysis can be conducted.

The ALS Requirements Document and Design Considerations [17] provides desired performance requirements and design considerations of the different subsystems. ALS functionality is decomposed into several areas including environment control, emergency response, resource provisions, waste management, and support for extravehicular activities. The document lists quantitative requirements that can help direct research efforts for satisfying these functions. For example, crew nutritional needs, life support requirements such as atmospheric pressure, gas composition, and water requirements, and design guidelines for the biomass production systems are all specified

The Baselines Values and Assumptions Document [2], or BVAD, provides more quantitative information on overall technology processes and subsystem requirements. The BVAD is a common database or shared information source to be used by modelers and systems analysts which consists of data gathered from all ALS research areas. This information is intended to begin to integrate the various ALS databases into a shared set of information that can be utilized by researchers in the various disciplines. It provides a range of data for top-level process characterization. For example, for plant growth, the BVAD lists information on average and maximum reported growth rates, harvest indices, and nominal and optimal growth conditions for each ALS candidate crop. For various waste processing unit processes, the flow rates of material inputs and outputs are defined and resource requirements such as power, mass, and volume are provided.

SIMA has also been developing an ALS Equipment database [18]. The purpose of this database is to be a unified and consistent source of ALS equipment information. This type of information can be utilized for the sizing of ALS equipment for subsequent analyses. By providing a single source of data, SIMA hopes to achieve several goals: (1) minimize repeated time-consuming data searches, (2) minimize data transfer errors by taking advantage of the electronic format, (3) add further credibility to analyses by providing a consistent source of documented and carefully reviewed data, (4) improve the understanding of the necessary system design assumptions required to incorporate a given technology into a system, and (5) capture knowledge regarding technology performance and determine the distinguishing factors between technologies that perform similar purposes.

#### 5.11.4 Examples of Systems Analysis and Integration for ALS Systems

In this section, several examples of systems analysis work applied to the BPC and ALS systems in general are described.

#### **ACESYS**

Following the structure presented in Figure 2, most systems analyses efforts concentrate on the items B, C, and D with an understanding that B is significantly affected

by A, and D is heavily driven by E. In most cases, identifying useful B and D becomes a great challenge. The information directly obtained in A may not be readily usable for C. The needs of E may not be clearly communicated to make D completely defined. A substantial amount of effort may be needed to reach workable formats for B and D. Finally, there remains the question of how to go from B to D as seamlessly as possible.

In order to facilitate overall systems analyses and integration for ALS systems studies, an interactive web site was developed with the intention to provide a central platform including all essential aspects of ALS systems analysis (Figure 3). The ACESYS platform represents an integrated automation-culture-environment (ACE) analysis system developed to facilitate and automate information gathering, processing, analyzing, and presenting [19]. The goal was to establish concurrent science communication capabilities within the ALS community and develop methodologies and tools for value-added information processing using the Internet cyber environment. Key components included:

- an input mechanism for research project information,
- a set of databases for storing discipline oriented information,
- a display capability for presentation of data,
- an on-line discussion forum, and
- JAVA applets and software programs for processing information.

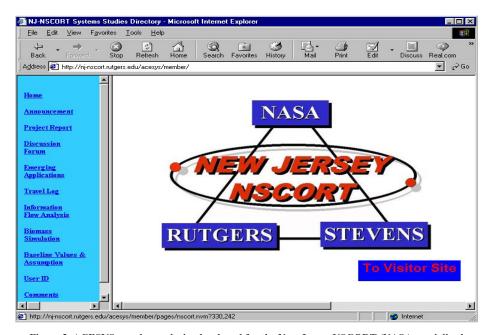


Figure 3. ACESYS member web site developed for the New Jersey NSCORT (NASA specialized Center of Research and Training), a five-year research project for the development of bioregenerative life support systems. In the left column are various options for relevant announcements, research project reports, discussion threads, software applications, and informational database access.

#### Information Flow Analysis

Another novel approach in information technology developed during the NJ-NSCORT project included a modeling tool that could be used by ALS project management to ensure selection of appropriate research projects that cover the entire range of ALS research needs. The information flow analysis (IFA) developed by Rodriguez [20] consisted of a JAVA applet that categorized the various research projects according to four decision making metrics: completeness, functionality, integratibility, and comprehensiveness/redundancy.

Project researchers and ALS management would be able to input information regarding existing and proposed research projects through the IFA web site and could then retrieve information such as, for example, how projects interact based on the decision metrics. Completeness describes the progress of a project along its projected timeline. Functionality describes the usefulness of a project with respect to the overall goals and requirements of an ALS system. Integratibility gauges the potential for a project to interact with other NASA projects. Comprehensiveness/redundancy quantifies how well a project addresses the breadth of NASA's ALS systems goals, and the redundancy of having similar projects addressing any particular requirement or goal. Equipped with this information, NASA administration could evaluate the quality of current research and chart future research needs. Researchers preparing project proposals could determine what ALS needs their proposals should address. Although Rodriguez [20] presented a proof of concept (by determining which proper questions such a decision support tool should address), it clearly furthered efforts to develop a highly useful tool to assist in management decision for ALS research or other system analysis areas.

#### Process-Level Models

Process-level crop models have been used for analyzing the BPC subsystem. Such models (also referred to as explanatory or mechanistic models) quantify individual plant processes, and then integrate them to calculate daily growth and development (e.g. [21]). Plant processes may include vegetative and reproductive development, photosynthesis, transpiration, growth and maintenance respiration, leaf area growth, vegetative partitioning, primary and secondary reproductive growth, and so on. An approach within the ALS community has been to modify field versions of existing models for use in the analysis and optimization of ALS crop production, e.g. [7,8].

A specific example of the usefulness of detailed crop modeling efforts for ALS applications is *phasic control*, the specification of a series of different environmental conditions during a crop's life cycle, with the goal of optimizing some aspect of productivity [22]. A focus in the ALS community is on maximizing the edible yield averaged over the duration of the crop life cycle (e.g., maximizing grams of edible biomass m<sup>-2</sup> day<sup>-1</sup>). These units are useful for ALS applications where crops can be successively planted and harvested. Simulation studies using process-level models are thus ideal for focusing research efforts in such applications.

There is naturally a gradation from simple to complex under the scope of what may be considered process-level models for the purposes of this section. An example of a simple crop model that has found use within the ALS community is the *energy cas-cade model* [23]. This model, which was originally developed for wheat, calculates daily crop growth rates using the following trends: a linear increase in canopy light absorption from emergence through canopy closure, a constant (maximum) light absorption after canopy closure, a constant canopy photosynthetic efficiency through the onset of senescence, then decreasing linearly thereafter until crop maturity, and a constant respiratory efficiency. This basic model has been coupled with features of more detailed process-level crop models to produce simplified models for several ALS candidate crops [24]. These models still maintain some of the functionality of the more detailed models, but the simplification makes them easier to incorporate in analyses at the ALS system level.

It may be practicable for certain applications to reduce process-level crop models even further. For instance, available data and crop model predictions have been used to generate quasi-empirical relationships for simulating daily white potato, wheat, and soybean growth and development as dependent on the environment (i.e., light integral, aerial carbon dioxide concentration, and air temperature) [6]. These models were then integrated into a model-based predictive control logic, which could be used to identify an optimal set of environmental setpoints at each 24-hour period. The control logic was incorporated into a Visual Basic computer platform and it was demonstrated, via simulations, that such an approach would be useful for attempting to compensate for system disturbances on the planned production schedule in the BPC.

As an example of the power of this approach, Figure 4 illustrates the time history of the tuber development of a white potato crop. A simulated 30% reduction in the

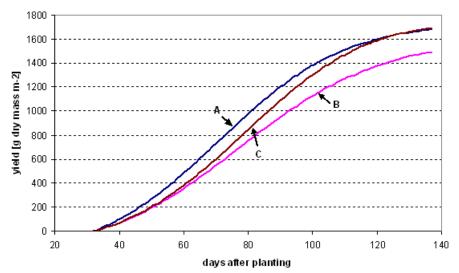


Figure 4. Systems analysis application based on model-based control of potato growth for the BPC. The three curves represent: (A) nominal tuber growth, (B) tuber growth following a disturbance in light intensity, and (C) tuber growth resulting from the control action.

desired light intensity value occurred during the 15<sup>th</sup> day of production and lasted for twenty days. Curve A shows the model-predicted growth curve without the disturbance. Curve B shows the predicted result when the disturbance is not corrected for, while curve C shows the results on the growth curve when the control logic is used to adjust air temperature, atmospheric carbon dioxide concentration, and light intensity following this disturbance. In this example, there was less than a 5% deviation between curves A and C at crop maturity. Thus, in addition to being used to simulate nominal and off-nominal conditions, such models are particularly useful for "what if?" simulation studies, such as for failure analyses. When included in systems studies, even simple crop models may identify issues that need to be addressed through more detailed modeling studies and/or specific experiments.

#### Top-Level Models

The top-level modeling approach generally includes multiple subsystems, their components, and interactions. The focus tends to be less detail oriented and more on the interface between processes. For ALS systems, various top-level models have been constructed. The general approach is to decompose the system into major subsystems including: (1) the biomass production component (BPC), (2) waste processing and resource recovery (WPRR), (3) food processing and nutrition (FPN), (4) the crew, and (5) the subsystem interfaces. Other top-level approaches may focus on defining performance characteristics of particular subsystems. For example, Goudarzi and Both [25] developed a method to predict crew performance during space missions and incorporated it into a technique to estimated costs associated with crew time.

A top-level model of an ALS system has been developed by Rodriguez [26], which focused on these subsystems. This model was supported by several subsystem modeling projects including an FPN model developed by Hsiang [27], and a BPC model developed by Fleisher et al. [12]. Models for the crew [28,29], WPRR [30], and the system interfaces were also developed for incorporation in the top-level model.

A back-end database of supporting information has also been developed for this model. This provides the flexibility to be able to simulate a wide array of ALS scenarios. This database was designed to store information regarding all the different components that may be included in an ALS system. The model was designed in such a way as to allow the user to select which parts are to be included in a simulation. For example, a crop database has been developed describing the candidate ALS crops, and the different types of growth chambers where these crops may be grown. This would allow for analyses of growing area and crop mix under given constraints, for instance. The model and database may be used in a similar fashion to vary aspects of the other subsystems as well. Users may select from an array of crew members, solid waste treatment systems, air revitalization systems, wastewater treatment systems, meals, and galley equipment in addition to the BP subsystem aspects when conducting their analyses. Given this flexibility, the back-end database has been termed an "ALS Warehouse," where users can go to select the parts for their ALS system simulation studies.

The development of this dynamic top-level ALS system model thus allows for analyses of design considerations, such as the sizing of system components and buf-

fers. It is also possible to ensure subsystem integration is as seamless as possible. Since the model captures the inherent dynamic issues of the system, components can be sized for peak rather than for average demands. By varying simulation scenarios it may also be possible to find new ways to lower peak demands, thereby reducing the overall size of the system.

For example, Figure 5 illustrates simulated results for daily crew time (left y-axis) and the production of edible biomass (right y-axis) in the BPC. Mission time is on the x-axis. In this scenario, BPC production areas for each crop are set at 25.63 m<sup>2</sup> for wheat, 42.51 m<sup>2</sup> for soybean, 5.73 m<sup>2</sup> each for white potato and sweet potato, and 2.87 m<sup>2</sup> for each salad crop (lettuce, spinach, and tomato). The manner in which the planting of each crop is scheduled will influence crew time requirements for cultural tasks. The simplest case is a non-staggered planting schedule, in which all crops are planted on the same day as in Figure 5. A large requirement for crew time occurs during the first few days of the simulation due to the fact that all the crops are planted at the same time. Requirements for crew time remain low after this point until the harvesting of the wheat crop on day 79. The harvest effects can be seen in the sudden drop in the edible dry mass within the plant growth system (the edible mass is moved from the BPC to the food processing components during harvest). A new wheat crop is planted on the following day. A tomato crop harvest can also be noted with the drop in edible dry mass on day 85. This, however, does not produce a discernible increase in crew time requirements, as the relative production area of the tomato crop is small compared to that of wheat. A soybean crop harvest can be noted on day 97 along with the peaks in crew time due to the larger size of the soybean crop.

In general, for the case of a non-staggered planting schedule, large peaks in crewtime coincide with harvest and planting dates, while there are relatively low requirements for crew labor between such times. This scenario may pose problems with man-

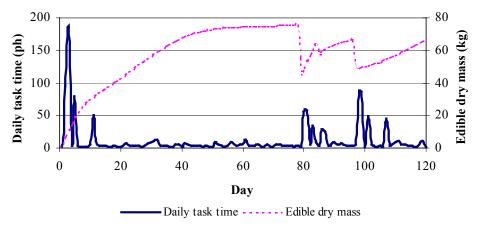


Figure 5. Effect of wheat cropping on nutrient and conditioning water storage simulated by the top level ALS model. Note that the model does not distinguish between developmental stages of the crops; thus, edible dry mass production is approximated throughout the production cycle.

aging the life support system because the sudden demand for crew time may affect the ability of the crew to finish tasks in other ALS components. Different scheduling approaches can, however, be simulated for the BPC using the top level ALS model so that such peaks can be smoothed out while ensuring that the required amount of edible biomass is still generated.

Similar efforts have been developed by other researchers. Drysdale et al. [31] evaluated different combinations of life support technology, such as re-supply, physicochemical regeneration, bioregeneration, and in situ resource utilization for a Mars mission, and determined that the efficiency of using regenerative life support technology increases along with mission duration. A complimentary study was conducted by Levri and Vaccari [32]. Their model was composed of five components: food and crew, crop, air, solid waste, and water. Mass flow, volume, crew time requirements, and power and cooling needs were simulated on a dynamic basis so that peak and offpeak demands could be estimated for missions of differing objectives, such as a 100-day Mars mission versus a 40-day lunar mission.

Wu and Garibay [33] developed an ALSS simulator to study energy, water, food, and air movement/utilization. Four basic components were simulated, including crew, crops, air recovery, and water recovery. Simulation results were primarily used to evaluate the performance of control strategies for managing energy and resource flow throughout the station. Their work indicates another use for systems studies and analysis in the design and testing of potential control algorithms for complicated life support technologies.

Wheeler [34] developed a simple metabolic model of food, carbon dioxide, and oxygen flow that included components for crew, biomass production, and waste recycling. The model was used to evaluate different strategies for treatment of inedible plant biomass with the objectives of maintaining system closure (with regards to mass) and optimizing carbon dioxide and oxygen balance in the life support system. The study concluded that appropriate waste treatment strategies varied depending on the degree of system closure desired.

#### 5.11.5 Conclusion

Advanced Life Support (ALS) systems consist of physical, chemical, and biological technologies and unit processes that, when properly integrated, will provide life support for long-term manned space exploration. Many research projects have been conducted over the past several decades in support of this goal. A substantial amount of experimental data has been generated as a result. In order to plan, design, assemble, and operate an ALS system, however, methodologies for extracting useful information from this data for systems analyses integration are of absolute necessity. The development of systems analysis tools and methods for processing information becomes an effective approach for this purpose.

Due to the multi-disciplinary nature of the research, however, much of the experimental results have yet to be organized into knowledge bases. Thus, the processing and organization of information becomes a critical step. Several documents have been developed in recent years by NASA's ALS community to support information proc-

essing and systems analysis techniques for the purpose of utilizing research information for systems design and integration. This interest in promoting systems efforts has also promoted systems related research for ALS applications. Examples include the development of an Internet platform to facilitate discussion and exchange of information between researchers, the construction of computerized tools that facilitate research project management so as to ensure that all vital areas of the ALS research program are addressed without significant overlap, and the development of decision support systems to support the design and operation of subsystems, such as the biomass production component. Mathematical modeling is still one of the prime foci of systems analysis work, particularly with respect to process-level models for the biomass production component and top-level models of the entire ALS system. Future needs in this area include continued development of crop models robust enough for estimating the direction and magnitude of changes in canopy gas-exchange, harvest index, and production scheduling due to off-nominal conditions that could occur in ALS systems.

As an engineering project evolves from concept to design and implementation, the complexity of the required knowledge and details increases. The use and processing of information will become more critical as long-term manned space exploration becomes more of a possibility. Future efforts will enhance the ability to integrate information and knowledge from various research efforts, perform systems analyses, finalize design and operation procedures, and evaluate system level interactions and performances. Systems studies and analyses will continue to be an integral part of the success of NASA's ALS program.

#### References

- 1. Averner, M. M. 1989. Controlled ecological life support systems. *Lunar Base Agriculture: Soils for Plant Growth*, eds. D. W. Ming, and D. L. Henninger. 173-191. Madison, WI: American Society of Agronomy.
- 2. Hanford, A. J. 2002. Advanced life support systems modeling and analysis project baseline values and assumption document. *JSC 39317, Crew and Thermal Systems Division*. Houston, TX: NASA-Lyndon B. Johnson Space Center.
- 3. Barta, D., J. M. Castillo, and R. E. Fortson. 1999. The biomass production system for the bioregenerative planetary life support systems test complex: Preliminary designs and considerations. *Proc. of the 29<sup>th</sup> International Conference on Environmental Systems*. SAE Technical Paper Series No. 1999-01-2188. Warrendale, PA: SAE International.
- 4. Kacira, M., and P. P. Ling. 2001. Design and development of an automated and non-contact sensing system for continuous monitoring of plant health and growth. *Trans. ASAE* 44: 989-996.
- 5. Hashimoto, Y. 1989. Recent strategies of optimal growth regulation by the speaking plant concept. *Act Hort*. 260: 115-121.
- 6. Fleisher, D. H. 2001. Crop modeling for multiple crop production and control for advanced life support systems. Ph.D diss. Camden, NJ: Rutgers University.

- 7. Cavazzoni, J., T. Volk, and G. Stutte. 1997. A modified CROPGRO model for simulating soybean growth in controlled environments. *Life Support and Biosphere Science* 4: 43-48.
- 8. Fleisher, D. H., J. Cavazzoni, G. A. Giacomelli, and K. C. Ting. 2003. Adaptation of SUBSTOR for hydroponic white potato production in controlled environments. *Trans. ASAE* 46: 531-538.
- 9. Cavazzoni, J., and P. Ling. 1997. Coupling machine vision and crop models for closed-loop plant production in advanced life support systems. *Life Support and Biosphere Science* 6: 279-285.
- Russo, D., and D. Henninger. 2002. Advanced life support project plan. Crew and Thermal Systems Division Lyndon B. Johnson Space Center CTSD-ADV-348 Revison C, JSC Publ. No. 39168. Houston, TX: Lyndon B. Johnson Space Center.
- 11. Ting, K. C. 2001. Concurrent science and engineering approach to decision support for controlled environment plant production. *Proc. of International Symposium on Design and Environmental Control of Tropical and Subtropical Greenhouses*, eds. S. Chen, T. Lin, C. Yeh, and R. Chen, 85-94. Taipei, Taiwan: National Taiwan University.
- 12. Fleisher, D. H., K. C. Ting, M. Hill, and G. Eghbali. 1999. Top level modeling of biomass production component of ALSS. *Proc. of the 29<sup>th</sup> International Conference on Environmental Systems*. SAE Technical Paper Series No. 1999-01-2041. Warrendale, PA: SAE International.
- 13. Ting, K. C. 1998. Systems analysis, integration, and economic feasibility. *Robotics for Bioproduction Systems*, eds. N. Kondo, and K. C. Ting, 287-320. St. Joseph, MI: ASAE.
- 14. Ting, K. C., and G. A. Giacomelli. 1992. Automation-culture-environment based systems analysis of transplant production. *Transplant Production Systems*, eds. K. Kurata, and T. Kozai, 83-102. The Netherlands: Kluwer Academic Publishers.
- 15. Ting, K. C. 1997. Automation and systems analysis. *Plant Production in Closed Ecosystems*, ed. E. Goto, 171-197. The Netherlands: Kluwer Academic Publishers.
- Hanford, A. J., and M. K. Ewert. 2001. Advanced life support systems integration, modeling, and analysis reference missions document. CTSD-ADV-383, Crew and Thermal Systems Division. Houston, TX: Lyndon B. Johnson Space Center.
- 17. Lange, K. E., and C. H. Lin. 1998. Requirements definition and design considerations. CTSD-ADV-245 (Revision A) Crew and Thermal Systems Division. Houston, TX: Lyndon B. Johnson Space Center.
- 18. Lange, K., K. Stafford, and W. C. Lee. 2002. ALS equipment database development. Presentation to the Systems Integration Modeling and Analysis Project Element.
- 19. Kang, S., K. C. Ting, and A. J. Both. 2001. Systems studies and modeling of advanced life support systems. *Agricultural and Biosystems Engineering* 2: 41-49.
- Rodriguez, L. F. 1999. The information flow analysis of the research activities within the New Jersey NASA Specialized Center of Research and Training. MS thesis. Camden, NJ: Rutgers, The State University of New Jersey.

- 21. Penning de Vries, F. W. T., D. M. Jansen, H. F. M. ten Berge, and A. Bakema. 1989. *Simulation of Ecophysiological Processes of Growth in Several Annual Crops*. Wageningen, The Netherlands: Pudoc.
- 22. Cavazzoni, J., T. Volk, B. Bugbee, and T. Dougher. 1999. Phasic temperature and photoperiod control for soybean using a modified CROPGRO model. *Life Support and Biosphere Science* 6: 273-278.
- 23. Volk, T., B. Bugbee, and R. Wheeler. 1995. An approach to crop modeling with the energy cascade, *Life Support and Biosphere Science* 1: 119-127.
- 24. Jones, H. W., and J. Cavazzoni. 2000. Top-level crop models for advanced life support analysis. *Proc. of the 30<sup>th</sup> International Conference on Environmental Systems*. SAE Technical Paper Series No. 2000-01-2261. Warrendale, PA: SAE International.
- Goudarzi, S., and A. J. Both. 2004. Accounting for performance decrements in crew time calculations for space missions. *Proc. of the 34<sup>th</sup> International Conference on Environmental Systems*. SAE Technical Paper Series No. 2004-01-2362. Warrendale, PA: SAE International.
- 26. Rodriguez, L. F. 2002. A dynamic object-oriented advanced life support system model. PhD diss. Camden, NJ: Rutgers, The State University of New Jersey.
- 27. Hsiang, H. H. 2002. Top-level modeling of a food processing and nutrition (FPN) component of an advanced life support system (ALSS). MS thesis. Camden, NJ: Rutgers, the State University of New Jersey.
- 28. Goudarzi, S., and K. C. Ting. 1999. Top level modeling of crew component of ALSS. 29<sup>th</sup> International Conference on Environmental Systems. SAE Technical Paper Series No. 1999-01-2042. Warrendale, PA: SAE International.
- 29. Goudarzi, S., J. Cavazzoni, and A. J. Both. 2002. Dynamic modeling of crew performance for long duration space missions. 32<sup>nd</sup> International Conference on Environmental Systems. SAE Technical Paper Series No. 02ICES-196. Warrendale, PA: SAE International.
- Rodriguez, L. F., S. Kang, J. A. Hogan, and K. C. Ting. 1999. Top-level modeling of waste processing and resource recovery component of an ALSS. 29<sup>th</sup> International Conference on Environmental Systems. SAE Technical Paper Series No. 1999-01-2044. Warrendale, PA: SAE International.
- 31. Drysdale, A. E., M. K. Ewert, and A. J. Hanford. 2003. Life support approaches for Mars missions. *Adv. in Space Res.* 31: 51-61.
- 32. Levri, J. A., and D. A. Vaccari. 2004. Model implementation for dynamic computation of system cost for advanced life support. *Adv.Space Res.* 34: 1539-1545.
- 33. Wu, A. S., and I. I. Garibay. 2004. Intelligent automated control of life support systems using proportional representations. *IEEE Trans. Sys. Man and Cybern. Part B.* 34: 1423-1434.
- 34. Wheeler, R. M. 2003. Carbon balance in bioregenerative life supports systems: Some effects of system closure, waste management, and crop harvest index. *Adv. in Space Res.* 31: 169-175.

# 6

# Management and Decision Support Systems

# 6.1 Farm and Crop Management Systems

K. O. Sindir

Abstract. Knowledge is becoming the new jewelry of life and there is an ongoing change towards knowledge-based societies and economies in which research, innovation, and knowledge are the major components. Knowledge-based agricultural production, through the adoption of information and communication technologies, is becoming necessary in order to respond to numerous challenges of consumer demands on food safety and security, sustainable development, environmental protection, sustainability of farmers' revenues, and a competitive advantage and power in both domestic and global markets. Farm and crop management integrated with information systems will allow farmers not only to maximize profits or minimize costs, but also to deal with issues surrounding quality and the value of production.

**Keywords.** Farm management, Crop management, Information systems, Decision support systems.

#### 6.1.1 Introduction

The world is currently encountering a revolution in information and communication technology, which could bring significant changes to human life. Knowledge is becoming the new jewelry of life and there is an ongoing change towards knowledge-based societies and economies in which research, innovation, and knowledge are the major components. Knowledge-based agricultural production, through the adoption of information and communication technologies, is becoming necessary in order to respond to numerous challenges of consumer demands on food safety and security, sustainable development, environmental protection, maintenance of farmers' revenues, and competitive power both in domestic and global markets. Some of the developments in attaining a knowledge-intensive agriculture include use of models on PCs or on the Internet, electronic data interchange which facilitates data exchanges between farmers' management systems and their partners, multimedia tools (such as the Internet and CD-ROMs), and e-commerce [1].

This chapter provides a brief description of information systems in agriculture and their roles in farm and crop management.

#### **6.1.2 Definition of Farm Management**

The major task of farm managers is to allocate resources, mainly in order to maximize the farm's profit or minimize the costs of doing work. However, problems of environmental pollution, sustainability, globalization of trade, increase in consumer demands on food safety, and recent developments on information and communication technologies have all made management of farms a highly complex issue.

Challenges and opportunities on and off the farm generate a changing agenda for farm business management and farm families in the rural sector. Conceptualizing farm businesses as a mix of *management*, *resources*, and *family* strengthens appreciation of new structures and strategies, ties in with "triple bottom line" thinking, and reflects the shift from farm policy to an array of policies focusing on social, environmental, and economic aspects of contemporary rural life [2].

Competitiveness on both domestic and global markets requires farmers to manage their productive resources more efficiently, and to become more effective business managers.

On the one hand, increasing consumer demands for safer and more secure food, and public and regulative pressure to protect the environment, has brought new concerns in farm management. Access to timely, proper, and correct information on activities and the inputs and outputs of farm businesses also must be of utmost importance, not only for farmers, but also for agro-traders, agro-industrialists, and consumers. Growers are required to keep up-to-date records to demonstrate that all production activities comply with food safety regulations and to help trace the history of products from farm to final consumer. These have all become increasingly complex issues for farms in a rapidly changing business environment which includes their competitors, customers, wholesalers, and suppliers.

On the other hand, policies, regulations, rules, and any other legislative arrangements also affect their decisions on farm and crop management.

Since farms are not the same size and type, there is no typical model for management. A farmer essentially has five main basic management functions to achieve the goals and objectives of his farming business [3]:

- Planning—Setting daily priorities and schedules, recognizing problem areas and looking for alternative solutions, making financial plans, looking for alternative cropping plans, establishing the overall enterprises for the business, and developing the farming business;
- Organizing—Establishing an internal structure of the roles and activities required to meet the farm's goals, coordination of efforts among people, deciding who reports to whom, determining the functions and authorities of positions, establishing the work routines and standard operating procedures (SOPs);
- 3. *Managing human resources (staffing)*—Recruiting and/or hiring workers, training and evaluating workers;
- 4. *Directing*—Delegation of authority, assigning responsibilities to workers, motivating the workers, establishing good communication with workers, dealing with workers' concerns;

5. Controlling—Keeping track of expenses and income, monitoring the records and accounts of operations within the business, comparing rates of production and levels of performance or productivity, monitoring production processes and making changes as necessary, keeping track of work routines.

Directing and controlling anything implies that there are options open and that decisions to be taken. Organizing and operating a farm business requires a manager to make and carry out numerous decisions of various kinds. Making decisions is not difficult if all options are well-identified and known. That means correct, meaningful, informed, realistic, realizable decisions must always be based on information. The major function of a farm manager is to process this information in terms of translation, coordination, recording, digestion, correlation, and reproduction. However, information has no merit in itself; that is, it may be of interest but it serves a purpose only if it is used. Its main use is in the decision-making process.

#### 6.1.3 Data, Information, and Knowledge

It is usually difficult for many to differentiate between the terms "data," "information," and "knowledge." According to Russell Ackoff, a systems theorist and professor of organizational change, the content of the human mind can be classified into five categories [4]:

- · data, or symbols;
- information, which is data that are processed to be useful; provides answers to "who," "what," "where," and "when" questions;
- knowledge, which is the application of data and information; answers "how" questions;
- understanding, which is the appreciation of "why;" and
- wisdom, or evaluated understanding.

Ackoff indicates that data is raw; it simply exists and has no significance beyond its existence (in and of itself). Data can exist in any form, usable or not, and does not have meaning of itself, whereas information is data that has been given meaning by way of relational connection. This meaning can be useful, but does not have to be. Knowledge, on the other hand, is an appropriate collection of information, such that its intent is to be useful and it is a deterministic process.

Managers usually need information of the following types [5]:

- summaries from the routine processes of the enterprise (operational staff need continuous and detailed information, whereas managers need only a summary of operational data);
- information on exceptional events;
- facility to find ad hoc information;
- time series information, concerned with the past and the future;
- comparative external information, focusing on information not only from internal sources of the enterprise but also external sources; and
- contextual or environmental information.

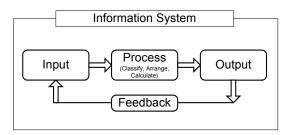


Figure 1. Structure of an information system.

#### 6.1.4 What Is an Information System?

Laudon and Laudon [6] define the information system technically as a set of interrelated components that collect (or retrieve), process, store, and distribute information to support decision making and control in an organization. In addition to supporting decision making, coordination, and control, information systems may also help managers and workers analyze problems, visualize complex subjects, and create new products.

The components of an information system are input, processing, and output. Input activity captures or collects the data as a source of raw information within the farm or its environment. Processing is the activity of giving meaning to this raw data, i.e., converting it into information which is useful and meaningful to the farmer. Output, as the final stage of the information system, transfers the processed and meaningful information to the people who require it or to the appropriate activities. Output returned to appropriate people within the farming business for evaluation and correction is called *feedback*.

Information systems are part of an organization, in this context a farming enterprise. Organizations usually have a structured hierarchy and formal standard operating procedures (SOPs), i.e., formal rules for accomplishing tasks that have been developed to cope with expected situations.

Powerful computers, software, and networks, including the Internet, have helped organizations become more flexible, eliminate layers of management, separate work from location, and restructure work flows, giving new powers to both line workers and management.

Information systems in a farm enterprise can be classified mainly into four major types:

- 1. *Transaction processing systems*—Sorting, listing, merging, and updating all kinds of transactions, invoices, records of daily farming activities and events. Transaction processing systems perform and records the daily routine transactions necessary to conduct the farming business.
- 2. Farm automation systems—Document management, scheduling, communication, including computer systems, such as word processing, electronic mail systems, and scheduling systems, which are designed to increase the productivity of farmers in the office. All farm automation systems could help farmers keep in

- contact with their suppliers, information providers, and agricultural research institutes, and follow up with recent developments in their field of business.
- 3. Management information systems (MIS)—Information systems at the management level that serve the functions of planning, controlling, and decision making by providing routine summary and exception reports. They are systems designed to provide daily production information at the individual animal level or crop production level that is of potential value in making management decisions [7].
- 4. Decision support systems (DSS)—Information systems that combine data and sophisticated analytical models or data analysis tools to support semi-structured and unstructured decision making for farm managers for a wide range of applications including business and organizational management, health, water resources, environment, crop/animal production, transportation systems, etc., e.g., decisions on feed mixture, when to apply pesticide for various crops, how to realize a competitive advantage.

Numerous MIS and DSS have been developed to help farmers improve their managerial decision making. The value of these systems is affected by goals and competence of the farmer(s), and characteristics of the farm, such as the size of the farm. Several are discussed below.

#### 6.1.5 Farm Management Information Systems (FMIS)

Information and decision making are inseparable [8]. A system for providing information is vital to a business decision-making process [9]. Farm decision-makers use information from a wide range of sources, but one of the most valuable sources of specialized information about the farm operation is provided by a *farm record system* (FRS) [10,11]. An FRS can include financial and production record types [12]. The system may be as simple as a cash book, or so large and complex that it requires the processing capabilities of a computer to maintain it efficiently. Information provided by an FRS can be passed on to individuals or organizations outside the farm business, such as accountants to prepare tax returns and bankers to support loan applications, or used within the farm gate to support the business decision-making process [13,14].

A farm management information system (FMIS) is conceptualized as an FRS that provides information to support farm business decision making [14]. If an FRS only provides information to individuals or organizations outside the farm business, then it is not a FMIS. Record-keeping methods are defined as those bookkeeping or accounting skills that are needed to set up and operate a variety of financial and production accounts or report types common to farm businesses.

A computer is conceptualized primarily as a management tool. It is a system of hardware and software elements capable of supporting a farm record system and performing analysis on the data. The adoption and use of a computer enables farm businesses to operate with a larger number of record types and perform more extensive and complex analyses than would be possible using manual procedures alone. The adoption of record keeping methods (i.e., establishment of a farm record system) is seen as the first level of sophistication and the first step toward establishing a FMIS. The use of information provided by a farm record system is seen as the second level of

sophistication and defines a manual FMIS. The adoption and use of a computer in support of a FMIS adds another layer of sophistication. The increase in sophistication of the FMIS is a contiguous process. Each step upwards in the level of sophistication is the next stage in the evolution of FMIS. Using the concept of levels of sophistication in FMIS, it is possible to segment farm businesses into four groups [14]:

- 1. No farm record system—All farm businesses have some compulsory reporting requirements, such as taxation. A farm record system is defined as maintenance of at least a basic cash book. Farm businesses in this category do not operate a farm record system, but are presumed to keep at least a box of check stubs and receipts to pass on to their accountants at the end of the financial year.
- Farm record system (FRS)—Farm businesses in this segment do operate a manual farm record system, but information provided by it is not used to support decision making.
- 3. Manual FMIS—In this segment information from a farm record system is used in farm business decision making. The farm record system is maintained and data analysis is performed using predominantly manual methods, but may be assisted by some type of electronic calculator or other items of basic office equipment
- 4. *Computer-supported FMIS*—It is not presumed that the farm record system will be maintained solely on computer. Most farm businesses in this segment will use a combination of computer and manual methods.

As mentioned before, information systems, comprising hardware and software, involve the use of digital information and control systems which collect (or retrieve), process, store, and distribute information to support decision making and control, to reduce risk and uncertainty, and to assist with the implementation and evaluation of improved business management strategies. As Blackmore [15] defines, information technology improves the knowledge base and increases the ability to control the production system in order to reduce risk and uncertainty, better identify the extent of variation in the system, and thereby formulate strategies that explicitly address this variation and support the processes to implement these strategies. In this respect, information technology is essentially concerned with risk management.

From a management point of view, [15] classifies information technology into two broad categories: (1) information management and decision support, including methods for the collection, assembly, logging and retrieval of data, and methods for data analysis, interpretation and decision making; and (2) implementation, i.e., monitoring and control systems.

It has become necessary for farm managers to implement information systems and the relevant technology on their farms in order to survive and be a partner of their sector, and to carry on with their profitable cost-effective business.

Relative to other business activities, the manager of a farm contends not only with the risks of the marketplace, but also with the uncertainties of the production system, given its reliance on natural and biological processes. There are a large number of external, uncontrollable factors that influence the technical and economic performance of the farm system, such as prevailing weather conditions or the incidence of pests and

diseases. Decisions on cropping practices are required "up front" without full knowledge of future events and resultant outcomes. Furthermore, the response of the system to actions taken cannot be predicted for sure, and decisions made at one point in the production cycle have considerable, albeit uncertain, influences on later outcomes and decisions [15].

#### 6.1.6 Decision Support Systems (DSS)

There are many definitions of DSS, but they usually fall into one of two categories: narrow or broad [16]. The narrow definition takes the view that DSS are interactive computer programs that use analytical methods and models to help decision-makers formulate alternatives for large unstructured problems, analyze their impacts, and then select appropriate solutions for implementation [17]. The DSS will essentially solve or give options for solving a given problem. The decision process is structured in a hierarchical manner, the user inputs various parameters, and the DSS essentially evaluates the relative impact of doing x instead of y.

The broader definition incorporates the narrow definition above but also includes other technologies that support decision making, such as knowledge or information discovery systems, database systems, and geographic information systems (GIS) [18]. With a broader perspective DSS: (a) are computer-based, (b) are an information and/or technology transfer agent, (c) contribute to option selection, and (d) aid decision making irrespective of whether the solution is generated by the system itself or independently deduced by the user from the information provided.

Decision support systems, which process information, estimate the efficiency of possible decisions and reduce uncertainty in management, can help farmers in evaluation and selection of the best possible decisions. DSS provide the farmer or consultant a possibility to choose proper variants of farm management by evaluating dynamically changing weather conditions, new technologies, changes in fields, and general agricultural policy.

Table 1 provides some examples of decision support systems in use. There are numerous other DSS models in use. A list of models with their descriptions and links to their sources can be found at <a href="http://www.wiz.uni-kassel.de/model\_db/models.html">http://www.wiz.uni-kassel.de/model\_db/models.html</a>. However, many of those IT-based decision support systems have difficulty in reaching high user numbers. One reason may be that the developers lack a user perspective, and also that majority of farmers make decisions intuitively [19,20].

Farmers' decision making is mostly viewed as a series of linear steps. Johnson et al. [21], for instance, identify six steps of decision making: (1) problem definition, (2) observation, (3) decision, (4) analysis, (5) action, and (6) bearing responsibility. Öhlmer and Öhlmer et al. [19,20] identify four separate functions of decision making: (1) problem detection, resulting in detection of a problem or not; (2) problem definition, resulting in choice of options for further development; (3) analysis and choice, resulting in choice of one or more options; and (4) implementation, resulting in output consequences and responsibility bearing. Each of the four functions has four subprocesses: (a) searching information and paying attention to relevant information; (b)

Table 1. Some types of agricultural DSS currently in use or under development (compiled from [18]).

DSS	Target Audience and Environment	DSS Category	Description
CABI Compendium	Scientists, researchers, and extension staff; worldwide	Database- and knowledge-driven	Pest management for various crops
CROPWAT	Scientists and irrigation engineers; worldwide	Model- and database-driven	Irrigation management
DSSAT (Decision Support System for Agrotechnology Transfer)	Scientists and researchers; worldwide	Database-driven, model-driven (CERES, CROPGRO, and CROPSIM systems)	For various crops, analyzes and displays outcomes of simulated agronomic experiments
HOWWET	Researchers, extension agents, and growers; Australia	Database-driven	Plant-water interactions
MAS (Multi-Agent System for Natural Resource Management)	Scientists, researchers, and extension workers; worldwide	Model-driven	Uses CORMAS programming software for simulations of various natural resource management outcomes
NuMass (Nutrient Management Support System)	Scientists and researchers; worldwide	Model- and database- driven	Soil nutrient management for various crops including rice, maize, cassava, etc.
WEPP (Water Erosion Prediction Project)	Scientists, researchers, and extension workers; worldwide	Model-driven	WEPP is a process- based, distributed parameter, continuous simulation, water erosion prediction model for use on a PC

planning (which was included only in the analysis and choice phase) and forecasting consequences of the new information; (c) evaluating consequences and choosing alternatives; and (d) bearing responsibility of the choice. A farmer uses more or less intuition and analysis in his decision making. Regardless of whether he uses intuition or analysis, he goes through all four of the functions listed above. The difference is in how he performs these functions.

Simulation-based models are one of the most commonly used among the systems of supporting decision making of farmers. Simulation can be defined as a process of building, analyzing, and using theoretical and experimental results in order to summarize a body of knowledge, to make predictions, or to understand system dynamics. The stage of building consists of defining the problem and the system, collecting knowledge (processes, parameters, and direct observations), developing a model concept, translating it into a mathematical model, and converting this into a computer program [22]. It is essential that there are two sources of information about the system at hand:

direct measurements (observations) and indirect statements on the system, produced by the model.

Simulation models have been developed in almost all aspects of agricultural production. However, a model is useful only if it accurately represents some of the system behavior, and useless if it does not. Many simulation models in agriculture, economics, environmental science, and ecology are based on a body of knowledge which consists of theories that are not generally accepted, debatable or controversial hypotheses, questionable simplifications, and a bundle of implicit or ambiguous assumptions. The problems with insufficient and imprecise system knowledge cause model uncertainty, which can be traced back to model structure, uncertain inputs, and uncertain parameters [23].

## 6.1.7 Record Keeping and Traceability of Farm Products

The recent concerns about the risk of food safety and food contamination have also increased the necessity of investing in information systems in order to keep records of the foodstuff's past and the methods applied during the production process. Traceability, in this respect, can be defined as the ability to trace the history, application, or location of any entity by means of recorded identifications. It the possibility to ensure, at any stage of the food chain, that the path of a food and the relevant information about it are known. The types of information required for the tracing of the product involve:

- 1. *Product identification*—A unique means, as simple as possible, to identify a food or batch thereof;
- 2. Product information—The raw materials used, how it was produced and changed (if appropriate), where and when it came from, and where and when it was sent (one step backward and one step forward); and the controls to which the product has been subject (if relevant); and
- 3. Linkages—Between product identification and product information.

This information is generated for the purpose of food safety and/or of fair practices in food trade, and may be used, as appropriate, by industry, government and other third parties. It is recorded by each business involved and stored, within an appropriate time frame, in a way making fast and easy retrieval possible. Information systems have an important role in collecting, retrieving, storing, and distributing the product information and identification.

## 6.1.8 Information Technology in Agricultural Mechanization

Until the last decade, depending on the power source and the stage in the mechanization process, there have been mainly three levels of technology considered for use in agriculture: hand-tool technology, animal-draught technology, and mechanical-power technology. Voss [24] elucidated the following four stages in the mechanization process:

- 1. Hand tools are used with a very little capital investment;
- 2. Human labor is supplemented by animal power for primary and secondary tillage and for pumping water;
- 3. Mechanical power is introduced to some but not all operations; and

4. All operations in the crop production are completely mechanized with power equipment (no electronic components are included).

However, none of the above-mentioned technologies and the stages of the development of agricultural mechanization could have dealt with in-field variability of crop and soil conditions. Today, technology has reached a level that allows a farmer to measure, analyze, and deal with in-field variability that previously was known to exist but wasn't manageable. The ability to cope with variations in productivity within a field and maximize yields has always been a desire of the farmer, especially the farmer with limited land resources. The recent development of microprocessors and other electronic technologies are new tools available to help all farmers reach this goal. This new approach of farming is called *precision farming* (or *precision agriculture*) and the technology behind it is called *variable rate application technology*.

In accordance with these recent developments in information and communication technologies (ICT), the following two stages of mechanization can be added to those listed above:

- 5. Intermediate level of information systems (IS) and information technology (IT) use in agricultural production. At this level, a farm owns a personal computer and software capable of supporting stock keeping, historical records, and analyzing what-if models. A monitoring system is installed on a farm tractor to display the forward speed, PTO speed, distance traveled, fuel consumption, and work rate. In addition, sprayers may have a control and data logging facility. No spatial information is required or used.
- 6. Precision farming as an advanced level of ICT application in agriculture. This level comprises the IT components of the fifth level but with enhanced capabilities, providing full spatial understanding and treatment of agricultural operations, e.g., soil and yield mapping, tractors equipped with agricultural bus systems, GPS-based instrumentation systems, and variable rate application technologies. The major concerns of this level of technology use are increases in yield, reduction of inputs and their costs, and the environment.

Farms, especially in developing countries, are generally small in scale, which causes low productivity and inefficient use of production inputs. The precision farming approach, with its potential for reducing environmental impacts and increasing productivity, is being considered a new revolution in agriculture. Precision farming takes the variability of the field into account in terms of soil fertility, yield, pest distribution, soil compaction, etc., and allows the application of the right amount of the right inputs at the right location. However, in order to benefit from this advanced technology to save inputs, energy and environmental impacts, some extra investments should be made which require an appraisal study for various farming conditions.

However, it is also important to know that information technologies, in terms of hardware and software and information systems, are in a very dynamic state, and therefore the information provided here is a picture of the past and current time, not necessarily the future.

#### 6.1.9 Conclusion

Knowledge-based agricultural production, through the adoption of information and communication technologies, is becoming necessary in order to respond to numerous challenges of consumer demands on food safety, security and traceability, productivity, sustainable development, environmental protection, maintenance of farmers' revenues and competitive power both in domestic and global markets.

According to Gelb [25], adoption of information and communication technologies as a unique challenge in rural areas has long been a specific public concern with regional, national, and international strategic significance. However, the rate, efficiency, and scope of ICT adoption are slower than expected. This could be expressed in reduced farmer profitability, an under-utilized contribution by extension, and reduced consumer welfare. Relevance, various human-element issues, an "immature" infrastructure, and proficiency in information systems seem to be common major adoption constraints, besides the individual impediments and unrecognized benefits and cost of ICT

## Acknowledgements

There is considerable cited information provided within this text which is compiled from books or papers of various scientists dealing with farm management and/or information technology in agriculture. I express my acknowledgments to all those cited in the text and listed within the references without whom this article could not be written.

## References

- 1. Waksman, G. 2001. ICT in farm management: Impact of the internet. *Proc. of Third Conference of the European Federation for Information Technology in Agriculture, Food and the Environment,* 567-572.
- 2. Ronan, G. 2002. Delving and divining for Australian farm management agenda: 1970-2010. *Agribusiness Review* 10: Paper 7, ISSN 1442-6951.
- 3. USDA. 1989. Farm management—How to achieve your farm business goals. *Yearbook of Agriculture—1989*. Washington, D.C.: United States Department of Agriculture.
- 4. Ackoff, R. L. 1989. From data to wisdom. *J. Applied Systems Analysis* (16): 3-9. (from the paper by Bellinger, G., D.Castro, A.Mills, Data, Information, Knowledge, and Wisdom, http://www.systems-thinking.org/dikw/dikw.htm)
- 5. Vernon, R. 2001. Knowing where you are going: Information Systems for Agricultural Research Management. The Hague, Netherlands: International Service for National Agricultural Research.
- 6. Laudon, K. C., and J. P. Laudon. 2000. *Management Information Systems: Organization and Technology in the Networked Enterprise*, 6<sup>th</sup> ed. Upper Saddle River, NJ: Prentice Hall International, Inc.

- 7. Verstegen, J. A., A. M., and R. B. M. Huirne. 2001. The impact of farm management on value of management information systems. *Computers and Electronics in Agriculture* 30: 51-69.
- 8. Kallman, E. A., and L. Reinharth. 1984. *Information Systems for Planning and Decision Making*. New York, NY: Van Nostrand Reinhold.
- 9. Kast, F. E., and J. E. Rosenweig. 1981. *Organization and Management: A Systems and Contingency Approach*, 3<sup>rd</sup> ed. New York. NY: McGraw-Hill.
- Osburn, D. D., and K. C. Schneeberger. 1983. Modern Agricultural Management: A Systems Approach to Farming, 11-12. Reston, VA: Reston Publishing Co.
- 11. Schnitkey, G. D., M. T. Batte, and E. Jones. 1991. Farm accounting systems and information usage. *Rev. Agric. Econ.* 13: 103-117.
- 12. Sonka, S. T. 1983. *Computers in Farming: Selection and Use,* 16-22. New York, NY: McGraw-Hill.
- 13. Scudamore, P. H. 1985. Developing on-farm computers. *Proc. Farm Electronics and Computing: An Int. Symp. Royal Agricultural Society of England*, 23-35.
- 14. Lewis, T. 1998. Evolution of farm management information systems. *Computers and Electronics in Agriculture* 19: 233-248.
- 15. Blackmore, B. S. 2000. Using information technology to improve crop management. Weather and agro-environmental management. *Proc. of the papers presented at the AgMet Millennium Conference*, 30-38.
- 16. Bakker-Dhaliwal, R., M. A. Bell, P. Marcotte, and S.Morin. 2001. *Decision Support Systems (DSS): Information Technology in a Changing World*, IRRN 26.2, 5-12. Los Bafios, Laguna, Phillipines: IRRI.
- 17. Watkins, D. W., and D. C. McKinney. 1995. Recent developments associated with decision support systems in water resources. *Rev. Geophysics* 33: (Suppl.).
- 18. Power, D. J. 1997. What is a DSS? *DSSTAR*, The online Exec. Journal for Data-Intensive Decision Support October 21, 1(3).
- 19. Öhlmer, B. 2001. Analytic and intuitive decision making—Swedish farmers' behaviour in strategic problem solving. *Proc. of Third Conference of the European Federation for Information Technology in Agriculture, Food and the Environment*, 609-614.
- Öhlmer, B., J. Flodin, D. Hagström, T. Karlsson, B. Noren, and E. Seremba. 2001. User perspective on decision support systems for farms—The case of Agriwise. Proc. of Third Conference of the European Federation for Information Technology in Agriculture, Food and the Environment, 543-548.
- 21. Johnson, G. L., A. M. Halter, H. R. Jensen, and D. Thomas. 1961. *A Study of Managerial Processes of Midwestern Farmers*. Ames, IA: Iowa State Press.
- 22. Scholten, H., and M. W. M. van der Tol. 1998. Quantitative validation of deterministic models: When is a model acceptable? *Proc. of the Summer Computer Simulation Conference, SCS*, 404-409, (ISBN 1-56555-149-4).
- 23. O'Neill, R. V., and R. H. Gardner. 1979. Sources of uncertainty in ecological models. *Quantitative Validation of Deterministic Models: When is a Model*

- Acceptable? Proc. of the Summer Computer Simulation Conference, SCS. H. Scholten and M. W. M. van der Tol, 1998, 404-409.
- 24. Voss, C. 1975. Different forms and levels of farm mechanization and their effect on production and employment. *Meeting of the FAO/OECD Expert Panel on the Effects of Farm Mechanization on Production and Employment*, 4-7.
- 25. Gelb, E. 2005. ICT Adoption in agriculture and rural development. *Theme Report*, 5<sup>th</sup> EFITA/WCCA Conference.

# 6.2 IT for the Design of Animal Husbandry Buildings

J. Gartung, W. Hartmann, F. Preiß, K. Uminski, and E. Witzel

Abstract. Due to the continually increasing number of design drafts required for building permits, construction planning has become much more complicated. In contrast, the time available for planning is being constantly reduced. For this reason, the use of information technology in construction and cost calculation has been a focus for a long time. Routine tasks can be easily completed and information rapidly exchanged with networks of partners and databases.

Methods developed in Germany in the area of cost calculation for agricultural buildings are described. They primarily address documentation, preparation, and processing of building cost data according to a construction cost network system. Linked methods and classification systems allow construction costs to be managed from the detailed level of unit prices for construction services, to the building elements, through to the generally classified steps of cost blocks. Construction cost data for selected farm buildings has been documented and is available via the Internet and CD-ROM, including a database that includes 107 barn models.

**Keywords.** Construction costs, Building software, Computer planning, CAD systems, Construction cost network system, Construction cost database BAUKOST, Cost blocks, Barn, Slurry, Auxiliary facilities, Animal housing, Dairy cattle, Barn.

## 6.2.1 Construction Planning

Although the buildings and permanent structures needed by farms are special structures, they are built by planners and construction companies in Germany based on the same principles, laws, regulations, codes, and guidelines as other buildings. Building planning has become increasingly complex through the constantly increasing number of requirements to have a plan approved, but the planning time should be constantly shortened. That is why efforts have concentrated on using information technology to help in planning.

As in all areas of daily life, the electronic media have proven themselves a help. Routine tasks, in particular, can be easily and quickly fulfilled with the support of information technology. Through networks with partners and databases around the world, information of almost every type can be exchanged in a very short time.

From the initial building idea through to the completion of the constructed object, many considerations must be made. Here, qualified professionals from very diverse areas are needed in order to appropriately meet the requirements for animal and environmental protection, work safety, economic viability, form and construction. In particular, site selection and consideration of many aspects for the appropriate positioning of the building on the farm have special significance for the new construction of buildings for animal husbandry.

All business negotiations in the planning and building process take place at the initiative of the owner. The owner determines the size and characteristics of what is constructed and sets the basic framework conditions. His wishes are then transformed into plans for buildings by architects, professional planners, and other specialists. The realization of the plans is carried out by a general construction firm or multiple construction firms in different fields. Therefore, many parties and tasks must be coordinated, organized, and controlled in the planning stage.

First, the planning of necessary features takes place. Generally, this responsibility rests with the owner. For this purpose he obtains individual information from professional via the Internet as well as from public or private consultants. This early phase of all building planning processes is particularly important because here all developments for all later results of the construction planning will be initiated. For this purpose the farm manager and the consultants have many IT-supported systems and information available [1].

In Germany, as a rule, an architect is assigned to plan the object as soon as the owner develops clear ideas. On the basis of the Regulation of the Fees for Services of Architects and Engineers (HOAI), a contract is signed with the architect. The basis for the extent of architectural services is the service portrait of the object planning. It includes the services of the contractors for new constructions. The basic services are subdivided into nine service phases: basic calculations, preplanning, draft planning, licensing planning, implementation planning, preparation for the assignments, cooperation in the assignments, object monitoring, and object supervision.

The classification described basically holds true for the planning, construction, and documentation of all types of buildings and permanent facilities. Farm buildings are no exception. But the production processes—as in industrial construction—determine the character and basic concept of the farm building. A narrow dovetailing of the previously mentioned need-planning is an absolute goal, because architects and engineers must take the requirements of the animals, the products, the production and working processing, the work environment, environmental protection, etc., under consideration in all planning steps.

Depending on the planning tasks, very different IT programs and systems are implemented as work aids. In addition to professional software such as CAD and AVA, and considerations such as static, heat, and sound insulation, special measures are also taken such measuring barn climate or emissions from animal husbandry.

#### 6.2.2 Construction

According to the traditional picture of the German construction industry, the owner has the idea and makes the assignment, which the planners transform into drawings

and detailed work plans. The construction company carries out the tasks listed in the plans. The division between planning and construction is still maintained at this time. Building teams, which make early connections between planners and builders, can lead to short construction times and low-cost solutions.

In Germany, the current task division between planners and realization has changed to some extent. Instead of architects, who as object planners work as equal partners with the professional planners (statisticians, electricians, ventilation planners, etc.), the form of general planning is becoming much more prevalent. One speaks of a *general planner* if the contract assignor assigns all or at least most of the planning, coordinating, organization, supervision, consultancy, and monitoring to a single contractor.

To carry out the actual construction, at this time the division between craftsmen (masons, carpenters, roofers, electricians, etc.) still exists. But other forms of cooperation are increasingly being created. Instead of single contractors, general contractors are emerging. The total creation of a building is called "turnkey completion." In this type of construction, conventional building processes are replaced with more rational building processes. The use of pre-fabricated large format parts is common. Here it can be advantageous if the detailed planning is done by the general contractor and not by the architect.

In the construction of barn buildings and multi-purpose buildings, the percentage of "turnkey facilities" is already at about 40%. In this case, the general contractors are the construction companies, which, in addition to providing the self-built building shells, also deliver and install the technical equipment with the help of subcontractors. Another possibility is the use of equipment providers who subcontract the building construction.

From the first planning considerations through to the beginning of operations, a broad range of information is shared among the participants. The goal must be to use previously gained data and make it available for the use phase, that is, for the entire life-cycle of the building through to demolition.

Today networks like *Business to Business (B2B)* are available to farms to exchange data and information. The manufacture and processing of construction parts in the finishing and production plants is automated in many companies. Software for construction site logistics supports on-site work.

As an example the use of information technology in construction and cost calculations is described as it is used in the Institute of Production Engineering and Building Research at the German Federal Agricultural Research Center (FAL) in the following sections.

#### 6.2.3 Construction and Cost Calculation with the Help of Construction Software

Building cost data for the new construction of selected farm buildings has been studied at the Institute since the end of the 1970s. Since 1982, the cost orientation values have been published in the classification steps of cost blocks [2] by the Association for Technology and Structures in Agriculture (KTBL) in KTBL Handbook [3]. In addition the data flows into the price guides of ALB Hessen. The most important Internet statements on the implemented cost block method, the procedures for process-

ing the most important types of barns, and selected cost values can be found at the address http://www.bb.fal.de [4] or http://www.ktbl.de [5]. In addition, since early 2001, a database with construction cost data on farm buildings is available from KTBL, the model of which is based almost exclusively on the database of the Institute. The Internet offering of KTBL is described in Section 6.2.4.

#### Construction Cost Data—Methodical Procedures

In order to compile construction cost data for barn building systems, it is necessary to construct building models, to calculate the number of individual construction parts and services, and to link them with unit prices. Appropriate barn types and sizes are selected for this purpose.

It is also necessary to establish constructions, building designs, and building part measurements. The functional measures must be in accordance with the individual requirements for husbandry appropriate for animals and optimal work processes. At the same time it must be observed that the selected building design is inexpensive and also that possibilities for personal contributions to the building are limited.

Figure 1 gives a schematic overview of the work processes from the building construction through to the presentation of results. The first step is the construction with

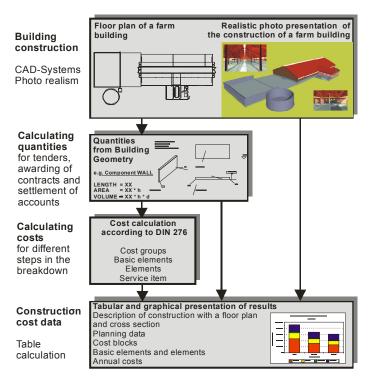


Figure 1. Steps from the construction of buildings through to the representation of the results.

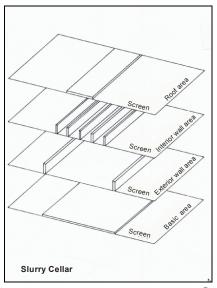
the help of a CAD system. In the Institute, a CAD program with a 3-D volume model is used. In the development of the building model floor plans, different views, cross-sections, and axonometrics are actively linked with each other. This method of working with volume models also has the advantage that during the construction period, total amounts of construction and equipment parts are documented and are available for further processing. On the basis of these amounts derived from building geometry, the cost calculation takes place with a program for tenders, assignments, and calculations (AVA). The costs are calculated on the basis of the German Standard DIN 276 "Costs of Construction."

For each building model a tabular and graphic presentation of results is made. With the help of a tabular calculation program spreadsheet, a large amount of data is brought together and reflected in tables and graphics.

## Building Construction with the Help of a CAD System

The building construction takes place with the program ARRIBA CA3D http://www.rib.de [6], which supports the development and construction of buildings with combined 3-D/2-D functions.

The specific program screen or layer technology is an important part of the planning system. For the building construction the screens are marked according to the same classification system. As in the cost calculation according to DIN 276, a separation of the construction body in various levels is undertaken for processing. In barns this subdivision is as a rule only the slurry cellar and the ground floor. The building parts set in the floor plan are already done in three dimensions behind the building model during the processing (Figure 2). For example, all exterior wall areas are



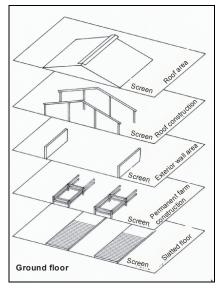


Figure 2. ARRIBA® CA3D screen technology.

assigned to the appropriate external wall screen and all parts of the roof are assigned to the roof screen. An infinite number of screens can be created, processed, and combined.

The CAD program offers catalogs of finished construction parts for different basic construction parts such as foundations, walls, ceilings, or roofs. For agricultural and functional parts these catalogs are generally not adequate. That is why it is necessary to compile building parts in a self-created catalog. This procedure is very complicated because every point has to be mathematically defined, but is unavoidable.

Based on the example of a fattening pig barn for 1000 animals, the following illustrates the advantages of the volume model. Through a layering of the individual floor plan screens, a finished floor plan is created (Figure 3). Cross sections, views, and perspectives can be presented using the volume model (Figure 4).

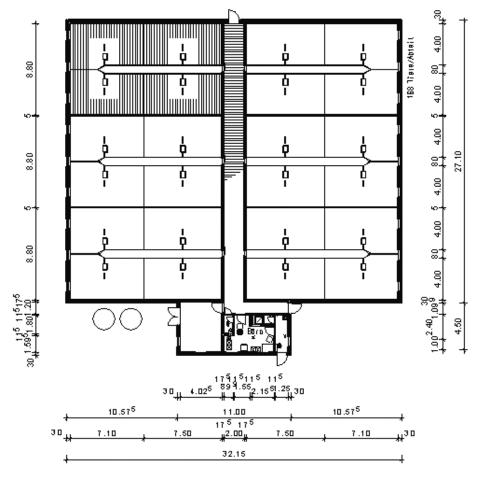


Figure 3. CAD screen: Floor plan of a 1000-slot fattening pig barn.

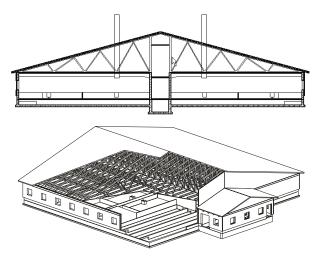


Figure 4. Volume model: Automatic creation of cross sections and perspective presentations.

## Calculating Amounts and Costs

From the constructed volume model the amounts of building and functional parts that are needed can be derived for the subsequent cost calculation. The calculation of costs begins with the adoption of individual prices for construction services in invoices for constructed objects. In various classification steps these individual positions are used for building elements.

All data flows into various databases (catalogs) and forms the basic data for all building models to be drafted. The project and part catalogs, the cost element catalog, the service directories, and the according price catalogs must be constantly updated and expanded for this purpose. A continuous updating of the database is required, but despite the high administrative and time requirements, the database offers the advantage that the root data can be called up and evaluated both in the space planning as well as in the calculation of costs as object data.

The documentation and preparation of construction costs requires the systematic classification of databases. For a cost element catalog, coding is used that is oriented on the organizational system of DIN 276 "Costs in construction" [7].

The cost framework of the element catalog is, as in the standard, three levels of cost division. These are marked with three-digit numbers. In the first level of the cost classification, the total costs are broken down into the following seven cost groups:

100 land

200 furbishing and hook ups for water and electricity

300 building—building construction

400 building—technical facilities

500 external facilities

600 equipment and artwork

700 related costs

In the second level, general building elements are included and at the third level, building elements. Since every building element can occur in different formats, a description of the type of format is added to the fourth and fifth positions of the code. The sixth and seventh positions describe the type of format and include more exact information on the material and exact measurements.

#### **Building Cost Network**

In this system, three different methods and classification systems are drawn together:

- the cost block method for cost estimates and comparisons of alternative solutions in the pre-draft phase;
- the building element method for cost estimates and cost calculations in the drafting phase; and
- format-oriented classification according to service areas for the construction format

On the basis of these databases, it is possible to work with the construction costs from the finely classified level of unit prices for construction services, through to the building elements, through to the general classification levels of the cost blocks with computer technology using the construction costs network (Figure 5) [1,4].

For model calculations at the Institute, first all service positions in the main service directory are adopted and then included in the appropriate unit prices in the price data. Then the compilation of appropriate building elements for the selected barn models takes place, out of the cost element catalog. Each building element is comprised of one or more service positions.

Each building variation, for which the amounts can be calculated from the volume model, can take place with the help of unit prices from the price data of the cost calculation. The permanent classification structure makes it possible to aggregate the data from the levels of elements through to the general elements through to cost groups and to summarize total costs

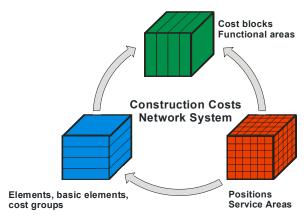


Figure 5. Construction costs network system.

In addition, the network allows the preparation of cost block data. Cost blocks are function-related building parts or construction part groups. Divided into BARN, MANURE, FEED, MILK, EGGS, and AUXILIARY FACILITIES, the cost blocks give the total investment necessary and relate to the individual use unit again. In fattening pig houses, as can be seen in Figures 3 and 4, only the cost blocks BARN, MANURE, FEED, and AUXILIARY FACILITIES come under consideration.

Thus, the cost block BARN includes building constructions like the construction pit, foundation, exterior and interior walls, ceilings, and roofs as well as constructional built-in elements, such as cement slatted floors or technical facilities for wastewater, water, gas, heat, ventilation facilities, heavy current, and specific-use facilities, particularly barn details like box divisions and animal slots.

The cost block MANURE includes constructions and technical facilities for the transport, homogenization, and storage of manure, liquid manure, and slurry. This cost block includes the construction parts of the buildings and, where applicable, the related external facilities.

The cost block FEED includes constructions and technical facilities for feed storage, preparation, and feeding in the building and external facilities.

The cost block AUXILIARY FACILITIES includes constructions and technical facilities in the hygienic and office area as well as for the loading ramp in the building and external facilities

#### Construction Cost Data

After the planning, construction, and subsequent calculation of numerous construction costs are available, the preparation of cost factors and the derivation of planning characteristics takes place with a tabular calculation program, Microsoft Excel. This takes place with the spreadsheet BAUKODA [1], which was developed by the Institute of Production Engineering and Building Research especially for this purpose.

Since construction cost data is calculated above all for a certain barn model, a drawing for each variation, as a rule consisting of a cross-section, barn floor plan, and slurry canal system, is made available for comparison. Furthermore, the planning characteristics (according to the German Standard DIN 277) such as the gross ground area, the construction area, or the gross capacity are calculated and placed in relation to the main use area. In addition, agricultural planning characteristics, such as barn area or slurry storage per animal are important as a measure for cost comparisons. These inputs are necessary for comparisons of buildings [8] and in particular for the selection of comparative objects.

The construction cost data of the Institute was given to the KTBL for further publication. The publication on the Internet is described in the following sections.

## 6.2.4 Cost Data for Farm Building

For the planning of farm buildings, inputs about investment needs and annual construction costs are an important prerequisite. Complete constructions are generally not available in the catalog and at fixed cost. In order to help determine the economic viability of an investment at the outset, and to secure financing, one must have planning

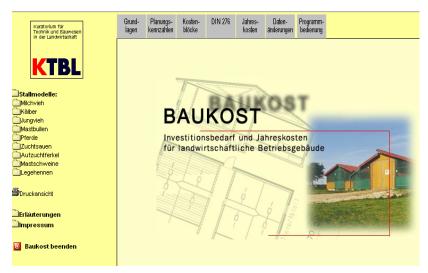


Figure 6. Construction cost calculation with the BAUKOST database from KTBL.

data and comparative data from other, similar constructions. For this purpose, KTBL has provided building cost values for different buildings and permanent facilities via the computer program "BAUKOST" on the Internet (Figure 6) since the end of 2002 and offered on CD-ROM since 2003 [5].

## The BAUKOST Database from KTBL

The provision of construction cost data makes cost calculation in the planning and evaluation stage easier, in particular for building consultants but also for other building professionals (architects, surveyors, farmers, banks, insurance, etc.).

The database makes construction cost data for farm buildings available. It is almost exclusively based on the database of the Institute of Production Engineering and Building Research of the FAL and on model barns of the Institutes for Agricultural Engineering, Construction and Environmental Technology of the Bavarian State Institute for Agriculture [4,5]. It is assumed that standard barn models are built by external firms without own labor. At the moment 107 barn models are available (Table 1).

Type of Production	Number of Barn Types
Dairy Cattle	16
Calves	6
Heifers	19
Fattening Bulls	10
Horses	14
Breeding Piglets	6
Fattening Pigs	16
Laying Hens	15
Broilers	6

Table 1. Overview of barn models in BAUKOST.

Through the selection of the available barn models, different husbandry practices and equipment variations are taken into consideration. The unified data base makes possible comparisons between different husbandry practices and herd sizes for one type of building [9].

The construction investment needs are classified in BAUKOST and made available as described in Section 6.2.3. All prices are based on wholesale prices and do not include value added tax.

## Design and Use of BAUKOST

The barn models in Table 1 are shown in a navigational tree. After selecting a product direction, the available types can be called up based on husbandry type and herd size. In a main window the barn models are listed and described with keywords. All information buttons are shown by clicking on a barn type. Thus it is possible to call up a drawing for each barn model as presented in Figure 7. Also, all previously mentioned individual information is visible and printable from the computer screen.

Building drawing (Figure 7)	Cost groups	Annual costs
Building description	Cost general elements	Change barn size
Planning characteristics	Cost elements	Change price level
	Cost blocks	

To compare the costs of various barns, the data from up to three models can be called up simultaneously. The display makes clear whether there is a cost reduction in the case of larger herds and how much cost difference there is between different husbandry practices and associated equipment. The results can be presented as shown in Table 2.

In the cost breakdown by cost blocks, the building is broken down into functionrelated buildings parts of part groups as in Section 6.2.3. A further subdivision of the cost blocks into three use periods (long-term, mid-term and short-term) makes possible a differentiated calculation of annual building costs from depreciation, interest, insurance, repairs, and operational costs. Through changes in the present approaches in the spreadsheet, annual costs adjusted to individual situations, such as length of use, repairs, interest, or insurance are possible.

Table 2. Presentation of results for BAUKOST.

			Reference
		Absolute	Size
1	Construction costs broken down according to cost groups	X	Per animal
	Building-Building construction		
	Building-Technical facilities		
	External facilities		
2	As in 1, but further broken down into general building elements	X	Per animal
3	As in 2, but further broken down according to building elements	X	Per animal
4	Cost breakdown classified according to cost blocks and depreciation	X	Per animal
	deadlines		
5	Calculation sheet to derive the annual costs	X	

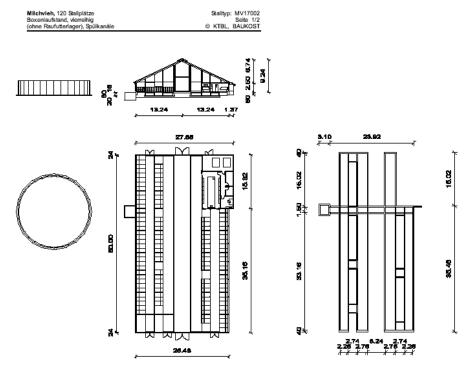


Figure 7. Description of barn models via floor plans and cross-sectional drawings on the basis of the example of a 120-slot lying-box barn for dairy cattle.

In addition to the previously mentioned display possibilities, BAUKOST also offers the possibility to adjust herd size and price level. For most models in BAUKOST there are "similar buildings," which are models of different sizes with the same construction style, equipment, and husbandry method. Comparable models can be displayed simultaneously. To a limited extent a building size can be freely chosen. For this purpose an interpolation or extrapolation of the investment needs can be calculated for a similar model. This is a purely mathematical calculation of costs.

Furthermore, the price level can be changed via freely selectable factors. With the help of an official building index, adaptations can be undertaken to adjust for time and economic differences. Even regional influences can be included with greater or lesser charges.

In the CD version it is also possible to change, delete, or expand individual elements of a model. Amount, unit price, and building description can be freely chosen. Thus, flexibility increases by using construction cost data, but it requires a professional use and good professional knowledge. These changes result in a new model that can also be edited in the drawing and planning figures. Building drawings cannot be created for own models.

## Example: Dairy Barns

With BAUKOST, it can be determined what influence the husbandry practice and the style of building have on the investment needs and the annual construction costs for a targeted selection of barn models with the same size herd. This becomes clear in a comparison of dairy cattle barns with different husbandry practices and herd sizes.

Table 3. Comparison of cost blocks for dairy barns with 188 slots under different husbandry practices.

- Table 5. Co			•	i 100 siots unuel u		
	2 × 2 rows ( flushing car	cubicle housing, no feed storage), nals, 188 animal ce level 2000	rows (no fee	deep litter barn, 2 d storage), tractor removal, 188 , price level 2000	2 rows (no tractor many	sloped floor barn, o feed storage), ure removal, 188 , price level 2000
Tyma	Cost in Euros		Cost in Euros		Cost in Euros	
Type	Total	Per Animal Unit	Total	Per Animal Unit	Total	Per Animal Unit
Total costs	653,239	3,474.68	556,521	2,960.22	542,987	2,888.23
Cost block BARN	289,799	1,541.48	297,178	1,580.73	260,853	1,387.52
Long-term investments	183,321	975.11	272,252	1,448.15	236,433	1,257.62
Mid-term investments	75,797	403.18	22,058	117.33	22,058	117.33
Short-term investments	30,681	163.20	2,868	15.26	2,362	12.56
Cost block MANURE	175,836	935.30	84,239	448.08	106,647	567.27
Long-term investments	107,121	569.79	0	0.00	0	0.00
Mid-term investments	60,534	321.99	78,870	419.52	101,278	538.71
Short-term investments	8,181	43.52	5,369	28.56	5,369	28.56
Cost block FEED	5,305	28.22	5,562	29.59	5,562	29.59
Long-term investments	0	0.00	0	0.00	0	0.00
Mid-term investments	3,608	19.19	3,865	20.56	3,865	20.56
Short-term investments	1,697	9.03	1,697	9.03	1,697	9.03
Cost block ANIMAL PRODUCT	182,299	969.68	169,542	901.82	169,925	903.86
Long-term investments	64,863	345.02	52,106	277.16	52,489	279.20
Mid-term investments	0	0.00	0	0.00	0	0.00
Short-term investments	117,436	624.66	117,436	624.66	117,436	624.66

An initial economic consideration and general cost estimates are often provided by a comparison of function-related construction part groups on the basis of cost blocks. Table 3 shows that, for example, the cost block MANURE should be set at almost double the level for a laying-box free-range barn with liquid slurry practices as for a deeplitter barn of the same size with two area bays. The cost block FEED shows only minimal differences. With this breakdown of costs, the cost differences related to husbandry practices and herd size can be seen in a very early planning phase. Cost-effective plan changes for entire functional areas can be more easily judged than is possible with values related to ground areas and spatial areas.

Figure 8 shows a comparison of four different barn models with different husbandry practices but the same number of animals, the differences between the cost blocks BARN, MANURE, FEED, and ANIMAL PRODUCT. Figure 9 gives an overview of

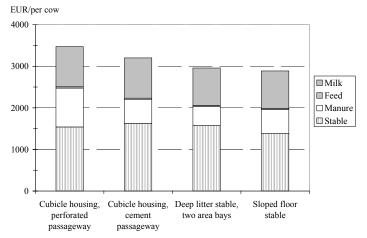


Figure 8. Investment demands for dairy barns with 188 slots under different husbandry methods broken down into cost blocks.

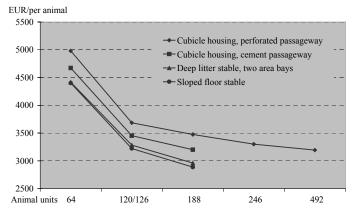


Figure 9. Investment demands for dairy barns per animal slots.

the total investment needs of building models available for dairy cattle husbandry in BAUKOST. Cost differences between husbandry practices are illustrated, and a perhead cost reduction with increasing herd size can clearly be seen. It is equally clear that in large herd sizes the per-head cost reduction is less.

## 6.2.5 Further Data and Program Offers

BAUKOST is a database developed especially for agricultural building measures. A range of further programs and databases is also available targeted to the entire building industry and containing construction cost information for farm buildings to varying extents.

The building cost information center (BKI) of German architectural chambers maintains a similar classified, extensive database, but this hardly contains farm buildings. The data is available in handbooks, on CD-ROM, and online (www.bki.de) [10].

The German Ministry for Transport, Building and Housing publishes normal manufacturing costs especially for the value assessment of buildings. This tabular work also includes data for farm buildings (www.bmvbw.de) [11].

The ALB Hessen publishes annual "Price guidelines for new and renovated farm buildings and rural houses." This comprehensive collection mainly contains individual prices and is suited in particular for assessing the value of individual services as well as renovations or extensions. The data is available in print as well as electronically (www.alb-hessen.de) [12].

In Switzerland the Agroscope FAT Tänikon Eidgenössische Forschungsanstalt maintains a building cost collection of farm buildings. This includes cost statistics per animal slot, per building volume or per use spatial unit based on finished and paid for buildings. The mean value per cow slot, per fattening slot, per m³ slurry storage or per m³ construction volume are updated via a construction cost index. A PC program is under development with price guidelines to calculate construction costs and assess annual costs (www.fat.ch) [13].

In the framework of farm planning, the construction costs and the resulting annual building costs in the network with the production process labor and feed and other costs can be seen. For this purpose KTBL has made a comprehensive data collection available containing a CD-ROM for the calculation of machinery and labor costs [5].

#### References

- 1. Gartung, J., and F. Preiß. 1997. KTBL-Arbeitspapier 237. Informationstechniken für das landwirtschaftliche Bauen. *Baukostenermittlung mit dem elektronischen Arbeitsblatt BAUKODA für Windows*, 23-46.
- Gartung, J. 1988. Methods for the cost-determination of agricultural building and facilities. Wooden Buildings in Agriculture: Papers presented at a Seminar of Section II of CIGR, ed. C. M. Dolby, 20. Swedish University of Agricultural Sciences, Department of Farm Buildings, Lund, Sweden: Sveriges Lantbruksuniversitet.

- 3. KTBL. 2002/2003. Taschenbuch Landwirtschaft, Daten für betriebliche Kalkulationen.
- Gartung, J., and K. Uminski. 2004. Bauökonomik im landwirtschaftlichen Bauen im FAL-Informationssystem, Institut für Betriebstechnik und Bauforschung. Bereich Gebäude und bauliche Anlagen. Accessed on December 22, 2004, from http://www.bb.fal.de.
- 5. Frisch, J., and E. Witzel. 2004. Stallbaukosten-Datenbank im KTBL-Informations system, *Fachthema Stallbau*. Accessed on December 22, 2004, from http://www.ktbl.de.
- 6. RIB Software AG. 2004. ARRIBA® planen *und* ARRIBA® CA3D *unter Produkte*. Access ed on December 22, 2004, from http://www.rib.de.
- 7. Normenausschuss Bauwesen (NABau) im DIN Deutsches Institut für Normung e.V. Juni 1993. *Deutsche Norm "Kosten im Hochbau, Building costs"*. *DIN 276*.
  - 8. Bockisch, F.-J., and J. Gartung. 2002. Construction costs for the manufacture of concrete solid and liquid manure storage containers. Eds. N. de Belie and B. Sonck. *Concrete for a sustainable agriculture: Agro-, aqua and community applications; Proc. of the 4<sup>th</sup> International Symposium, 35-41.* Ghent, Belgique: Ghent Univ.
  - 9. Gartung, J., and K. Knies. 1999. Investment demand for hen sheds. *Agricultural Engineering Research Agrartechnische Forschung* 5: 135-144.
  - 10. Riering, E. 2004. BKI-Baukostendatenbank in *BKI-Fachinformationen*. Accessed on December 22, 2004, from http://www.bki.de.
  - 11. Bundesministerium für Verkehr, Bau- und Wohnungswesen. 2002. *Normalherstellungskosten in Anlage 3, Wertermittlungsrichtlinien 2002 (WertR 2002)* im BMVBW-Informationssystem. Bauwesen, Städtebau und Raumordnung. Arbeitshilfen, Leitfaden, Richtlinien. Accessed on December 22, 2004, from http://www.bmvbw.de.
  - 12. Arbeitsgemeinschaft für Rationalisierung, Landtechnik und Bauwesen in der Landwirtschaft Hessen e.V. 2003/2004. Richtpreise für Neu- und Umbau landwirtschaftlicher Wirtschaftsgebäude und ländlicher Wohnhäuser. ISBN 0945-4985.
  - 13. Hilty, R., L. Van Caenegem, and D. Herzog. 2002. *FAT-Preisbaukasten*. Accessed on December 22, 2004, from http://www.fat.ch, http://www.fat.admin.ch/d/veran/praes/2004/hir1bd.pdf.

# 6.3 Monitoring, Prediction, and Control of the Micro-Environment

D. Berckmans and E. Vranken

Abstract. Each living organism (human, animal, plant) is living within an imperfectly mixed fluid (e.g., moist air) and is responding to its local micro-environment. To realize an efficient control of the biological response, continuous monitoring, prediction, and control of the micro-environment to variations of control inputs is required. This section describes how information technology can help us to realize active control of 3-D micro-environments around living organisms.

**Keywords.** 3-D micro-environment, Imperfect mixing, Biological response, Ventilation rate control, Airflow pattern control.

## 6.3.1 Evolution of Agricultural Buildings

In the past, livestock houses had mainly a shelter function. Early scientific research in this field showed the effect of climate on animal performance [1] leading to engineering concepts in the construction of livestock houses. Later, the practice of livestock keeping without a direct link to the use of outside land began. This made the function of livestock houses evolve from a shelter function to a production tool, with a main focus on animal production (meat, growth, milk, eggs). Traditional livestock houses became production units at a large scale. The farmer kept bigger groups of animals and was no longer able to "know" his animals individually. The scale of greenhouses increased in a similar way. Greenhouse climate control needed more sophisticated controllers to guarantee good growing conditions.

The introduction of electronics in process control units led to more automatic control equipment in livestock houses, greenhouses, and storage rooms. The next step was the appearance of computers and digital controllers, with several companies becoming specialized suppliers with expertise in sensors, ventilation systems, and controller hardware. It has to be noted, however, that the development of most of this equipment was driven by technology companies not familiar with biological processes. Today, the technical capacity to make controllers or computers reliably is common and these companies are searching for ways to apply information technology to biological processes.

# 6.3.2 The Importance of Micro-Environmental Control in Agricultural Buildings

It has been known for a long time that the micro-environmental conditions (temperature, humidity, gas concentration, dust, etc.) in the occupied zone of an agricultural building (livestock building, greenhouse, or storage place) influence the response of animals, plants, or products, and consequently have an impact on growth rate, development, product quality, and health [2-8]. Table 1 gives a brief literature overview to demonstrate the influence of climatic parameters on health status, welfare, and performance of animals, plants, and biological products.

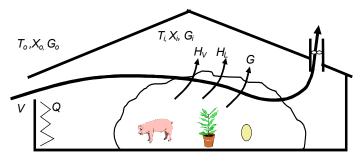
Authors	Environmental Variable	Observed Influence
Cannell and Thornley, 1998 [9]	Temperature and CO <sub>2</sub>	Leaf and canopy photosynthesis
Daudet et al., 1999 [10]	Wind speed	Influence on photosynthesis, stomatal conductance, transpiration rate (and other tree functions)
Hartung et al., 1994 [11]	Dust, micro-organisms	High dust and micro-organism concentrations result in respiratory diseases and reduced infection resistance
Hilbert et al., 1991 [12]	Carbon dioxide and daily photon-flux density	Influence on leaf nitrogen concentration and root:shoot ratio
Kittas et al., 2001 [13]	Greenhouse ventilation regime	Influence on the microclimate and energy partitioning of a rose canopy during summer conditions
Nienaber et al., 1993 [14]	Temperature	Room temperature plays an important role in the eating behavior of pigs.
Ogilvie, 1993 [15]	Temperature, air speed	Under practical field conditions, within certain limits, the influence of temperature and air speed on the health status of pigs is not significant, at least in the case of ad libitum feed supply.
Stanghellini and van Meurs, 1992 [16]	Light, temperature and air humidity	Influence on crop transpiration
Yang et al., 1995 [17]	Aerial conditions	Effect on heat and mass exchange between plants and air in greenhouses

Table 1. Literature overview to demonstrate the influence of climatic parameters on the health status, welfare, and performance of animals, plants, and biological products.

The micro-environmental conditions in a biological production unit are the result of the internal production of heat, moisture, and gases and the removal of these elements from the occupied zone, mainly by means of air mass transport. Consequently, the purpose of a ventilation system is to control the micro-environment (temperature, humidity, gas concentration, dust, etc.) in the occupied region at an acceptable level corresponding to a maximal production efficiency and/or thermal comfort.

The three basic equations in control of the environment are the sensible heat balance, the moisture balance, and the gas balance (Figure 1). Consideration of these balances for an agricultural building show that control of ventilation rate is of great importance. Given the simplifying assumption of perfectly mixed air in the house and a homogeneous temperature under steady state conditions, then the balances for sensible heat, moisture, and gases can be determined and used to calculate the ventilation rate needed to control indoor air temperature, humidity, and/or gas concentration (Figures 1 and 2). In an efficient environmental control system, these three equations satisfied most of the time during the year.

In order to control the biological response of living organisms in a imperfectly mixed fluid, four logical steps must be realized: (1) control of the ventilation rate; (2) control of the airflow pattern; (3) control of the micro-environmental conditions; and finally (4) integration of the animal response in the control system. These four different steps in bioprocess control are shown in Figure 2.



$$H_{V} + Q = V \cdot \gamma \cdot c_{DA} \cdot \left(T_{i} - T_{O}\right) + \sum_{k} \cdot S \cdot \left(T_{i} - T_{O}\right) \tag{1}$$

$$\frac{H_L}{\varepsilon} + W = V \cdot \gamma \cdot (X_i - X_o) \tag{2}$$

$$G=V.(G_i-G_o) \tag{3}$$

 $H_V$  = sensible heat production (J/s),  $H_L$  = latent heat production (J/s), G = gas production (kg/s), Q = heat supply (Q > 0, Q < 0)(J/s), V = ventilation rate ( $m^3$ /s),  $\gamma$  = air density (kg DA/ $m^3$ ),  $c_{DA}$  = heat capacity of dry air (J/kg DA.K), k = heat transfer coefficient (J/s,  $m^2$ .K), S = surface area of heat exchange ( $m^2$ ),  $T_i$  = (desired) inside air temperature (K),  $X_i$  = inside absolute humidity (kg  $H_2$ 0/kg DA),  $G_i$  = inside gas concentration ( $kg/m^3$ ).  $T_O$  = outside air temperature (K),  $X_o$  = outside absolute humidity (kg  $H_2$ 0/kg DA),  $G_o$  = outside gas concentration ( $kg/m^3$ ).

Figure 1. Calculation of the sensible heat (1), moisture (2), and gas (3) balance.

A list of symbols is at the end of this section (p. 397).

To achieve an efficient micro-environmental control for such a complex process, control theory offers the possibility of model-based control. A mathematical model is used to predict how the process outputs will behave and this information allows more efficient control. Most of the models that can be found in the literature cannot be used directly for control purposes because they are too complex, because of the mechanistic equations, and most of them are only valid under steady-state conditions. With newly available software, an accurate model of such complex processes can be realized by using on-line mathematical techniques based on continuous monitoring of the inputs and outputs of the considered process [18].

This principle of model-based control can be applied at all four different levels indicated in Figure 2. An overview is given in the next sections.

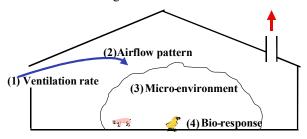


Figure 2. Four steps in bioprocess control.

## 6.3.3 Optimal Control of the Ventilation Rate

#### Importance of Ventilation Rate Control

Many authors have used the balance equations to determine the relationship between physical micro-environmental parameters and the ventilation rate through a ventilated building [19-22]. They all conclude that, especially at low ventilation rates, the effect of a variation of the ventilation rate on the resulting inside temperature, the inside air humidity, and the gas concentration is relatively strong. As an example, Figure 3 represents the relation between the inside temperature in a livestock building and the ventilation rate at different levels of outside temperature.

It is clear from the three basic equations in Figure 1 that they are linked by one variable: the ventilation rate. As a result, changing the ventilation rate has a direct effect on the resulting indoor temperature, humidity, and gas concentration. Consequently, the need for adequate control of the ventilation rate is formulated as an absolute criterion for good and stable internal climatic conditions.

However, from laboratory tests and field analysis of existing ventilation equipment it has been shown that most agricultural ventilation systems do not succeed in adequately controlling the ventilation rate and there are serious doubts about the economics of the beneficial influence of many automatically controlled mechanical ventilation systems [23-25]. It was concluded that, before one can introduce more sophisticated control algorithms in the field, good control of the ventilation rate through the building is required.

In addition to its influence on indoor climatic conditions, the ventilation rate also affects the emission of pollutants (such as ammonia and dust) from animal houses. Inside the houses, high ammonia concentration represents a potential health hazard to workers and animals [26]. High ammonia emission is associated with environmental and ecological damage, and [27,28] concluded that it was possible to reduce the yearly ammonia emission from a pig house by about 10% through adequate control of the ventilation rate throughout the year.

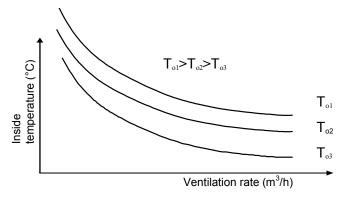


Figure 3. The relationship between the inside temperature  $(T_i)$ , the outside temperature  $(T_0)$ , and the ventilation rate (V) in a livestock building, assuming Q=0 in the sensible heat balance.

From literature and field research, it can be concluded that the ventilation rate is the most important parameter in climate control of agricultural buildings because of its determining character on the *global level of the internal climatic conditions*, their *internal spatial distribution*, the resulting *health status* of the biological material, the *production results*, the *pollutant emission*, and the overall *energy consumption* for ventilation and heating. Consequently, improving ventilation rate control in agricultural buildings is expected to have several benefits.

In practice, good control of the ventilation rate of a building is not straightforward. In most buildings, axial fans are used because they are relatively cheap and they can produce a wide range of ventilation rates. Their disadvantage, however, is the high sensitivity to pressure differences as caused by wind action, variation of air inlet positions, etc.

One way to achieve better control of the ventilation rate is negative pressure control, which is a very common technique in the US. A pressure sensor is used to control the position of the air inlets in such a way that there is a constant negative pressure of about 20 Pa. This assures that the fan is working in its stable working zone and under that condition there is a unique relationship between the fan's voltage and the ventilation rate

Another solution is to measure the ventilation rate continuously and to use this information in the feedback loop of a ventilation rate controller. This technique is often used in Europe and has some advantages over pressure control, as pointed out in the next section.

## Active Closed-Loop Control of the Ventilation Rate

The main control principle in today's agricultural applications is feedback control of inside temperature and open-loop control of all other variables, such as ventilation rate, humidity, emissions, airflow pattern, and biological responses (Figure 4a). Using information technology and introducing model-based control at several levels brings substantial advantages in the control of biological processes.

The integration of feedback in a traditional ventilation control system (Figure 4b) has some major advantages, as demonstrated in laboratory and field tests [22]. The main reason to use active feedback control of the ventilation rate is that the dynamic variation of the ventilation rate is in a totally different frequency range than the variation of outside and resulting inside temperature. Moreover, fluctuations in ventilation rate have a negative effect on the level and the stability of airflow pattern, 3-D microenvironment, emissions, and energy use.

In such a feedback control system, a sensor is needed that continuously measures the ventilation rate. For this purpose, several techniques are described in the literature, each of them having their specific accuracy. In general, ventilation rate monitoring can be categorized into two groups: direct and indirect methods. Indirect methods, such as tracer gas, CO<sub>2</sub> balance or heat balance methods, can give a fairly good indication of the ventilation rate over a longer period of time (e.g. several hours), but they assume

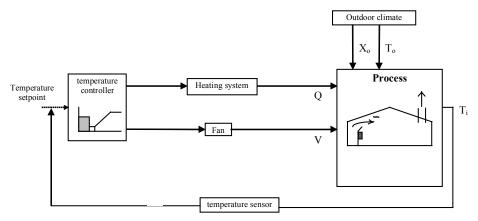


Figure 4a. Traditional open-loop climate control.

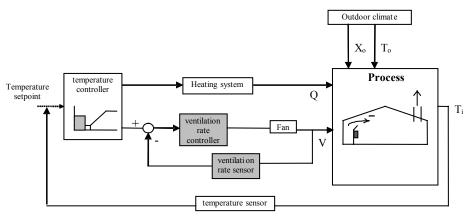


Figure 4b. Integration of a ventilation rate controller in a traditional climate control system.

steady-state conditions and a perfect mixing within the ventilated space. Since these conditions are not fulfilled under practical field conditions, measurement errors up to 40% often occur. In general, these indirect monitoring techniques are not accurate and not fast enough to be used for on-line monitoring of the highly fluctuating ventilation rate in ventilated structures. Moreover, most of them are too expensive for continuous control purposes.

On the other hand, direct methods try to monitor the velocity (distribution) in every ventilation opening and most of them are based on velocity point measurements in a ventilation section. Commonly used sensors are pitot tubes, hot wire anemometers, vane anemometers, etc. The main disadvantage of these direct measuring techniques is that they assume a homogeneous velocity distribution in the measuring section. However, in most applications of naturally or mechanically ventilated systems, the air-stream pattern can be very complex and turbulent and the measured air velocity

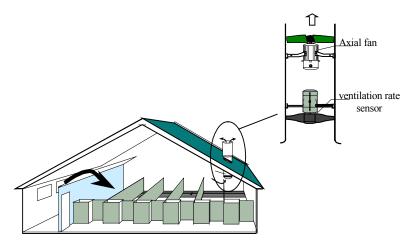


Figure 5. Principle of a free-running impeller covering the complete ventilation opening.

profile is strongly dependant on the static pressure difference over the ventilation opening. A second problem to overcome is the temporal fluctuation of the velocity profile. Especially in larger inlet or outlet sections, pressure fluctuations as a result of wind action can be considerable. A ventilation rate sensor that can be used under these conditions must have a fast response (time constant of 3 seconds, optimally) to be able to measure the overall ventilation rate as a result of these fast fluctuations.

From these findings it is clear that the velocity profile in the desired measuring plane is very complex and that a good airflow measurement technique in such a case requires a unique sensor design. For continuous and accurate ventilation rate measurements under field conditions, as needed for control purposes, in principle only a free-running impeller covering the whole ventilation opening gives satisfying results. Given an optimal blade design and construction requirements, an accuracy of 1% of full scale was found in a working range between 200 and 14 000 m³/h in combination with pressure differences between 0 and 120 Pa [29]. Figure 5 shows how such a free-running impeller can be mounted in the chimney of a mechanically ventilated agricultural building.

The integration of an accurate free-running impeller into the feedback loop of a ventilation control system significantly improved the stability of ventilation rate and the resulting indoor climatic conditions, as shown by comparison field tests [29]. During this field trial, the ventilation rate was monitored simultaneously in two identical pig house compartments with different fan control systems. Comparison of the measured ventilation rate in compartment 2 (open-loop voltage control) and compartment 1 (feedback ventilation rate control) during one week is illustrated in Figure 6. For both compartments, the desired ventilation rate was 300 m³/h. From this example, it is clear that a closed loop ventilation rate controller resulted in a more stable and accurate ventilation rate compared to the open-loop, voltage-based controller (standard deviation 11 versus 167 m³/h over the seven day period shown).

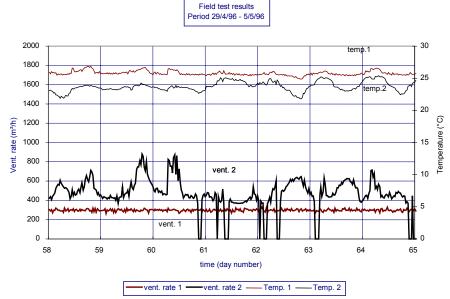


Figure 6. Measured ventilation rate and average indoor temperature with two different fan control strategies (1 = feedback ventilation rate control, 2 = open-loop voltage-based control).

## 6.3.4 Active Control of the Airflow Pattern in the Ventilated Space

In confined spaces, the distribution as well as the amount of air entering the room per unit of time has a great influence on providing optimal conditions for the living organisms in the space [20,30]. Airflow patterns in livestock buildings and warehouses influence the distribution of air velocities, temperature, gas concentrations, and the release of gases from manure [11,31]. In today's agricultural buildings, the airflow pattern is not actively controlled. In most cases, indoor temperature measurements are used to adjust the inlet openings, but the relatively slow response of the measured indoor temperature on fast fluctuations of the airflow pattern do not permit an adequate control. Consequently, the measurement and control of airflow pattern have become important concerns of scientists in the last decades [32,35].

Direct determination of the airflow pattern for research purposes is possible using image processing techniques [36,37]. The techniques are based on the visualization of the two-dimensional airflow pattern using smoke as a tracer of the air entering the room (Figure 7). The accuracy of the prediction of the centerline in a test room with dimensions 8 m by 4 m by 2.5 m was 0.05 m when the camera was placed at a distance of 4 m from the test installation [38]. Despite its accuracy, this type of airflow pattern monitoring can only be used for scientific purposes and is no solution for continuous measurements as needed for control purposes.

Air velocity measurements have also been widely used for airflow pattern measurements [31,34,35,39,40]. Similar to the previous method, this measurement technique is accurate (1% to 5%), but it is expensive and can hardly be used in the dirty and corrosive environment of livestock buildings.

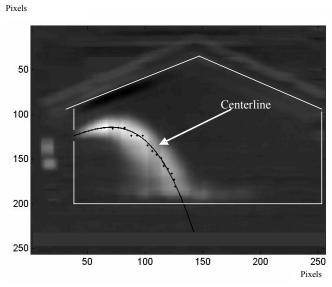


Figure 7. Determination of air flow pattern centerline and drop distance with image analysis.

Airflow pattern quantification in literature is mostly linked to Archimedes number, which is the ratio of buoyancy forces to initial jet momentum [41]. Taking into account the room and inlet dimensions, Randall and Battams [21] introduced the corrected Archimedes number  $(Ar_c)$  as:

$$Ar_{c} = \frac{C_{d}abBH(B+H)(T_{hs} - T_{0})}{V^{2}(T_{hs} + T_{0})}$$
(4)

where  $C_d$  = the discharge coefficient

a and b = the vertical and horizontal dimensions of the air inlet (m)

B and H = the width and height of the ventilated room (m)

V = the ventilation rate (m<sup>3</sup>/s)

 $T_o$  = the supply air temperature (K)

 $T_{hs}$  = the temperature (K) of the "heated surface," i.e., the temperature of the hottest element in the room (e.g., the heating element or the organism)

In theory, Equation 4 can be used to calculate the vertical inlet dimension a that corresponds to a desired Archimedes number for a given ventilation rate V and inlet temperature  $T_o$ .

The use of Archimedes number and drop distance as a parameter to characterize the flow pattern was suggested by Leonard and McQuitty [42]. Jet drop distance is defined as the horizontal distance from the inlet to the point where the jet reaches the occupational zone [43]. The correlation between drop distance and inlet  $Ar_c$  was studied for different inlet openings and inlet configurations provided by a guiding plate. It was concluded that  $Ar_c$  could fully represent the total effects on temperature differential and ventilation rate of drop distance but it could not represent the entire inlet open-

ing effects on drop distance and could not represent the configuration effect at all [39]. Another difficulty of using  $Ar_c$  is to calculate the effective opening area to find the vertical dimension of the inlet, a. Zhang et al. [39] used the relation between pressure difference and velocity of air at the inlet to calculate the area. However, this indirect measurement was subject to errors coming from accuracy of the instruments. Besides, in field situations the air velocity is not equally distributed over all air inlets because of wind action. In conclusion, the use of Archimedes number does not permit prediction of the airflow pattern in an accurate way under practical field conditions and consequently other measuring techniques must be developed.

In many commercial systems nowadays, inside pressure is adjusted to control the air inlets, but this system does not accurately control airflow patterns because it does not directly take into account the temperature difference between inside and outside, and provides a feed-forward control assuming pressure as the only disturbance.

Numerous investigators have found numerical simulations, such as computational fluid dynamics, to be valuable tools for indoor airflow prediction [33,44-46]. Although all numerical methods need experimental data against which they can be validated, there is a surprising dearth of such data [35]. The major difficulties are that setting up the model and identifying and specifying appropriate boundary conditions are difficult. Moreover, the calculations are very time consuming, require large amounts of computer memory, and, indeed, sometimes fail to converge to a solution [47-49].

It was concluded from the literature and our own observations that a low-cost and accurate sensor that can be used in the feedback loop of an airflow pattern controller was still missing and consequently a new sensor concept had to be developed.

A novel airflow pattern sensor was based on the idea of predicting airflow patterns from inlet temperature distribution [50]. It was shown that temperature sensors on an arc located at the inlet were sufficient to predict the trajectory of the air jet [51] and could be used as a low-cost airflow pattern sensor (Figure 8) for control purposes.

In this sensor concept, eight thermocouples were fixed on an arc with a radius of 0.3 m to predict the 2-D centerline of the airflow pattern in the room. The thermocouples were placed at -50°, -25°, 0°, 25°, 40°, 50°, 60°, and 75° to the center of the inlet

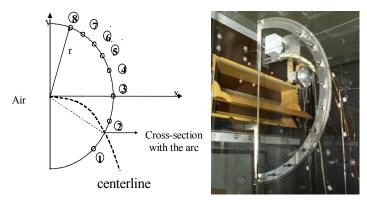


Figure 8. Schematic representation and picture of airflow pattern sensor at the air inlet.

and were named from  $T_I$  to  $T_8$ , respectively. The arc was positioned in a vertical plane perpendicular to the length of the slotted air inlet. From these temperature observations the 2-D trajectory of the airflow pattern and the drop distance (DD) was predicted by means of a mathematical model. Experiments in a full-scale test installation showed that a linear model structure (Equation 5) was sufficient to predict the drop distance with an accuracy of 0.73 m in a room with dimensions of 8 m by 4 m by 3 m.

$$DD = m_1 T_1 + m_2 T_2 + \dots + m_8 T_8 + m_9 T_{in} + m_{10} T_{room}$$
 (5)

where  $T_1, T_2, ... T_8 =$  arc temperatures

 $T_{in}$  = incoming air temperature

 $T_{room}$  = average room temperature

 $m_1, m_2, \dots m_{10}$  = fixed parameters for a specific inlet and room configuration

In a final step, this type of airflow pattern sensor was integrated in the feedback loop of a model-based airflow pattern controller. For this purpose, dynamic experiments were performed in a real-scale test installation. During the tests, valve positions were randomly changed in a stepwise manner, while the response of the changing airflow pattern (drop distance, *DD*) was measured with the airflow pattern sensor.

From the experimental dynamic input-output data, a dynamic model was obtained using a general discrete ARX model structure:

$$y(k) + a_1 y(k-1) + a_2 y(k-2) = b_1 u(k-nk) + b_2 u(k-nk-1) + b_3 u(k-nk-2)$$
 (6)

where y(t) = the output (drop distance) at time step t

u(t) = the control input (valve position) at time step t

This model was used in a next step to calculate the control parameters of a model-based airflow pattern control system. An example of the controller performance in the real-scale test installation is shown in Figure 9. Ventilation rate changes were used as

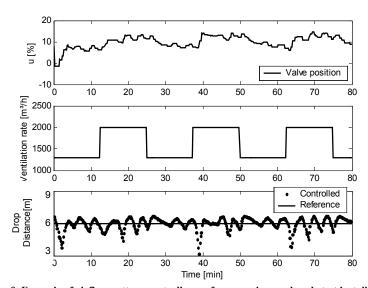


Figure 9. Example of airflow pattern controller performance in a real-scale test installation.

a disturbance variable. The controller showed only a small deviation from the optimal drop distance of less than 25 cm. It was concluded that the concept of an airflow pattern sensor offers new possibilities for on-line control of the fresh air distribution in a ventilated space.

#### 6.3.5 Model-Based Control of the 3-D Micro-Environment

A similar approach of model-based control was used to control the 3-D-microenvironment. A dynamic model is used that predicts the 3-D energy and mass transfer in a non-perfectly mixed ventilated space [18]. The model concept is shown in Figure 10. The ventilated space is imperfectly mixed in terms of temperature, humidity, gas concentrations, etc. It has been shown that in such a room well-mixed zones can be defined for temperature [22,38,52]. A well-mixed zone (WMZ) is a mathematical 3-D volume in which an acceptable temperature gradient is measured. Considering a WMZ in the space (Figure 10), the energy and mass flows to this specific zone can be considered. A part  $(V_c)$  of the total ventilation rate (V) that enters the room will reach the specific WMZ and, similarly, a part  $(O_c)$  of the total heat production (O) reaches the WMZ. To describe the dynamic behavior of micro-climate variables such as temperature in the WMZs, a hypothetic-deductive approach is used. In this approach, a model structure of minimum complexity and with sensible physical interpretation is first formulated by applying physical conservation equations to the WMZs. Once the model structure is formulated in this manner, physically meaningful model parameters are estimated from experimental time-series data, which are measured within the WMZs.

By using today's technology and sensors, this model can be implemented in a controller, resulting in more accurate control of the 3-D distribution of the microenvironment throughout the ventilated place. As a result, it can be used in a model-based control system providing 3-D-temperature distribution control. Figure 11 shows an example of the measured temperature distribution in a room before and after a control action in which the uniformity is optimized by a model-based control of the ventilation rate.

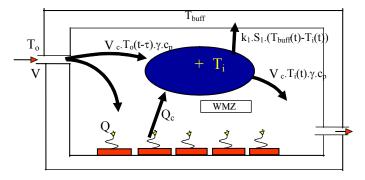


Figure 10. Well-mixed zone approach for the study of imperfectly mixed fluids.

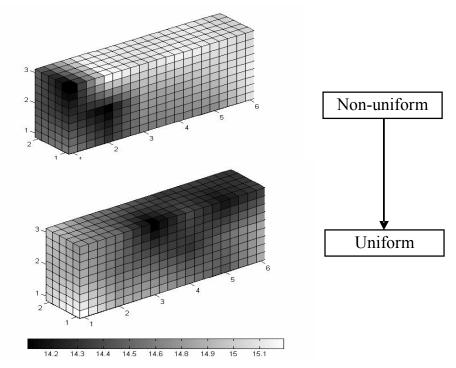


Figure 11. 3-D temperature distribution in imperfect mixed fluids before (top) and after (bottom) a control action. (Room dimensions 3 m by 2 m by 1.5 m.)

## 6.3.6 Model-Based Control of the Individual Biological Response

In this final step it is shown how efficient control of the environmental responses of animals, plants, and products can be achieved by information technology. Although the individual living organism itself is the most important component in an environmental control system, current climate controllers make use of set points that are assumed to be optimal for the "typical" animal, plant, or product. This approach does not always result in the expected performance or welfare since it over-simplifies the complex interactions between a living organism and its micro-environment. Since the final objective remains to promote—or even control—the individual performance or wellbeing, it is logical to measure on-line the biological responses to the process inputs and to integrate these in the control of the environment [53-55]. For animals, this control technology is termed integrated management system or precision livestock farming (Figure 12). The principles of this technology are based on accurate modeling of such complex processes by using on-line mathematical techniques on continuous monitoring of inputs and outputs of the considered process. It has been shown that the approach is applicable for humans, animals, and plants (see a list of projects at http:// labr.be).

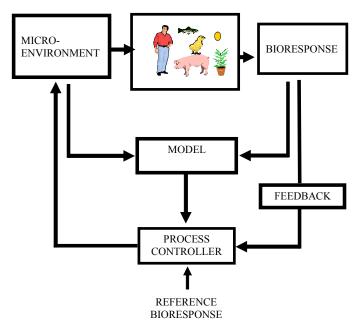


Figure 12. Schematic overview of an integrated management system or precision livestock farming system applied to the environmental management for living organisms.

The responses of living organisms can be physiological, behavioral, or related to production. In fact, any measure that can be related either directly or indirectly to performance and/or behavior, e.g. heart rate, heat production, metabolic rate, photosynthesis, etc., can be used as a response. Given (1) a target or reference value for the bioresponses, and (2) a means to sense these, then a process controller adjusts the microenvironment to produce the desired bio-response. At present, this concept has only been demonstrated in the laboratory, e.g., control of heat production in response to temperature or light intensity of chicken [55], but its potential to improve production conditions for animals, plants, and stored product is great.

Several sensing techniques are under development to measure the biological responses of living organisms. As an example, for laying hens behavioral variables such as scratching, eating, drinking, resting, and spreading the wings can be monitored online by means of image analysis [56]. In future, the on-line monitoring of the response of laying hens will allow prediction of their behavior and an improved control of their welfare by manipulation of the hen's micro-environment.

#### 6.3.7 Conclusions

Information technology can be used to control the micro-environment and the biological responses of living organisms at different levels. In an agricultural building, every individual animal, plant, or product responds to its local micro-environment. Controlling this 3-D micro-environment is an absolute condition for more optimal production results in term of quantity and quality. In this section it is shown that the

use of model-based control techniques in different logical steps results in an optimal environmental control, based on the on-line integration of biological responses into the control actions.

	List of Symbols
a, b	List of Symbols the vertical and horizontal dimensions of the air inlet (m)
B, H	the width and height of the ventilated room (m)
$C_d$	the discharge coefficient
$C_p$	heat capacity of dry air $(J/(kg_{DA} \cdot K))$
$H_V$	sensible heat production (J/s)
$H_L$	latent heat production (J/s)
$G_i$	inside gas concentration (kg/m³)
$G_o$	outside gas concentration (kg/m³)
G	gas production (kg/s)
k	heat transfer coefficient $(J/(s \cdot m^2 \cdot K))$
Q	heat supply to the ventilated space (J/s)
$egin{array}{c} Q \ Q_c \ S \end{array}$	heat supply to the considered well-mixed zone
S	surface area of heat exchange (m <sup>2</sup> )
T	time (s)
$T_i$	(desired) inside air temperature (K)
$T_{hs}$	temperature of the "heated surface" (K)
$T_1, T_2, T_8$	arc temperatures (K)
$T_{in}$	incoming air temperature (K)
$T_{room}$	average room temperature (K)
$T_o$	outside air temperature (K)
u(t)	the control input (valve position) at time step t
V	ventilation rate through the ventilated space (m³/s)
Vc	ventilation rate entering the considered well-mixed zone (m <sup>3</sup> /s)
$X_i$	inside absolute humidity (kg <sub>H2O</sub> /kg <sub>DA</sub> )
$X_o$	outside absolute humidity (kg <sub>H2O</sub> /kg <sub>DA</sub> )
y(t)	the output (deflection angle) at time step t
	air density (kg <sub>DA</sub> /m³)
$\tau$	time constant (s)
$C_{DA}$	heat capacity of dry air $(J/(kg_{DA} \cdot K))$
DD	drop distance
$m_1, m_2, m_{10}$	fixed parameters
$Ar_c$	Archimedes number
$\gamma$ $\tau$ $C_{DA}$ $DD$ $m_1, m_2, m_{10}$	air density $(kg_{DA}/m^3)$ time constant (s) heat capacity of dry air $(J/(kg_{DA} \cdot K))$ drop distance fixed parameters

# References

- 1. Kleiber, M. 1961. The Fire of Life. New York, NY: John Wiley.
- 2. Mount, L. E. 1968. *The Climatic Physiology of the Pig.* London, UK: Edward Arnold.

- 3. Monteith, J. L., and L. E. Mount. 1974. *Heat Loss from Animals and Man.* London, UK: Butterworths.
- 4. Verstegen, M. W. A., E. W. Brascamp, and W. Van der Hel. 1978. Growing and fattening pigs in relation to temperature housing and feeding level. *Canadian J. Animal Science* 58: 1-13.
- 5. Adams, P. 1991. Effect of diurnal fluctuations in humidity on the accumulation of nutrients in the leaves of tomatoes. *J. Horticultural Science* 545-550.
- Challa, H. 1990. Crop growth models for greenhouse climate control. *Theoretical Production Ecology: Reflections and Prospects. Simulation Monographs 34*, eds. R. Rabbinge, J. Goudriaan, H. Van Keulen, F. W. T. Penning de Vries, and H. H. Van Laer, 125-145. Wageningen, The Netherlands.
- 7. Van Henten, E. J. 1994. Greenhouse climate management: An optimal approach. Dissertation. Wageningen, The Netherlands: Wageningen Agricultural University.
- 8. Udink ten Cate, A. J. 1985. Modelling and (adaptive) control of greenhouse climates. PhD thesis. Wageningen, The Netherlands: Landbouwhogeschool.
- 9. Cannell, M. G. R., and J. H. M. Thornley. 1998. Temperature and CO<sub>2</sub> responses of leaf and canopy photosynthesis: A clarification using the non-rectangular hyperbola model of photosynthesis. *Annals of Botany* 82: 883-892.
- 10. Daudet, F. A., X. Le Roux, H. Sinoquet, and B. Adam. 1999. Wind speed and leaf boundary layer conductance variation within tree crown: consequences on leaf-to-atmosphere coupling and tree functions. *Agricultural and Forest Meteorology* 97: 171-185.
- 11. Hartung, J. 1994. Environment and animal health. *Livestock Housing*, eds. C. M. Wathes and D. R. Charles, 25-48. Cambridge, UK.
- 12. Hilbert, D. W., A. Larigauderie, and J. F. Reynolds. 1991. The influence of carbon dioxide and daily photon-flux density on optimal leaf nitrogen concentration and root:shoot ratio. *Annals of Botany* 68: 365-376.
- 13. Kittas, C., N. Katsoulas, and A. Baille. 2001. Influence of greenhouse ventilation regime on the microclimate and energy partitioning of a rose canopy during summer conditions. *J. Agricultural Engineering Research* 79(3): 349-360.
- 14. Nienaber, J. A., G. L. Hahn, et al. 1993. Eating behavior of swine as influenced by environmental temperature. *Proc. of the Fourth International Symposium of Livestock Environment*, 937-944.
- 15. Ogilvie, J. R. 1993. Operant response of weanling pigs to temperature and air speed. *Proc. of Fourth International Symposium of Livestock Environment*, 779-787.
- 16. Stanghellini, C. and W. T. M. Van Meurs. 1992. Environmental control of greenhouse crop transpiration. *J. Agricultural Engineering Research* 51: 297-311.
- 17. Yang, X., K. M. Ducharme, R. J. McAvoy, G. Elliott, and D. R. Miller. 1995. Effect of aerial conditions on heat and mass exchange between plants and air in greenhouses. *Trans. ASAE* 38(1): 225-229.
- 18. Berckmans, D., and M. De Moor. 1994. Linking a physical model to identification to model the air flow pattern and the 3-D energy and mass transfer in a ventilated

- space. Proc. of Fourth Int. Conference on Air Distribution in Rooms—ROOMVENT '94.
- 19. Carpenter, G. A. 1974. Ventilation of buildings for intensively housed livestock. *Heat Loss from Animals and Man*, J. L. Monteith, and L. E. Mount, 389-403. London, UK: Butterworths.
- Randall, J. M. 1981. Ventilation systems design. Environmental Aspects of Housing for Animal Production, ed. J. A. Clark, 351-369. London, UK: Butterworths.
- 21. Randall J. M., and V. A. Battams. 1979. Stability criteria for airflow patterns in livestock buildings. *J. Agricultural Engineering Research* 24(4): 361-374.
- 22. Berckmans D., and V. Goedseels. 1986. Development of new control techniques for the ventilation and heating of livestock buildings. *J. Agricultural Engineering Research* 33: 1-12.
- 23. Berckmans D., E. Vranken, R. Geers, and V. Goedseels. 1988. Efficiency of climate control equipment in pig houses. *Farm Building Progress* 93: 15-22.
- 24. Geers, R., V. Goedseels, D. Berckmans, and W. Huybrechts. 1984. Mortality, feed efficiency and carcass value of growing pigs in relation to environmental engineering and control; A case study of Belgian contract farming. *Livestock Production Science* 11: 235-241.
- 25. Maes, F., R. Geers, E. Vranken, V. Goedseels, and D. Berckmans. 1988. Environment control for growing finishing. *Pig International*, 6-8.
- 26. Carr, L. E., and L. W. Douglas. 1990. Empirical models to determine ammonia concentrations from broiler chicken litter. *Trans. ASAE*: 1337-1342.
- 27. Van Breemen, N., P. A. Burrouhg, E. J. Velthorst, H. F. Van Dobben, T. De Wit, T. B. Ridder, and H. F. R. Reijnders. 1982. Soil acidification from atmospheric ammonium sulphate in forest canopy throughfall. *Nature* 299: 548-550.
- 28. Ni, J.-Q., D. Berckmans, E. Vranken, and J. Coengrachts. 1994. A simulation model to evaluate ammonia emission from pig house. *Proc. of European Simulation Symposium ESS 94*, eds. A. R. Kaylan, et al., 13-15.
- 29. Vranken, E., and D. Berckmans. 1998. Control of the ventilation rate in agricultural buildings by using a newly designed ventilation rate sensor in the controller feed-back loop. *Computers in Agriculture*, 829-841.
- 30. Chen, Q., A. Moser, and P. Suter. 1992. A numerical study of indoor air quality and thermal comfort under six kinds of air diffusion. *Trans. ASHRAE* 98(1): 203-217.
- 31. Heber, A. J., C. R. Boon, and M. W. Peugh. 1996. Air patterns and turbulence in experimental livestock building. *J. Agricultural Engineering Research* 64(3): 209-226.
- 32. Gosman, A. D., P. V. Nielsen, A. Restivo, and J. H. Whitelaw. 1980. The flow properties of rooms with small ventilation opening. *Trans. ASME* 102: 316-323.
- 33. Chow, W. K., and W. Y. Fung.. 1996. Numerical studies on the indoor airflow in the occupied zone of ventilated and air-conditioned space. *Building and Environment* 31(4): 319-344.

- 34. Harral, B. B., and C. R. Boon. 1997. Comparison of predicted and measured airflow patterns in a mechanically ventilated livestock building without animals. *J. Agricultural Engineering Research* 66(3): 221-228.
- 35. Posner, J. D., C. R. Buchanan, D. Dunn-Rankin. 2003. Measurement and prediction of indoor airflow in a model room. *Energy and Buildings* 35(5): 515-526.
- 36. Janssens, K., D. Berckmans, and M. De Moor. 1996. Modelling the three-dimensional air flow pattern in a ventilated agricultural building. *Proc. of ROOMVENT '96, Fifth International Conference on Air Distribution in Rooms*, ed. S. Murakami, 441-442.
- 37. Van Brecht, A., K. Janssens, D. Berckmans, and E. Vranken. 2000. Image processing to quantify the trajectory of a visualized air jet. *J. Agricultural Engineering Research* 76(1): 91-100.
- 38. Berckmans D., K. Van de Weyer, and M. De Moor. 1993. Visualization and quantification of the airflow pattern using image analysis. *Proc. of the International Conference on Building Design, Technology and Occupant Wellbeing in Temperate Climates*, eds. E. Sterling, C. Bieva, and C. Collett, 207-216.
- Zhang, G. Q., J. Strom, and S. Morsing. 1996. Jet drop models for control of air jet trajectories in ventilated livestock buildings. *Proc. of Roomvent 1996, Fifth International Conference on Air Distribution in Rooms*, ed. S. Murakami, 305-316.
- 40. Bjerg, B., S. Morsing, K. S. Svidt, and G. K. Zhang. 1999. Three-dimensional airflow in a livestock test room with two dimensional boundary conditions. *J. Agricultural Engineering Research* 74(3): 267-274.
- 41. Koestel, A. 1955. Paths of horizontally projected heated and chilled air jets. *Trans. ASAE* 61: 213-232.
- 42. Leonard, J. J., and J. B. McQuitty. 1985. Criteria for control of cold ventilation air jets. ASAE Paper No. 85-4014. St. Joseph, MI: ASAE.
- 43. Zhang, G. Q., J. S. Strøm, and S. Morsing. 1992. Wall inlets for winter ventilation control. ASAE Paper No. 92-4531. St. Joseph, MI: ASAE.
- 44. Murakami, S., and S. Kato. 1989. Current status of numerical and experimental methods for analyzing flow field and diffusion field in a room. *Building Systems: Room Air and Air Contaminant Distribution*, ed. L. L. Christianson, 36-56.
- 45. Chen, Q. 1995. Comparison of different k-ε models for indoor airflow computations. *Numerical Heat Transfer*, *Part B* (28): 353-369.
- 46. Fan, Y. 1995. CFD modeling of the air and contaminant distribution in rooms. *Energy and Buildings* 23(1): 33-39.
- 47. Hong, T., and Y. Jiang. 1997. A new multizone model for the simulation of building thermal performance. *Building and Environment* 32(2): 123-128.
- 48. Dascalaki, E., M. Santamouris, M. Bruant, C. A. Balaras, A. Bossaer, D. Ducarme, and P. Wouters. 1999. Modeling large openings with COMIS. *Energy and Buildings* 30(1): 105-115.

- 49. Ren, Z., and J. Stewart. 2003. Simulating air flow and temperature distribution inside buildings using a modified version of COMIS with sub-zonal divisions. *Energy and Buildings* 35(3): 257-271.
- 50. Berckmans D., E. Vranken, and K. Janssens. 2001. Method and apparatus for determining a flow pattern of a fluid in a space. *International Patent WO* 01/01082.
- 51. Vranken, E., S. Schuysemans, A. Van Brecht, and D. Berckmans. 2002. On-line model based control of the airflow pattern in ventilated spaces. *Proc. of AgEng 2002, International Conference on Agricultural Engineering*.
- 52. Janssens, K., D. Berckmans, and A. Van Brecht. 2000. Modelling the spatiotemporal temperature distribution in an imperfectly mixed ventilated room, *Proc.* of Roomvent 2000. 7<sup>th</sup> Int. Conference on Air Distribution in Rooms, 223-228.
- 53. Aerts, J.-M., D. Berckmans, P. Saevels, E. Decuypere, and J. Buyse. 2000. Modelling the static and dynamic response of total heat production of broiler chickens to step changes in air temperature and light intensity. *British Poultry Science* 41: 651-659.
- 54. Aerts J.-M., M. Lippens, G. De Groote, J. Buyse, E. Decuypere, E. Vranken, and D. Berckmans. 2003a. Recursive prediction of broiler growth response to feed intake by using a time-variant parameter estimation method. *Poultry Science* 82(1): 40-49.
- 55. Aerts, J.-M., C. M. Wathes, and D. Berckmans. 2003b. Dynamic data-based modeling of heat production and growth of broiler chickens: development of an integrated management system. *Biosystems Engineering* 84: 257-266.
- 56. Leroy, T., J. Ceunen, E. Vranken, E. Struelens, F. Tuyttens, B. Sonck, A. Janssen, K. De Baere, J. Zoons, and D. Berckmans. 2003. Developing a quantitative computer vision method for on-line classification of poultry behavior in furnished cages. ASAE paper No. 034006. St. Joseph, MI.: ASAE.

# 6.4 Information Technologies in Water Management

F. S. Zazueta, J. Xin, L. S. Pereira, and A. Musy

**Abstract.** An overview on the use of information technology in agricultural water management is presented. This includes examples of past uses of databases, software and modeling, expert systems, real-time expert systems, and fuzzy logic, knowledge delivery systems, emerging technologies, and large-scale water management systems.

Keywords. Computers, Agroinformatics, Agricultural software, Decision support.

#### 6.4.1 Introduction

Over the past twenty years, information technology (IT) has played an important role in transforming water management. Techniques used in the development, mobilization, conveyance, and distribution of hydraulic resources, on-field water use, and drainage evolved into complex, knowledge-intensive systems. IT affected strategic

direction, design, and day-to-day management of water systems. This section will focus on the use of on-field water management for agricultural and landscape production. Readers are advised to see the relevant chapters in Volume I of this CIGR Handbook series [1] for complementary information on processes modeled and the use of models referenced here.

As technology and knowledge about the interactions of water with plant, soil, and atmosphere increased, water management became increasingly sophisticated. What was once conceived to be a management system to control water and maintain a salt balance became a system that manipulates the physical, chemical, and biological properties of the environment where the plant is grown in order to increase production quantity and/or quality, and ultimately maximize the economic value of the enterprise in a sustainable manner [2]. This created the need to use IT in a diversity of forms and approaches from simple to sophisticated depending upon the orientation, size, and economic capabilities of the farm enterprises, as well as on the society where it will be applied.

#### 6.4.2 Technological Advances Affecting Water Management

The ability to manage water has been influenced by advances in hardware, software, their related core technologies, and improved basic and practical knowledge of agricultural production systems. The outcome of these advances can be summarized into four words: *faster*, *connected*, *smarter*, and *cheaper* [3]. Advances in technology in several areas that affected water management strategies and tactics include:

- sensors and control;
- database technologies;
- design and management software;
- expert systems, real-time expert systems and fuzzy logic; and
- knowledge delivery systems.

Developments in each of the above areas occurred concurrently evolving in complexity to the contemporary systems in use today, and more importantly in the integration of several technologies addressing particular water management problems.

During the late 70s and early 1980s, the appearance of mass-produced microprocessors led to the development of the microcomputer and dedicated processors. Almost immediately, applications of the technology could be found in field [4], landscape [5], and greenhouse production systems. Low-cost closed-loop control systems using microcomputers, power-line modulation and soil-moisture sensors were used in greenhouse and field production [6]. During this decade, numerous manufacturers produced specialized closed-loop computer-based controllers that quickly became adopted, particularly for greenhouses and pressurized irrigation systems. Over time, environmental sensors that included atmospheric, soil, and plant response were used to assess and control parameters such as radiation intensity, air temperature, air relative humidity, soil moisture, and nutritional properties of the water solution [7,8].

In field applications, closed-loop control systems making direct measurements of soil or plant conditions have not been widely used in commercial applications. Instead, water balance methodologies based on indirect estimations of crop water use seem to be prevalent. This is probably because of the difficulties associated with point measurements in a heterogeneous system, reliability, and cost of soil moisture or plant state sensors. Nevertheless, modest water budget methods evolved from pencil and pad [9] through spreadsheets, dedicated software, and local databases [10,11] into systems that acquire information from Internet-based weather station networks, implement sophisticated knowledge-based algorithms and apply water based on precision agriculture principles. However, this evolution is not evident everywhere and is only true for large commercial farms in high-technology countries or regions. For large areas around the world, applications of those water-balance models with local data still are innovative and constitute appropriate technologies that generate substantial improvements in water use and conservation field practice.

In addition to improvements in sensors, actuators, computer and networking hardware, and software, there were major advances in our understanding of cropped systems that led to improved tools for water management. Also, advances in mathematics and computer science have resulted in applications ranging from simple stand-alone software to knowledge-intensive software such as real-time expert systems [12], rule-based expert systems, neural networks [13], and modeling in a water management context. This understanding is reflected in the current biological simulation models [14] and water quality process modeling in the root zone [15,16]. These models are often used in conjunction with geographic information system (GIS) and global positioning system (GPS) technologies not only for day-to-day management of water in precision agriculture but also used for planning purposes that are aggregated from the field level into module, district, and watershed levels.

As networking technologies became ubiquitous, the ability to share information anywhere anytime became closer to reality. By the mid-1990s World Wide Web resources provided up-to-date information about best management practices, and real-time data resources made accurate on-time information available allowing increases in effectiveness and efficiency in using water, particularly for potentially catastrophic events, such as freezes and droughts.

By the late 1990s and early 2000s, improved understanding of global climate led to development of models capable of predicting climatic conditions based on sea surface temperatures. This resulted in a widening of the scope of the use in IT in water management [17,18]. Namely, while on-field water management had been essentially tactical, climate prediction augmented managerial capabilities to include strategic planning in water management. However, developments in climate prediction are slow and they are only available for limited areas such as those where the El Niño southern oscillation (ENSO) has a strong footprint.

#### 6.4.3 Database Technologies

Water resource management is increasingly dependent on databases and the infrastructure connecting them. *Database management systems (DBMS)*, which provide the data storage and retrieval technology, have become an indispensable tool for analyzing and managing water resource data. DBMS are excellent tools used in storing and retrieving data collected from hydrologic and weather stations such as those used for flood and drought monitoring and watch systems. These data may feed models directly and the information produced can be focused on various applications such as dam storage, reservoir management, and irrigation systems conveyance and delivery. Example tools for drought mitigation are given by Rossi et al. [19].

The Internet has made profound impacts on database applications. DBMS and Internet technology provide the capability for web-based client/server applications and water resources data can be made readily accessible through the Internet. Today, three types of databases, *relational*, *object-oriented*, and *spatial* databases, are commonly used.

#### Relational Database Management Systems (RDBMS)

Relational DBMS (RDBMS), such as SQL Server, Oracle, and IBM DB2, were designed to store relational data with strict rules concerning their relations. They are the most widely used DBMS today. Numerous RDBMS applications have been developed in the area of water resource management. For example, the ARS water database [20] is a collection of precipitation and stream flow data from small agricultural watersheds in the United States. AQUASTAT [21] is a global information system of water and agriculture. The objective of the system is to provide users with comprehensive information on the state of agricultural water management across the world. MetBroker [22] is a heterogeneous database that provides weather information in various locations worldwide through web services. The readily available weather data can be used for irrigation management and agricultural simulation models.

# Object-Oriented Database Management Systems (OODBMS)

RDBMS have been very successful in meeting the needs of today's many applications of IT to water use and conservation. However, RDBMS are cumbersome when dealing with complex data models. *Object-oriented database management systems (OODBMS)*, such as ObjectStore, GemStone, Jasmine, and Objectivity, have the strength of storing and retrieving objects rather than relations and tables. The increased emphasis on process integration is a driving force for the adoption of OODBMS. Typically, applications of OODBMS have the characteristic of managing complex and highly interrelated information. For example, Extension Digital Information Source (EDIS) [23] uses OODBMS to manage over 5,000 agricultural extension publications. Few object-based applications exist in water management. However, their number is likely to increase as the level of competence of professionals in water management increases in this area, and as relational models become unmanageable as attempts are made to model complex systems.

#### Spatial Databases

For water resource assessment and management, there is a need to manage geometric, geographic, spatial, and temporal data. *Spatial databases* are the underlying database technology for GIS such as ArcIMS and IDRISI. As GISs offer powerful tools for the collection, storage, management, and display of spatial and temporal data, they play a unique role for water resource management by analyzing multiple forms of spatial data such as water distribution, soil, crop, land use, and weather. A variety of ex-

ample applications are included in Musy et al. [24] and Crausaz and Musy [25]. Integration of RDBMS and GIS is also commonly used to solve agricultural problems. Various applications that are aimed at watershed and basin management, water quality, and agricultural decision support systems were developed and are available from many sources for landscape and commercial agriculture design. The extended use of spatial databases in agricultural water management is limited by basic data availability, such as soil and soil water data. However, several approximations may be used to overcome problems but they require testing or calibration [26].

#### 6.4.4 Software and Modeling

Design and management software were amongst the early applications of computers in water management. In the early 1980s software focused on specific tasks for design, scheduling, and maintenance. This included applications such as pipe diameter calculations, irrigation scheduling, chemical injection rate estimations, and computer control of irrigation systems. A collection of irrigation and hydrology-related software is maintained in the IRRISOFT database [27].

In the area of computer-aided management of on-field water resources, systems evolved that are comprehensive production management systems, which include irrigation scheduling and operation. Spike [28] described the development and application of a phenology-based citrus management system involving multiple aspects of commercial citrus production.

Use of dynamic modeling techniques in water management reaches back to the early applications of computers. Early applications of modeling focused on a component of a system. Models that describe saturated and unsaturated flows in ground water, free surface water flow in canals, and irrigation are abundant in the literature.

The modeling of irrigation canal systems advanced rapidly in the past two decades due to hardware improvements, particularly sensors and controllers [29,30]. Unsteady-state models were reviewed by Ritter [31]. These models are capable of simulating flow rate and depth variation with time [32]. However, using such models for canal network management is difficult and time consuming. Models based on simplified hydraulic descriptions for water delivery management were described by Skutsch [33]. More recently, models for optimized management using GIS tools or including economic considerations to support delivery decisions [34] were developed. Models coupling hydraulics and hydrologic features were adopted in many parts of the world [35,36], including to plan maintenance in conjunction with deliveries [37,38]. In the case of pressurized irrigation systems, several approaches are used to optimize pipe sizes and system layout. Models that are able to simulate the functioning of the systems and assess the respective performance [39] are becoming popular, including their use for design purposes [40].

Water balance models evolved into effective water management tools mostly as a result of improvements in crop water requirements estimation and information management tools such as DBMS and GIS. These models are used at different scales, such as a subunit, a farm, an irrigation district and watershed, or an irrigation system, namely when operating in GIS, through the Internet or other approaches. At the farm

level, models for irrigation scheduling are also used to simulate crop water relations, including yield [41,42]. They are also used in real time by adopting some kind of weather forecasts [43]. The water flux models [44] tend to become operative essentially for simulating the transport of solutes in the soil media, as for salinity management as in the SWAP model [45] or the more advanced non-uniform transient flow model HYSWASOR [46]. A new generation of powerful models, which are able to simulate water, chemical processes, plant uptake, growth and yield, constitute an essential part of today's modeling aimed at environmentally conscious crop and water management [47]. Among others there is the root zone water quality model RZWQM [48,49], which has very good potential for assisting in soil characterization, mainly for the selection of agronomic and irrigation management practices in irrigated agriculture [50].

Another group of farm models is related to irrigation system design. Several models are currently used for evaluation and design of surface irrigation systems, such as SIRMOD [51] and SRFR [52]. The design of sprinkler and micro-irrigation systems evolved from the typical hydraulic calculations to systems where irrigation performance level is targeted [53]. As technology evolved and CAD tools increased in capabilities, specialized software was developed. As a result, today's irrigation software design industry is mature and there are numerous offerings in the market that include features such as plant databases, 3-D modeling, intelligent distribution of components, and materials pull-down lists. A simple Internet search on "irrigation software" yields numerous hits. However, such models have a limited use at locations where data requirements cannot be met.

Drainage models are also numerous and diverse [54]. They may be used for design or to assess functioning of the systems. Many of these have evolved over time to include other components. Because drainage is often associated with salinity and nitrate management, these components are often added to such models [55]. DRAINMOD [56] simulates the hydrology of poorly drained soils at a field and watershed level. The model was modified to include components that model fate and transport of nitrogen and salinity [57-61]. In addition the model was linked to other models such as the DUFLOW canal and water quality model [62]. BASINS was developed for watershed and water quality studies, including point and non-point sources [63]. Also, water model components were used as sub-models in comprehensive crop models [64].

#### 6.4.5 Expert Systems, Real-Time Expert Systems, and Fuzzy Logic

Expert systems (ES) are a type of artificial intelligence that has been successfully applied to water management. ES use forward chaining or backward chaining techniques to match an outcome to a set of conditions normally connected by IF... THEN... rules. The main reason for using real-time ES is to reduce cognitive load on users. To reduce cost and efforts in developing ES, most ES applications are currently implemented using shells such as JESS [65], CLIPS [66], OPS5 and variants, although AI languages like LISP and PROLOG are occasionally used.

ES may be used to solve a wide range of problems in the domain of water resource planning and management. Numerous rule-based ES were developed during the last

decade. These included comprehensive management systems such as GOSSYM/COMAX [67] for cotton; Neper Wheat [68] for irrigated wheat management in Egypt; CORMIX [69], a rule-based ES for the analysis, prediction, and design of discharges into watercourses or the atmosphere. Others addressed very specific problems in water management such as biological clogging [70].

Real-time ES is less frequently implemented in the field. An example of a near real-time model is SIMSEM [71], which is used in an irrigation system in southern France to produce short-term information on planting dates and couples a simulation model with ES. These systems are characterized by being driven by real-time events and having a time-limited response requirement. CIMS [72] is a real-time ES that uses real-time soil moisture data and weather data for irrigation control and management.

Fuzzy logic is a superset of conventional Boolean logic that is a way of processing data by allowing partial set membership rather than crisp set membership or non-membership. Fuzzy logic provides a simple way to achieve a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. A fuzzy ES uses fuzzy logic instead of Boolean logic. In other words, a fuzzy ES is a collection of membership functions and rules that are used to reason about data. Unlike conventional ES, which are mainly symbolic reasoning engines, fuzzy ES are oriented toward numerical processing. Such approaches can be applied to water basin or reservoir management. Rules and constraints can be expressed in a fuzzy way and the results of the simulation are estimated for a better allocation of water. Examples of the technique in water management include the fuzzy evapotranspiration model [73].

### 6.4.6 Knowledge Delivery Systems

One of the key roles that IT has played is in information delivery, particularly to extension audiences. During the late 1970s, early knowledge delivery systems (KDS) relied on terminal access to central mainframe computers for access to limited sets of information. During the 1980s, as database and networking technology evolved, KDS became richer in content and delivered through high-capacity media such as CD-ROMs and was also available on private networks. When the Internet became open to commerce and the WWW was developed, an explosion in access to information ensued during the mid 1990s. Today the main mechanism to deliver knowledge is the WWW, with varying degrees of sophistication. On the low end are static web pages, while on the high end are full authoring and publications management systems based on object technologies that generate HTML or XML content for web delivery as one of several products.

The Extension Digital Information Source (EDIS) maintains more than 5,000 extension publications in electronic format. It consists of a centralized OODBMS from which deliverables, such as CD-ROMs, HTML, and PDF, are generated [74]. Rather than using conventional word processing or markup tools, specialists are provided with a tool that creates an interface directly with the OODBMS. Also, management of the publications process is implemented through a web-based document routing system. In this way, content is made available immediately upon completion of the au-

thoring and review/approval process. Deliverables are generated synthetically on demand to ensure quality and timeliness of the information.

### 6.4.7 Emerging Technologies

Following the trend of miniaturization of computer equipment over the past few years the capacity of the notebook computer has tended to migrate into handheld devices. With the convergence of personal digital assistants (PDAs) with wireless, better software development tools, and a broad base of users skilled in using desktop and online resources, there is a clear trend towards nomadic computing.

Early uses of PDAs focused on scheduling, notetaking, email, and storage of information such as telephone numbers and addresses. Maturity of PDAs was reached when they were able to run applications comparable in complexity to those executed on a notebook computer and to store substantive amounts of information. The term *pocket computers* is sometimes used to describe the devices, reflecting this new generation product. Their compact size is convenient for extension agents who often move from location to location and from the office to the field.

Current experience in the deployment of handheld computers in extension demonstrated 82% penetration of the target audience one year after initiation of the program [75]. Success was due largely to the parallel development of handheld applications identified by the target audience.

#### 6.4.8 Large-Scale Water Management

Several implementations of large-scale systems have been possible due to modern information technologies. The combination of hardware, software, low cost, and trained technicians have resulted in systems that allow accurate irrigation management (see 6.4.4) while addressing other needs. Examples of production systems that integrate several technologies follow.

The Water Conserv II Water Reclamation Facility in Orlando, Florida integrates databases, modeling, and computer control [76]. In this system wastewater has been treated and used for citrus irrigation since 1984. The system delivers an average 25 mgd through a pressurized pipe network that is centrally controlled by one operator. A total of 75 growers covering 4,300 acres (1,740 ha) are serviced. Excess water is used for aquifer recharge by application to rapid infiltration basins. Control and feedback information is provided by a wireless system connected to a central computer.

Full-scale irrigation district management systems integrating database systems, data acquisition, modeling, GIS, and expert systems have been developed. SICODE [77] is a comprehensive management tool used for accounting, crop, water, and planting from permits to real-time water allocations. The system uses relational databases and the modeling component uses the DSSAT models for strategic crop and water allocations [78,79]. For tactical irrigation it implements a Penman-based water balance and real-time data from automated weather stations. Water allocations are made using a rule-based expert system and simulation results that can be displayed using GIS. The system was deployed over 15,000 ha with over 500 users. The SICODE databases are very comprehensive and contain users, soils, crop, agronomic recommendations, and other data

### 6.4.9 Where Are We Going?

Changes in technology will proceed with organizational and social change. The rapid evolution of hardware and software devices will increase consumer demand for timely and accurate information. If current trends in wireless broadband technologies continue, it is likely that, in the medium term, demands (beyond data) for on-field best management practices will increase rapidly. These developments, coupled with the rapid advances that are being made in human/computer interfaces, are likely to make IT ubiquitous, enabling access to dramatically increased amounts of data, tools for analysis, and decision support systems.

As technology evolves the perennial problem of creating classes of technology haves and technology have-nots seems to be exacerbated. This raises some important issues: How can the technology gap be reduced or closed? What mechanisms should be put in place to ensure that small farmers can benefit directly or indirectly from IT?

Clearly the solution to this problem is beyond the scope of this paper. Nevertheless, access to the technology is an important component of the solution, including issues related to affordability, viability, reliability, and support infrastructure. Finally, it is not the technology itself that is important, but rather how it is used to solve problems of water management and production in the cultural context where it is applied. Using IT in a beneficial way will require a transformation of practice that can only be achieved through a participative process that involves all stakeholders with a focus on education.

### References

- 1. van Lier H. N., L. S. Pereira, and F. R. Steiner, eds. 1999. *CIGR Handbook of Agricultural Engineering, Vol. I: Land and Water Engineering.* St. Joseph, MI: ASAE.
- 2. Zazueta, F. S. 1995. International developments in microirrigation. *Proc.* 5<sup>th</sup> Int. *Irrigation Congress*, 314-324. St. Joseph, MI: ASAE.
- 3. Gallaire, H. 1998. Chapter 4: Faster Connected and Amarter. *Technologies for the 21<sup>st</sup> Century*, 47-76. Paris, France: Organization for Economic Co-operation and Development (OECD).
- 4. Phene, C. J., and T. A. Howell. 1984. Soil sensor control of high-frequency irrigation systems. *Trans. ASAE*, 392-396.
- 5. Augustine, B. J., and G. H. Snyder. 1984. Moisture sensor-controlled irrigation for maintaining bermudagrass turf. *Agron. J.* 76: 848-850.
- 6. Zazueta F. S., A. G. Smajstrla, and D. S. Harrison. 1984. Microcomputer control of irrigation systems. I: Hardware and software considerations. *Soil and Crop Sci. Soc. Fla. Proc.* 44: 123-129.
- 7. Zazueta, F. S., and J. Xin. 1992. *Soil moisture sensors*. UF/IFAS EDIS Reviewed Extension Publications, EDIS document number EH226. Gainesville, FL: Florida Extension Service.
- 8. Mead, R. 1998. Soil moisture instrumentation: Sensors and strategies for the 21<sup>st</sup> century. *Irrigation J.*, Sept/Oct.

- 9. Choate, R. E., and D. S. Harrison. 1977. Irrigate by the accounting method. *IFAS Circular 431*. Gainesville, FL: University of Florida.
- 10. Pereira L. S., A. Perrier, M. Ait Kadi, and P. Kabat, eds. 1992. *Crop-Water Models*. Special issue of the ICID Bulletin. New Delhi, India: ICID.
- 11. Pereira L. S., B. van den Broek, P. Kabat, and R. G. Allen, eds. 1995. *Crop-Water Simulation Models in Practice*. Wageningen, The Netherlands: Wageningen Pers.
- Xin, J., F. S. Zazueta, A. G. Smajstrla, T. A. Wheaton, J. W. Jones, P. H. Jones, and D. D. Dankel II. 1997. CIMS: An Integrated real-time computer system for citrus microirrigation management. *Applied Engineering in Agriculture* 13(6): 785-790.
- Zhang, Q., S. Cetinkunt, T. Hwang, U. Pinsopon, M. A. Cobo, and R. G. Ingram. 2001. Use of adaptive control algorithms for automatic calibration of electrohydraulic actuator control. *Applied Engineering in Agriculture* 17(3): 259-265.
- 14. Jones, J. W., and J. T. Ritchie. 1990. The use of crop models in irrigation management. Monograph Chapter: *Management of Farm Irrigation Systems*, eds. G. J. Hoffman, T. Howell, and K. H. Solomon. St. Joseph, MI: ASAE.
- 15. Pereira L. S., and M. R. Cameira. 1999. Unsaturated zone leaching models: A user's view. *ICID J.* 48(2): 1-9.
- 16. Ma L., L. R. Ahuja, J. C. Ascough, M. J. Shaffer, K. W. Rojas, R. W. Malone, and M. R. Cameira. 2001. Integrating system modeling with field research in agriculture: Applications of the root zone water quality model (RZwQM). *Advances in Agronomy* 71: 233-292.
- 17. Hansen, J. W., A. Hodges, and J. W. Jones. 1998. ENSO influences on agriculture in the southeastern United States. *J. Climate* 11: 404-411.
- Royce, F. S., J. W. Jones, and J. W. Hansen. 2001. Model-based optimization of crop management for climate forecast applications. *Trans. ASAE* 44(5): 1319-1327
- 19. Rossi, G, A. Cancelliere, L. S. Pereira, T. Oweis, M. Shatanawi, and A. Zairi, eds. 2003. *Tools for Drought Mitigation in Mediterranean Regions*, WSTL 44. Dordrecht, The Netherlands: Kluwer.
- ARS. 2003. ARS Water Database. Washington, D.C.: United States Department of Agriculture. Accessed February 2003, from http://hydrolab.arsusda.gov/wdc/ arswater.html.
- 21. FAO. 2003. Review of world water resources by county, Water Report No. 23, Rome, United Nations. Accessed February 2003, from http://www.fao.org/waicent/faoinfo/agricult/agl/aglw/aquastat/main/index.stm.
- 22. Laurenson, M. R., A. Otuka, and S. Ninomiya. 2002. Developing agricultural models using MetBroker mediation software. *J. Agric. Meteorol.* 58: 1-9.
- 23. Beck, H. W. 2001. Agricultural enterprise information management using object databases, Java and CORBA, *Computers and Electronics in Agriculture* 32: 19-147.
- 24. Musy, A., L. S. Pereira, M. Fritsch, eds. 1999. *Emerging Technologies for Sustainable Land Use and Water Management*, Presses Polytechniques et

- Universitaires Romandes, Lausanne, 71 p. and CD-ROM.
- 25. Crausaz, P.-A., and A. Musy. 1997. GESREAU: un outil d'aide à la gestion des eaux par une modélisation du territoire. *Revue Internationale de Géomatique* 7(2): 127-139.
- Heinemann, A. B., G. Hoogenboom, and R. T. Faria. 2002. Determination of spatial water requirements at county and regional levels using crop models and GIS. An example for the State of Paraná, Brazil. *Agric. Water Manag.* 52: 177-196.
- 27. Stein, T. M. 2000. IRRISOFT Database on Irrigation and Hydrology Software. Accessed March 2003, from http://www.wiz.uni-kassel.de/kww/irrisoft/irrisoft i.html.
- 28. Spike, P. 1998. The decision information system for citrus. *Keynote Address*. 7<sup>th</sup> *Int. Conference on Computers in Agr*. St. Joseph, MI: ASAE.
- 29. Malaterre, P. O. 1995. Regulation of irrigation canals. *Irrigation and Drainage Systems* 9(4): 297-327.
- 30. FAO. 1994. Irrigation Water Delivery Models. *Proceedings of the FAO Expert Consultation*, Water Report No. 2.
- 31. Ritter, W. F., ed. 1991. Irrigation and Drainage. *Proc. Nat. ASCE Conference*. Reston, VA: ASCE.
- 32. Kumar P., A. Mishra, N. S. Raghuwanshi, and R. Singh. 2002. Application of unsteady flow hydraulic model to a large and complex irrigation system. *Agric. Water Manag.* 52: 49-66.
- 33. Skutsch J. C., ed. 1994. *Irrigation Water Delivery Models*. FAO Water Reports 2. Rome, Italy: FAO.
- 34. Reca, J., J. Roldán, M. Alcaide, R. López, and E. Camacho. 2001. Optimisation model for water allocation in deficit irrigation systems. Part II: Application to the Bambézar irrigation system. *Agric. Water Manage* 48: 117-132.
- 35. Singh R., J. C. Refsgaard, L. Yde, G. H. Jorgensen, and M. Thorsen. 1997. Hydraulic-hydrological simulations of canal-command for irrigation water management. *Irrig. Drain. Syst.* 11(3): 185-213.
- 36. Santhi, C., and N. V. Pundarikanthan. 2000. A new planning model for canal scheduling of rotational irrigation. *Agric. Water Manag.* 43: 327-343.
- 37. Nawazbhutta M., B. A. Shahid, and E. J. van der Velde. 1996. Hydraulic model to prioritize secondary canal maintenance inputs: results from Punjab, Pakistan. *Irrig. Drain. Syst.* 10(4): 377-392.
- 38. van Waijjen E. G., W. W. H. Hart, M. Kuper, and R. Brower. 1997. Using a hydro-dynamic model to plan maintenance activities and improve irrigation water distribution: application to the Fordwah distributary in Onjab, Pakistan. *Irrig. Drain. Syst.* 11(4): 367-386.
- 39. Lamaddalena, N., and J. A. Sagardoy. 2000. Performance analysis of on-demand pressurized irrigation systems. FAO Irrigation and Drainage Paper 59. Rome, Italy: FAO.
- 40. Pereira, L. S. 2003. Policies for Water Savings in the Yellow River Basin: A DSS Applied to Ningxia and Shandong. INCO-DC Project n ERBIC18CT970170.

- Final Report Papers and Publications (CD-ROM).
- 41. Pereira, L. S., A. Terrier, M. Ait Kadi, and P. Kabat, eds. 1992. *Crop-Water Models*. Special issue of the ICID Bulletin, New Delhi, India: ICID.
- 42. Pereira, L. S., B. van den Broek, P. Kabat, and R. G. Allen, eds. 1995. *Crop-Water Simulation Models in Practice*. Wageningen, The Netherlands: Wageningen Press.
- 43. Gowing J. W., and C. J. Ejieji. 2001. Real-time scheduling of supplemental irrigation using a descision model and short-term weather forecasts. *Agric. Water Manag.* 47: 137-153.
- 44. Fernandez J. E., C. Slawinski, F. Moreno, R. T. Walczack, and M. Vanclooster. 2002. Simulating the fate of water in a soil-crop system of a semi-arid Mediterranean area with the WAVE 21 and the EURO\_ACCESS-II models. *Agri. Water Manag.* 56: 113-129.
- 45. van Dam J. C., J. Huygen, J. G. Wesseling, R. A. Feddes, P. Kabat, P. E. V. VanWalsum. 1997. SWAP Users Manual: Simulation of transport Processes in Soil—Water—Atmosphere—Plant environment. Wageningen, The Netherlands: Dept. of Water Resources, Wageningen Agric. University.
- 46. Homace M., R. A. Feddes, and C. Dirksen. 2002. Simulation of root water uptake. III. Non-uniform transient combined salinity and water stress. *Agric. Water Manag.* 57: 127-144.
- 47. Ma, L., and H. M. Selim. 1997. Physical non-equilibrium modeling approaches to solute transport in soils. *Advances in Agronomy* 58: 95-150.
- 48. USDA. 2006. RZWQM Root Zone Water Quality Model. Accessed February 2006, from http://gpsr.ars.usda.gov/ products/rzwqm.htm.
- 49. Ahuja, L. R., K. W. Rojas, J. D. Hanson, M. J. Shaffer, and L. Ma, eds. 1999. Root Zone Water Quality Model, Modeling Management Effects on Water Quality and Crop Production. Highlands Ranch, CO: Water Resources Publications, LLC.
- 50. Cameira, M. R., L. Ahuja, R. M. Fernando, and L. S. Pereira. 2000. Evaluating field measured soil hydraulic properties in water transport simulations using the RZWQM. *J. Hydrology* (236)1-2: 78-90.
- 51. ISED. 1989. Surface Irrigation Simulation Software. User's Guide. Logan, UT: Irrigation Software Engineering Div., Dept. Agricultural and Irrigation Eng., Utah State University.
- 52. Strelkoff, T. 1993. SRFR, A Computer Program for Simulating Flow in Surface Irrigation Furrows-Basins-Borders. Phoenix, AZ: USDA-ARS Water Conservation Lab.
- 53. Pedras, C. M. G., and L. S. Pereira. 2001. A simulation model for design and evaluation of microirrigation systems. *Irrigation and Drainage* 50(4): 323-334.
- 54. Zimmer, D. (guest editor) 1995. Towards a better applicability of drainage models. *Irrigation and Drainage Systems* 9(3).
- 55. Kupper, E., E. P. Querner, J. A. Morabito, and M. Menenti. 2002. Using the SIMGRO regional hydrological model to evaluate salinity control measures in an irrigation area. *Agric. Water Manag.* 56: 1-15.

- Skaggs, R. W. 1980. DRAINMOD: Methods for design of drainage-water management systems for soils with high water tables. Ft. Worth, TX: United States Department of Agriculture, Soil Conservation Service, South National Technical Center.
- 57. Konyha, K. D., and R. W. Skaggs. 1992. A coupled field hydrology—Open channel flow model: Theory. *Trans. ASAE* 35(5): 1431-1440.
- 58. McCarthy, E., J. W. Flewelling, and R. W. Skaggs. 1992. A hydrologic model for a drained forested watershed. *J. Irrigation and Drainage Engineering* 118(2): 242-255.
- Kandil, H. M., R. W. Skaggs, S. Abdel-Dayem, and Y. Aiad. 1993. DRAINMOD-S: Water management model for irrigated lands, crop yield and application. Subsurface Drainage Simulation Models, Transactions of Workshop, 15th Congress ICID, ed. E. Lorre, 257-275.
- 60. Amatya, D. M. 1993. Hydrologic modeling of drained forested lands. PhD thesis. Raleigh, NC: North Carolina State University.
- 61. Breve, M. A. 1994. Modeling movement and fate of nitrogen in poorly drained soils. PhD thesis. Raleigh, NC: North Carolina State University.
- 62. Fernandez, G. P., G. M. Chescheir, R. W. Skaggs, D. Amatya, and F. Birgand. 1997. Development and field testing of a watershed scale DRAINMOD based hydrology and water quality model. St. Joseph, MI: ASAE.
- 63. U.S. Environmental Protection Agency (EPA). 2002. Better Assessment Science Integrating Point and Nonpiont Sources (BASINS). Accessed February 2003, from http://www.epa.gov/waterscience/basins/basinsv3.htm.
- 64. IBSNAT. 1992. Decision Support System for Agrotechnology Transfer. Version 3.5. Honolulu, HI: University of Hawaii.
- 65. Friedman-Hill, E. 2003. JESS. Sandia National Laboratories. Accessed February 2003, from http://herzberg.ca. sandia.gov/jess/.
- 66. Riley, G. 2002. CLIPS, NASA. Accessed February 2003, from http://www.ghg.net/clips/CLIPS.html.
- 67. Hoges, H. F., F. D. Whisler, S. M. Bridges, K. R. Reddy, and J. M. McKinion. 1998. Simulation in crop management: GOSSYM/COMAX. *Agricultural Systems Model and Simulation*, eds. R. M. Peart, and R. B. Curry, 235-281. New York, NY: M. Dekker, Inc.
- 68. Rafea. A., and J. Striklen. 1996. Neper Wheat. Accessed February 2003, from http://potato.claes.sci.eg/claes/ wheat.htm.
- 69. Jirka, G. H., R. L. Doneker, and T. O. Barnwell. 1991. CORMIX: A comprehensive expert system for mixing zone analysis of aqueous pollutant discharges. *Water Science and Technology* 24(6): 267-274.
- Zazueta, F. S., W. Huisman, and A. G. Smajstrla. 1996. An expert system for diagnosis and treatment of bacterial clogging in microirrigation. *Proc. FLAIRS*, 410-414. Key West, FL: FLAIRS.
- 71. Leenhardt D., and P. Lemaire. 2002. Estimating the spatial and temporal distribution of sowing dates for regional water management. *Agric. Water Manag.* 55: 37-52.

- Xin, J., F. S. Zazueta, A. G. Smajstrla, T. A. Wheaton, J. W. Jones, P. H. Jones, and D. D. Dankel II. 1997. CIMS: An integrated real-time computer system for citrus microirrigation management. *Applied Engineering in Agriculture* 13(6): 785-790.
- 73. Odhiambo, L. O., R. E. Yoder, D. C. Yoder, and J. W. Hines. 2001. Optimization of fuzzy evapotranspiration model through neural training with input-output examples. *Trans. ASAE* 44(6): 1635-1633.
- 74. Beck, H. W. 2002. Overview of Formal Languages for Ontologies. AOS Workshop. Gainesville, FL May 2002. Accessed February 2003, from http://fao.org/agris/aos/Workshops/ ThirdWorkshop/Beck/beckintro/.
- 75. Zazueta, F. S., P. Vergot III, H. W. Beck, and J. Xin. 2002. Use of personal digital assistants (PDAs) in agricultural extension programs. *Proc. of the Third Asian Conference on Information Technologies in Agri. Ed.*, 358-362. Beijing, China: Mei Fangquan, China Agricultural Scientech Press.
- 76. Cross, P. 2003. Welcome to water Conserv II. Accessed February 2003, from http://www.waterconservii.com/ index.html.
- 77. TecTra/DICARTU. 1985. Manual del Usuario. SICODE: Sistema Computacional Para la Distribucion Eficiente del Agua de Riego. TecTra/DICARTU. Guadalajara, México.
- 78. Roost, N., A. Musy, J. M. Lance, and Y. Cui. 2001. Framework of a prototype DSS for improved irrigation management in the Yellow River Basin. *International Conference on Agricultural Science and Technology, Vol. 6: Information Technology for Agriculture, ICAST*, eds. M. H. Wan, L. J. Han, T. W. Lei, and B. J. Wang, 406-412. Beijing, China: Ministry of Science and Technology.
- 79. Roost, N. 2003. Strategic options analysis in surface irrigation systems: Integrated modeling for efficient, productive and equitable water use. Thesis report No. 2641.

# 6.5 Geographical Information Systems

A. R. Mohamed Shariff

Abstract. Data analysis is an important aspect of agriculture, as it helps increase productivity. The geographical nature of agricultural information requires positional data to be accounted for, in addition to the attribute data. Spatial or geographical information systems are systems for collecting, storing, enhancing, and analyzing the spatial or geographical data sets. Thus, a GIS can be used to solve a multitude of position-related agricultural problems. This section provides an overview of GIS, followed by three specific examples of the use of GIS in agriculture: the distribution of disease or stressed plants in durian plantations, the distribution of bulk density and its relationship to paddy yield, and the determination of sea areas with high aggregation of chlorophyll-a that enables the identification of potential fishing zones.

**Keywords.** Geographical information systems (GIS), Spatial information systems (SIS), Agricultural GIS, Durian plantations, Precision farming, Remote sensing in agriculture, Potential fishing zones (PFZ).

### 6.5.1 Definition of a Geographical Information System

A geographical information system has the capability of collecting, storing, analyzing, and providing an output, with respect to geographically referenced data. The system need not be computer based. The formal term *GIS* gained prominence with the increase in computing power that made computing tasks involving large spatial datasets possible and practical. An agricultural GIS and field-level geographic information system, or FIS [1], is a system that can be used for assembling, storing, manipulating, and displaying spatially referenced agricultural data [2]. The term *spatial* is preferred to *geographical* in this definition because the environments involved in the agricultural sector are not confined to the simple geographic environment but include space within mills and agricultural processing factories, better described with the term *spatial*. However, the term *GIS* rather than *SIS* (Spatial Information System) is retained here for ease of understanding.

GIS has gained in popularity as it helps to solve a multitude of problems and serves a diverse audience. The UK government enquiry committee stated that the technology of GIS is as important to spatial analysis as the invention of the microscope and telescope were to science, the computer to economics, and the printing press to information dissemination [3]. In the agricultural arena, GIS is utilized, for example, with variable rate and yield monitoring technologies in the creation of soil nutrient maps, and in integrated pest management systems. In some of these systems, GIS is used in conjunction with other spatial technologies such as the Global Positioning System (GPS) and remote sensing.

#### 6.5.2 GIS Models

For a GIS to work, two basic types of information about the geographical entities need to be known: the geometric/location information of the object and the attribute or descriptive information that characterizes it. These two types of information are input into the GIS with a unique identifier that serves to link the location information with the attribute information. The uniqueness of this identifier must be ascertained in the planning stage of a GIS, as duplicate or non-unique identifiers will give inaccurate or confusing results.

There are two major types of GIS models, the vector and the raster model. A *vector GIS* represents real-world features with the basic geometries of a polygon, line, or point. For example, area features such as a plantation or a ranch can be represented by a polygon, a linear feature such as irrigation canal or river can be represented by a line, and wells or trees can be represented as points. As such, a vector GIS uses graphic data structure that represents the points, lines, and areas of geographical space by exact X and Y coordinates. On the other hand, a *raster GIS* uses graphic data structures that create discrete quantities of space that are grouped in a series of uniformly

shaped cells. A more recent development has been the integration of the vector and raster models into integrated systems [4].

### 6.5.3 Components of a GIS

The major components of a GIS are the:

- data input and verification subsystem,
- data storage and database management subsystem,
- data transformation and analysis subsystem, and
- data output and presentation subsystem.

#### The Data Input and Verification Subsystem

The GIS operates with the central processing unit (CPU) as its main engine for computation. The data input and verification subsystem is utilized in acquiring data and feeding them into the GIS. These data may be from available sources such as existing hardcopy maps, aerial photos, satellite imageries, or existing digital data obtainable from a multitude of agencies. However, if the data are not already available, then field surveys and measurements need to be carried out or commissioned to acquire these data. Data acquisition is the most time-consuming and expensive activity in the setting up of a GIS. Some new entrants to the GIS arena may be misguided in overlooking the complexity of the data requirements.

If data is available in hardcopy maps or in images, then it will need to be converted to digital form. This is achieved through the use of digitizing tables and pucks for the case of hardcopy maps or through the use of scanners in the case of aerial photography. The choice of digitizing tables and scanners depends on the resolution of the existing data and the intended accuracy of the database [5].

#### The Data Storage and Database Management Subsystem

Data is typically stored on hard disks, including portable types such as diskettes, CD-ROMs, DVDs, and flash drives. For long-term storage, as in the case of soils and yield data, the storage medium and environment should be considered. Experience of the 1980s from a pioneer computer-assisted land surveying system in Malaysia (CALS Johor) showed that data stored on tapes and kept in a controlled environment, over some years, may still be prone to fungus attacks. This problem can be overcome by recopying the saved files every few months onto different storage devices.

Database management systems are the software components of the storage system that handle the storage, integrity, and security of the datasets. Off-the-shelf software is available for this purpose and the current utilization of relational database models are still prevalent with the trend towards object-oriented designs (see, for example, Section 6.4.3).

### The Data Transformation and Analysis Subsystem

A software module that converts, for example, the data into different formats or different projections and coordinate systems carries out data transformation. It is particularly important as the current shift is towards positioning using global positioning systems (GPS) that use the World Geodetic System (WGS) of referencing while applica-

tion in agricultural holdings will most likely be using the local coordinate system. Proper transformation and adherence to a common reference system for a particular job is necessary to avoid errors in subsequent spatial analysis. The data transformation system also helps in the exchange of data as it allows a user to integrate data from multiple sources, which may have different referencing systems, into the user's working dataset.

A common approach to spatial analysis is through the use of the map overlay method. The overlay of spatial data in different layers allows for the geometric intersection process between objects in these layers to take place. This analysis is further enhanced with the ability to query the related attribute data of these geometries that are stored in relational data tables [6]. However, spatial analysis is not restricted to the overlay techniques as, for example, external statistical software can also be used to query and analyze the GIS dataset.

# The Data Output and Presentation Subsystem

Output from a GIS is typically shown on the computer monitor or as hardcopy from printers and plotters. For output that will be used in the field, considerations have to be given to the type of material on which the plotting will be done. Durable material is recommended for rugged field use.

### 6.5.4 GIS User Interfaces for Agricultural Applications

There are currently no GISs specifically designed for farmers with minimal computer skills. For practical in-field GIS applications, GIS designers need to emphasize GIS software that is easy to install, akin to plug and play, without any hassle of referring to the user manuals. Drop-down menus that have replaced command lines are positive user-interface developments for farmers, as are the pen-pointing devices that have replaced manual keyboards in palm-held devices such as PDAs. Issues of culture and language between the designers and users of the GIS software will also need to be addressed so that the technology can be best utilized without causing undue frustrations to the farmers. Although some headway has been achieved in user interfaces in general, there is a wide area for further development and improvements in designing the most appropriate user interfaces for farmers, particularly those who do not otherwise use computers. For literate farmers who already use new technologies in their work, there will also be gains from user-friendly technology.

#### 6.5.5 Examples of GIS Applications in the Agricultural Sector

GIS has a great flexibility in being used in a variety of agricultural applications. Three of these applications, involving the management of a durian plantation, planting of paddy, and prediction of potential fishing zones, are briefly described below.

### GIS for Durian Plantation Management

A GIS-based management system has been introduced for durian fruit plantations [7]. In this system, the geometry of the durian plantation is acquired from existing land registry documents while the location of physical features on the ground, such as rivers, roads, huts, durian trees, and water tanks, are surveyed and updated in the database. GPS is used for these surveys where it is practical. In situations when there is no

satellite signal under the tree canopy, conventional survey techniques with a laser-based distance measuring device such as Minimeter and tape have been used. Attribute data that characterize the durian plantation, such as diseases, pests, height of grass, grade of durian fruits, and fruit yield per tree, were collected. At the same time, different attributes were also identified for each feature. This geometric and attribute data was processed with GIS software.

A unique addressing system (unique identifier) for the trees had to be devised. In this case the plantation was divided into several regions and each region was given a unique alphabet. The trees in each region were then given a row number followed by their position in that row. Thus a tree in region W, situated in row number 5, and positioned 7<sup>th</sup> in the row is given a unique identifier of :W0507 where W is the unique region alphabet, 05 is the row number with a field width of 2 digits, and 07 is the position in the row with a field width of 2 digits. This identification was also painted on the tree to ease identification of the tree on the ground.

Upon completion of the database several relevant queries could be performed to aid the management of the plantation. Examples of these queries include the display of areas that have a high yield, areas that have a disease background, areas with high grass (knowledge of which will help alert the workers about the potential of encountering snakes and other natural dangers in these terrains), areas infected with aggressive ants (knowledge of which will enable the workers to take adequate protective steps prior to venturing into these regions).

The disease map is shown in Figure 1; areas in the southeast and north are seen to have a history of critical diseases. Figure 2 shows the yield map. The map legend differentiates between trees that have provided yield for more than 10 fruit seasons and trees that have yielded fruits between 5 and 10 fruit seasons. Interestingly the yield map also shows that areas in the southeast and north have a high yield. This is important information for the plantation management because a possible reason that the trees are giving a high yield could be due to the stress caused by the disease. Thus, the management could use this information to take appropriate remedial action before the trees become more critically diseased.

### GIS in Precision Farming for Increasing Rice Yield

GIS is an important component of precision farming and research, which can be used to increase crop yield. Roy et al. [8] mapped spatial variation within the paddy field using data of yield and soil parameters and then integrated it with positional information from GPS. The grid-point location map was produced from data recorded with a Leica GPS 500. Corrections were applied to this data using the base data received from the Survey and Mapping Department, Malaysia.

The GPS coordinates were originally recorded in the WGS84 coordinate reference frame, but these were transformed to the Malaysian Rectified Skew Orthomorphic (MRSO) coordinate system (the mapping system used in Malaysia). The attribute data of the physical soil parameters/properties (soil bulk density, cone index, and moisture) were recorded in a grid pattern after harvesting the paddy. These data were then input into a GIS database.

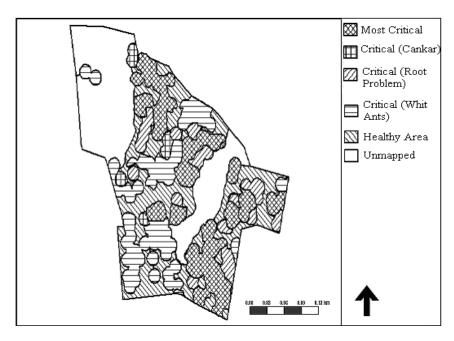


Figure 1. Map of overall disease.

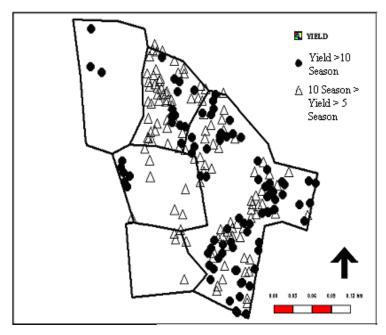


Figure 2. Yield map per durian tree.

Figures 3 to 5 show the results of the spatial analysis of the effects of the measured soil parameters on the paddy yield. The study revealed that the soil bulk density and

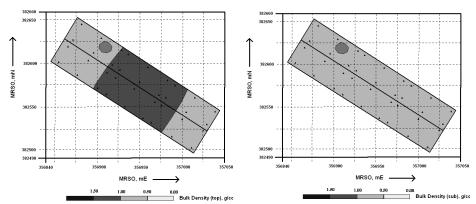


Figure 3. Spatial distribution of bulk density in topsoil (left) and subsoil (right).

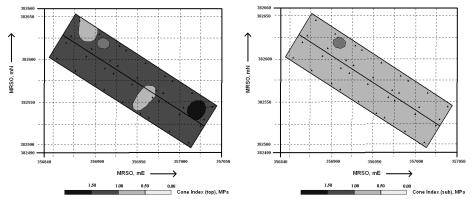


Figure 4. Spatial distribution of cone index in topsoil (left) and subsoil (right).

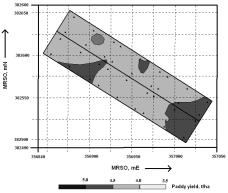


Figure 5. Spatial distribution of paddy yield.

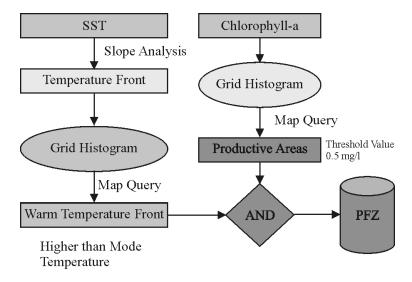


Figure 6. Flowchart for determining a potential fishing zone (PFZ).

cone index have a negative effect and soil moisture has a positive effect on the paddy yield. The spatial maps indicate that the middle portion of the plot needs improved soil loosening. From this, it could be recommended that adding more organic matter and effective rotation could help overcome these problems.

#### GIS Using Satellite Imagery Data as Input

The beneficial use of spatial imagery in agriculture for crop management has been known as early as 1929 when aerial photography was used to map soil resources [9]. Research on the use of GIS to create a satellite-based fish forecasting system [10] demonstrated the viability of remote sensing data as input into a GIS. This study analyzed the correlation between the surface phytoplankton distributions against the variation of the oceanographic conditions and showed that the upwelling area in east coast of Peninsular Malaysia exhibited higher chlorophyll-a concentration.

Figure 6 shows the flowchart for this project. Geometric data were collected from the satellite imageries of ADEOS/OCTS and SeaWIFS, and an oceanographic survey was carried out to collect further attribute data. Attribute data on the sea surface temperature (SST) were determined from the imagery and the monthly chlorophyll-a data were extracted from the ADEOS/OCTS Global Map Data Set and the SeaWIFS imagery. These data were then imported into GIS software for further analysis. These input parameters were later analyzed to produce the Potential Fishing Zone (PFZ) map.

The map in Figure 7 shows the sea surface temperature classification results while the map of the chlorophyll-a distribution is given in Figure 8. Finally, a map that can guide the fishermen to potentially rich fishing grounds is derived (Figure 9).

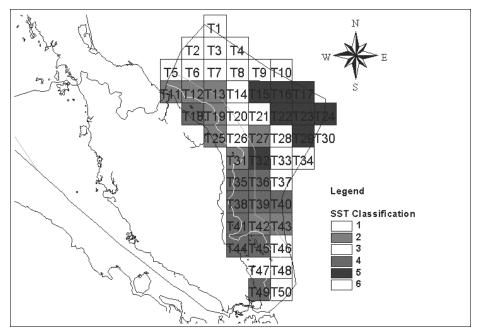


Figure 7. Map of SST classification result.

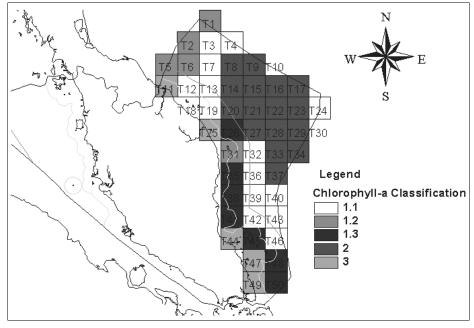


Figure 8 Map of the chlorophyll-a classification.

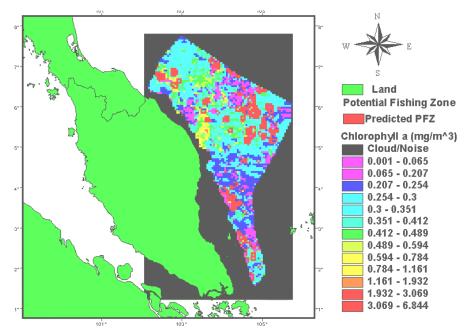


Figure 9. Potential fishing zone for the period 21-27 September 2001.

#### 6.5.6 Research Directions in Agricultural GIS

As more countries industrialize and progress, less manpower will be available for farms. Farmers in more and more countries will need to rely on automation and robotics for the optimum management of their farms. In these areas, GIS will play a prominent role in enabling many of the space-related tasks; for example, early versions of semi-automatic machines for the harvesting of oil palm fruit bunches are already being introduced in Malaysian palm oil plantations. GIS will be used in automation and robotics applications of farm tasks. For example, agricultural robots can be given tasks such as grass cutting, weeding, and tilling with simple instructions from the farmers in a simple natural language. These instructions will contain spatial terms that are easy to use for the farmers while at the same time the programming of these terms using a macro language for robot programming will be enhanced. Work in these directions is already in progress [11].

Due to anticipated lack of trained and skilled manpower, tasks such as determining ripeness of fruits or other products that are suitable for harvesting will also be automated. Image analysis based on the fruits' color or use of other sensors to determine parameters such as moisture content will be used to determine fruit ripeness.

An area of research that will receive great attention is the development of real-time mapping through the integration of GIS with GPS and inertial navigation systems (INS) and sensors to improve the practice of precision farming in the field. The explo-

sive growth of the wireless telecommunications market will take mobile computing to greater heights with the use of wireless agricultural GIS.

Improvements in the resolution of satellite images, aided by low-cost, high-accuracy GPS portable devices for establishing accurate location information and ground truth, will lead to easier and more effective monitoring of plant health, and thus more effective management of farms can be expected. This leap in the utilization of spatial technology will profoundly affect the running of farms. The biggest beneficiaries of these technologies will be farmers who are able to master these technologies while having a good sense and working knowledge of their farms.

#### References

- 1. Kluitenburg, M., and C. Redulla. 1999. Making GIS a versatile analytical tool for research in precision farming. *Computers and Electronics in Agriculture* 22: 221-231.
- Neményi, M., P. Á. Mesterházi, Z. Pecze, and Z. Stépán. 2000. The role of GIS and GPS in precision farming. *Computers and Electronics in Agriculture* 25: 107-120.
- 3. DoE. 1987. Handling geographic information: Report of the Committee of Enquiry, chaired by Lord Chorley. London, UK: HMSO.
- 4. Fonseca, F. T, M. J. Egenhofer, and P. Agouris, P. 2002. Using ontologies for integrated geographical information systems. *Trans.GIS* 6(3).
- 5. DeMers, M. N. 2003. *Fundamentals of Geographic Information Systems*, 3<sup>rd</sup> ed. New York, NY: John Wiley and Sons.
- 6. Chrisman, N. 2002. *Exploring Geographic Information Systems*, 2<sup>nd</sup> ed. New York, NY: John Wiley and Sons.
- Hin, T. C., A. R. M. Shariff, and S. K. Roy. 2002. GIS based durian plantation management system. *National Seminar on Precision Farming*, Kuala Lumpur, Malaysia.
- 8. Roy, S. K., A. R. M. Shariff, and D. Ahmad. 2003. Spatial variability of soil compaction and its effect on paddy yield. Technical Paper, Department of Biological and Agricultural Engineering, Universiti Putra Malaysia.
- 9. Santhosh, K. S., L. Soizik, G. M. Casady, and G. A. Seielstad. 2003. Remote sensing applications for precision agriculture: A learning community approach. *Remote Sensing of Environment* 88: 157-169.
- Knee, T. C., S. Mansor, M. H. Ibrahim, A. R. M. Shariff, and A. R. Mahmud. 2000. Studies of sea surface temperature and chlorophyll-a variation using remote sensing in east coast of peninsular Malaysia. *Proc. of the 4<sup>th</sup> Annual Seminar of Geoinformation Engineering*, Serdang, Malaysia.
- 11. Blackmore, S., and R. Shariff. 2005. Natural language semantics to describe agricultural robotic behaviours. *First Asian Precision Agricultural Conference*.

# 6.6 3-D Animation and Virtual Reality

F. Xiong, X. Zhao, and Y. Zhang

Abstract. Computer graphics has become a very active field of information science. With the adoption of many advanced technologies, computer graphics, especially 3-D animation and virtual reality, has bright prospects. In this section, we introduce some basic aspects, especially about the visual plant, which is a hotspot in agriculture informatics. First we introduce some concepts in 3-D animation and popular modeling technologies, rendering, and motion control, then a brief description of virtual reality. Finally we introduce some applications of 3-D animation and virtual reality in agriculture, especially achievements in virtual plants, including modeling methods and modeling tools for virtual plants.

Keywords. 3-D animation, Virtual reality, Virtual plant.

#### 6.6.1 Introduction

Computer graphics has become a very active field of information science. It is a practical technology of using computers to display, generate and process graphs or images. Computer graphics has been used in many aspects of society now, such as industry, military, scientific research, commerce, education, art, amusement, movies, TV, and even family life. With the adoption of many advanced technologies including multimedia, image processing, computer vision, artificial intelligence, networks, etc., and the convenience and low price of graphic devices and the emergence of novel software tools, computer graphics is becoming more and more popular. Virtual reality and 3-D animation have bright prospects.

#### 6.6.2 3-D Animation

Computer animation is an interesting technology that produces sequential motion graphs or images based on conventional animation. It has experienced two phases: 2-D animation and 3-D animation. 2-D animation, also called *computer-assisted animation*, can substitute for the repeated labor of traditional manual animation (e.g., drawing, rendering, copying, scaling, moving, rotating, and so on). Being based on image frames, it is hard for 2-D animation to interpolate between two keyframes automatically, and to give users a realistic and real-time feeling. 3-D animation is based on geometric models, composed of vertexes, lines, and planes with 3-D depth information. With the development of image and graphics devices, software and multimedia technology, and graphics theory, 3-D animation is bringing us more and more exciting applications.

The basic technologies of 3-D animation include modeling, rendering, and motion control. There are numerous software packages for 3-D animation.

#### **Modeling Technologies**

Modeling means how to apply suitable models to represent different graphic objects. Description of an object includes geometry information and topology information. Geometry information consists of geometry data such as size, shape, and position; topology information consists of frame information, which reflects how edges,

vertexes, and planes connect with each other. 3-D modeling mainly contains three basic technologies: polygon-based modeling, curved surface-based modeling, and parameter-based modeling. Polygon-based modeling is the simplest one. It regards an object as a set of polygons. Curved surface-based modeling uses quadric surfaces like cylinder, cone, sphere, ellipsoid, and spatial grids composed of controllable points to define the shape of a non-flat surface object. Parameter-based modeling can be described by the geometry and structure parameters of an object.

Natural objects can be classified into regular and irregular objects. For regular object, Euclidian geometry can be used to express its modeling, and this is called geometry modeling. However, Euclidian geometry can hardly be used to describe irregular objects. For most irregular objects procedural modeling methods, such as fractal dimension, shape grammars, and particle systems, are adopted.

The *fractal geometry approach* differs from the classical geometry approach which abstracts the object as a set of zero-dimension points, one-dimension lines, two-dimension planes, and three-dimension volumes. It uses the fractal dimension to describe an object. Fractal geometry does not adopt mathematic formulae but procedures to create modeling, and an object is expressed by infinite self-similarity, and the details between the whole and the parts are repeated infinitely. The generation procedure is on the basis of certain rules using recursive operating to produce the partial details. Fractal modeling technology is suitable to describe those self-similar but vague objects, like rocks, seashores, mountains, water, plants, etc. For application in agriculture, fractal modeling can describe the growth of trees, grass, crops, etc.

The *shape grammar approach* is also a procedural way to generate objects' details. It depends on a group of production rules to change or increase the substructure on the basis of the origin to form the final structure. L-System, one of the typical approaches, is an efficient way to describe plants. For example, a tree can be described as a trunk with stems and leaves, and under the guide of its growth rules, the stems and leaves can be connected in a special way, then the tree is simulated.

The particle system approach is another way to simulate irregular and vague objects. It can perfectly reflect the dynamic and random properties of a object, and it is good at describing objects such as water, fire, clouds, smog, forests, and lakes, which would change with time. Its main idea is to use a lot of micro particles with simple shapes to simulate objects. These particles might be spherules, ellipsoids, cubes, or other shapes. Under certain rules, the size, shape, transparency, and color of particles can be changed with time, and the motions of the particles are mainly simulated by stochastic processes. Every particle will experience three stages: birth, growth, and death.

#### Rendering and Motion Control

Rendering makes the resultant images true to life. It is one of the key functions in 3-D animation. Rendering is implemented by proper algorithms according to geometric shape, relative position, and surface features of the object. Raycasting and raytracing are the main algorithms where hidden faces, material, texture mapping, shading, shadowing, and lighting need to be considered.

The common hidden algorithms mainly include the painter algorithm, Z-buffer, back-face detection, and scan-line Z-buffer. Z-buffer is widely used due to its simplicity. For the reality of 3-D animation, the colors of facets are determined by many factors, such as their reflection and transmission characteristics, light model, normal and the view from observer, etc. The object's surface is often covered with certain material; it must be processed with texture mapping. Shading mainly contains Lambert modeling, Gourand modeling, Phong modeling, etc. Specular, ambient, and diffuse reflections can also affect the reality of rendered object.

*Motion control* is another key technology in 3-D animation. It involves keyframe animation, kinematics, dynamics, morphing animation, role animation, script animation, camera motion, etc, and how to control animation interactively.

Interpolation is often used in keyframe animation to generate the intermediate frames. Keyframe animation has two types: shape-based and parameter-based. It mainly produces animation effects by parameter change and motion track. The animations controlled by kinematics theory or by dynamics theory are called kinematic animation or dynamic animation, respectively. They are often used to describe the motion and acceleration of objects. Morphing animation is to transform the object for the effect of exaggeration. Role animation uses body language or action to personalize the animation. Script animation uses script language to describe the keyframe motion of objects.

### 3-D Animation Software Tools

A lot of software tools for 3-D animation have emerged in the market. Some popular ones are 3D Studio, 3D Studio Max, Maya, Softimage/3D, PowerAnimator, etc. Among them, 3D Studio Max, Maya, and Softimage/3D are fashionable animation software platforms for PCs or workstations. These tools have powerful functions and friendly interfaces so that users can easily create a 3-D model, set the motion track, and define the texture and light model of an object, etc. All of them have the ability of fast rendering and easily setting resolution, color depth, alpha channel and antialiasing for the resulting animation, and provide various animation formats and compression technologies for storage.

In addition, Maya provides a novel operating interface that enables users to utilize MEL (Maya embedded language) to automate the repeat operations. 3D Studio Max can support several popular 3-D graphic interfaces, such as OpenGL and Direct 3D, which can be accelerated by a 3-D display card to get remarkable performance. 3D Studio MAX also is a multithread software with symmetric multiprocessor (SMP) ability, which can run in a multiprocessor computer and do parallel rendering in a network.

### 6.6.3 Virtual Reality

Virtual reality (VR), using computers and related multimedia equipments (e.g., video, audio, and tactile devices), aims to create a virtual tri-dimensional world, in which users can see, hear, manipulate, and interact with the virtual environment and can even feel the force fed back from it. Virtual reality involves many high technologies such as computer graphics, image processing, pattern recognition, computer vi-

sion, intelligent control, simulation, automation, sensor, audio processing and detection, mechanism, optics, network, and parallel processing, etc.

In general, the hardware devices used in a VR system include:

- *image displays and vision tracking systems* involving conventional computer monitor, head-mounted display, various position trackers (mechanical armatures, ultrasonic sensors, magnetic trackers, optical position trackers, etc.);
- auditory generators and trackers including audio synthesizers, audio detection, sound source locators; and
- haptical and force feedback devices ranging from the conventional mouse, trackball or joystick to force balls, manipulators, instrumented gloves, force feedback systems, haptic feedback interaction systems, etc.

There are many VR hardware devices in the market. However, the performances of most existing devices, especially the high-speed image processors, have not yet met the needs of VR. Research on how to apply the human psychology and physiology into VR doesn't have a clear result. In addition, expensive hardware might be another barrier to its widespread application.

Most VR software can run in PC-level computers by adding high-performance accelerator cards, while some VR software needs a workstation for running. One example of VR software is from Dimension International, which has developed a PC edition toolkits named VRT. It is a powerful development environment for virtual reality software to create an immersion world. Other examples of VR software are: VEOS (Virtual Environment Operating Sell), developed by Human Interface Technology Laboratory; World Tool Kit (WTK), developed by Sense8 Corporation; Minimal Reality (MR), developed by the University of Alberta; Generic Visual System (GVS), developed by Gemini; SkyWriter and Reality Engine, by SGI; Division Provision, by Division Corporation; and PhotoVR, by Straylight Corp.

Another important development in VR is VRML (Virtual Reality Modeling Language), which is a computer language used to build interactive 3-D animation in a browser, just like HTML (Hypertext Markup Language) for hypertext. It allows for 3-D modeling, scene structure, animation, sound, and sensor functions. In order to produce 3-D modeling, VRML needs a browser or plug-in, which can be freely downloaded, to interpret.

### 6.6.4 Applications of 3-D Animation and Virtual Reality in Agriculture

Computer modeling and visualization for plants is a hotspot in agriculture informatics. The *virtual plant* is computer simulation of the structural development and growth process of a plant in 3-D space. As an organism, the growth process of a plant and the influence of its environment are so complicated that it is difficult to simulate with fidelity, but researchers have made some valuable progress in this field.

### Computer Modeling Methods for Virtual Plants

Several models for generating graphs or images of plants have been put forward by computer scientists, botanists, and mathematicians. Depending upon those models, not only virtual plants, but also virtual fields or virtual environments can be produced

conveniently. Among these models, the L-system, A-system, iteration function system (IFS), particle system, and automation are representatives.

### L-System

*L-system* was first introduced by Lindenmayer [1] in 1968, and then developed by Prusinkiewicz et al. for plant modeling. L-system essentially is a string rewriting system, which uses an axiom and a set of productions to generate a string by finite recursive operations, and this string can be interpreted as the structure of plant. L-system is suitable for describing morphologic architectures of plants.

The simple L-system is termed D0L-system, where "D0" stands for deterministic. D0L-system is context-free and uses a turtle interpretation algorithm to generate a fractal image. A turtle can move forward, backward, rightward, or leftward. The turtle state is defined with five simple symbols: I, +, -, [, and ]. The I symbol means one unit length of movement in any direction. The direction is determined by the + and – symbols. The + symbol means turn left (counter-clockwise) while the – symbol means turn right (clockwise). The symbols [ and ] mean a stack. The [ symbol pushes the current state of the turtle into the stack, while the ] symbol pops a state from the stack and assigns it as the current state of the turtle. For example, through the use of production  $S \rightarrow A[-B][C][+D][E]$ , the right part of Figure 1 gives the successor of iteration from the left part predecessor S.

Context-sensitive L-system was introduced for the interaction between adjacent plant parts. Context-free L-system is just the special case of the general context-sensitive L-system. In order to add randomness to the growth process, the stochastic L-system was proposed. It can produce different output every time since the growth is not predictable or reproducible with the same axiom and rules. Open L-system was proposed by Mech et al. to simulate the mutual influence between a plant and its environment.

For simulating continuous phenomena, *parametric L-system* was presented. Figure 2 illustrates the growth process and the bloom sequence of a plant. Prusinkiewicz et al. then proposed timed L-system and differential L-system (DL-system) as formal frameworks for constructing models of branching in continuous growth process.

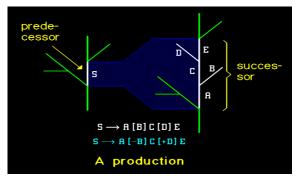


Figure 1. Example of L-system principle (by Prusinkiewicz).

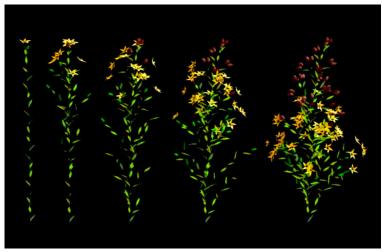


Figure 2. Growth process and bloom sequence of a plant (by Prusinkiewicz [2]).

L-system is powerful for building fractal objects, but in general users have difficulty in understanding it due to labor-intensive writing rules. Moreover, it is more complicated to simulate the growth of a plant with fidelity.

#### A-System

Honda [3] described the shape of a tree using the angle and the length of branching. Based on Honda's work, Anon [4] proposed another simulation model for plant shapes called *A-system* and four geometrical models: GMT1, GMT2, GMT3, and GMT4. Asystem can simulate visually different kind of plants by stochastically giving the angles of branches and the features of ramification structures, and can also simulate the effects of light, wind, and gravitation on the shape of a plant using attractor algorithms.

#### The Iteration Function System

The *iteration function system (IFS)*, introduced by Hutchinson in 1981, is also an effective method for plant simulation. Barnsley [5] and Demko adopted IFS to generate the fern leaf with exact self-similarity. Furthermore, Barnsley et al. proposed recurrent IFS, which could embody different self-similarity among parts of the plant. Their further work promoted IFS to develop quickly and made it become a convenient way to construct fractal geometry with any dimension. In 1991, Prusinkiewicz [6] and Hammel developed a language-restricted IFS, which generalized the original definition of the IFS by providing a means of restricting the sequences of applicable transformations.

### The Particle System

Using the *particle system*, Reeves [7] simulated sod and forest. In this stochastic model, the structured particle system was applied in describing the complicated natural scenery. When a tree is created, the trunk and the main branches are still generated by the traditional geometry model, but the small organs like leaves commonly use the particle system approach. So, it may not be suitable for the description of a whole plant. Figure 3 is a simulation of forest scenery by Reeves [7].



Figure 3. Forest scenery (by Reeves [7]).

#### Automation

De Reffye et al. [8] proposed *finite automation* as a method of plant morphogenesis modeling. The process of growth, dormancy, and death of plant is described by this finite automation using Markov chains and state transition graphs. A multiscale model of plant topological structures (MTG) was introduced by Godin et al. [9] based on the work of de Reffye. MTG can describe the topological structure using different time scales. Zhao et al. [10] gave a new method, named dual-scale automation, to simulate plant development. Dual-scale automation consists of macrostates and microstates. Through the combination and circulation of the macrostates and the microstates, the topological structure of plants is generated visually. It is built based on botany and agronomy, and can simulate plants growth more truly. Figure 4 is an example of the growth process of cotton simulated by this dual-scale automation.

Additionally, Oppenheimer [11] used a fractal method to simulate plants and trees in real time. Viennot [12] proposed a ramification matrix to model the structure of plant. He made use of this matrix to describe the number of ramified nodes and their

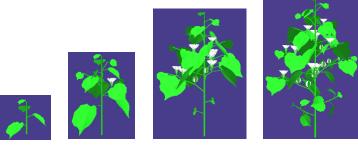


Figure 4. Simulated growth process of cotton showing 4 different periods [10].

relationship, and generated the fractal structure of plants through iteration. Greene [13] utilized voxel space to simulate the influence between the plant and its barrier, and modeled the stochastic growth processes of rambler in voxel space.

### Computer Modeling Tools for Virtual Plants

Virtual plants can be created by general programming languages such as Visual C, Delphi, Visual Basic, etc., according to plant growth information such as growth velocity. These languages have the ability of calling 3-D graphics interfaces like OpenGL or Direct3D. Another way to make virtual plants is to use 3-D animation platforms like 3D Studio Max, Maya, etc. They generally have powerful functions and friendly interfaces to help users create 3-D models or 3-D animations for many fields. However, these methods heavily depend on the dialogue between user and machine. It means the result will be too subjective and hardly faithful to the plant itself. In recent years, some software tools specially applied in simulating plants have appeared: Xfrog, SpeedTree, AMAP, GOSSYM/COMAX, CottonPlus, PlantStudio, Virtual Plants, Digimation-Tree Factory, and Tree Druid, etc. As examples, three tools are summarized below.

#### Xfrog

Xfrog plant modeling software was developed by Bernd Lintermann [14] from GreenWorks GbR (Germany). Many branching objects, including trees, bushes, etc., are generated conveniently using this software. It defines three groups of components to describe the structural and geometrical elements of plants. The first group of components creates graphical objects like stems, twigs, leaves, and other geometric primitives. The second group expresses the plant's structure, such as the definition of the branch orientation, the petal placement, etc. The third group denotes some global variables affecting the generated plant's overall shape. Each component offers a basic set of parameters that define a geometric primitive. Users may construct a plant image by clicking a graphical dialog, and have control over the modeling process.

Xfrog has many good qualities, including light and gravity tropisms and providing outputs to Rayshade, Maya, etc. However, it neither biologically simulates the growth and structure of plants, nor supports the vein structures of leaves, so it is difficult to generate actual 3-D leaves. Also, the trees generated by this software have inaccurate branching geometry, just intersecting cylinders instead of Y-junctions.

### Artic Monitoring and Assessment Programmer (AMAP)

Based on the birth and death of growing buds, De Reffye et al. [8] developed a procedural model that permits users to control the development of plants by some parameters. AMAP, developed by CIGR (France), provides the necessary methods for measuring, analyzing, and simulating the architecture, functioning, growth, and production of plants, crops, etc. AMAP has made great progress in the combination of growth mechanism models and visualization models of plants.

AMAP has several sub-models including AMAPmod, AMAPsim, AMAPpara, and AMAPhydro to provide different functions. AMAPmod model is mainly used to build development models of plant based on sufficient spatial information. According to the plant model built by AMAPmod, AMAPsim model simulates the growth process of

plants with 3-D visualization. AMAPpara model describes long-term plant growth as the cumulative output of the cyclic interaction between plant ecophysiological function and architectural development. In the model of AMAPhydro, a hydraulic model of a plant is built to simulate the production and distribution process of nutrition within the plant and the influence of plant architecture on its growth.

#### Plant Studio

Plant Studio was developed by Kurtz-Fernhout [15] in Germany. In this studio, users can generate many sorts of plant. The attributes of meristem, ramification, anthotaxis, leaf, internode, flower, and fruit of plant are predefined in the form of components by which users can easily assemble a virtual tree. The results of Plant Studio can be exported in many common exchange formats such as 3DS, DFX, WRL, OBJ, POV, LWO, etc., so they can be further processed in other 3-D tools.

#### 6.6.5 The Future of Virtual Plants

Though great progress in virtual plants has been made, it is far from simulating real plant growth, which needs not only the plant topological and geometrical model information, but also the eco-physiological model information involved in the complex factors of plant physiology and plant ecology and the influence of environment. It's believed that integration of diverse disciplines and technologies will promote the development and practical application of virtual plants.

Virtual plants will play important roles in research and agricultural development. For example, virtual plants could carry out virtual experiments to save much time and money. These virtual experiments will help explore the growth rules of plants and help farm managers make proper decisions. Virtual plants may also promote research in plant physiology, plant ecology, etc., and their development in agriculture. Moreover, virtual plants are significant in advanced technology and knowledge for agricultural technicians and farmers. It can be predicted that the progresses in virtual plants will play a growing role in agriculture informatics, and the performance of virtual fields, virtual landscapes, virtual farms, virtual bonsai, virtual environments, etc., will improve rapidly.

#### References

- 1. Lindenmayer, A. 1968. Mathematical models for cellular interaction in development, Parts I and II. *J. Theoretical Biology* 18: 280-315.
- 2. Prusinkiewicz, P., and A. Lindenmayer. 1990. *The Algorithmic Beauty of Plants*, 40-50. New York, NY: Springer Verlag.
- 3. Honda, H. 1971. Description of the form of trees by the parameters of the tree-like body: Effects of the branching angle and the branch length on the shape of the tree-like body. *J. Theoretical Biology* 31: 331-338.
- 4. Anon, M., and T. L. Kunii. 1984. Botanical tree image generation. *IEEE Computer Graphics and Applications* 4: 10-34.
- 5. Barnsley, M. F., and S. Demko. 1985. Iterated function systems and global construction of fractals. *Proc. Roy. Soc. London*, Vol. A 399, 243-275.

- 6. Prusinkiewicz, P., and M. Hammel. 1991. Automata, language, and iterated function systems. *Fractal Modeling in 3D Computer Graphics and Imagery*, ACM SIGGRAPH, Course Note C14, eds. J. C. Hart and F. K. Musgrave, 115-143.
- 7. Reeves, W. T., and R. Blau. 1985. Approximate and probabilistic algorithms for shading and rendering structured particle systems. *Computer Graphics* 19: 313-322.
- 8. De Reffye P., C. Edelin, J. Francon, M. Jaeger, and C. Puech. 1988. Plant models faithful to botanical structure and development. *Computer Graphics* 22: 151-158.
- 9. Godin, C., and Y. A. Carglio. 1998. A multiscale model of plant topological structures. *J. Theoretical Biology* 84(191): 1-46.
- 10. Zhao, X., P. de Reffye, F. Xiong, and B. Hu. 2001. The dual-scale automation model of virtual plant development. *J. Computers* Vol.24.
- 11. Oppenheimer, P. E. 1986. Real time design and animation of fractal plants and trees. *Computer Graphics* 20: 55-64.
- 12. Viennot, X. G., and G. Eyrolles. 1989. Combinatorial analysis of ramified patterns and computer imagery of trees. *Computer Graphics* 23: 31-39.
- 13. Greene, N. 1989. Voxel space automaton: Modeling with stochastic growth processes in voxel space. *Computer Graphics* 23: 175-184.
- 14. Lintermann, B., and O. Deussen. 1999. Interactive modeling of branching structures. *IEEE Computer Graphics & Applications* 19: 56-65.
- 15. www. Kurtz-Fernhout.com

# 7 Communication Issues and Internet Use

# 7.1 Dedicated Communication Systems and Standards for Agricultural Applications

H. Auernhammer and H. Speckmann

**Abstract.** Modern agricultural technology is controlled by electronics. Therefore, machines and implements represent intelligent process units with the ability to communicate internally, with other units, and with the management. A comprehensive and trouble-free usage of these new communication options is only possible when worldwide accepted communication standards are developed and used.

The beginning of standardized communication can be found in the point-to-point connection standards between tractor and implement, DIN 9684/1 and ISO 11786. They define how tractor sensor signals have to be provided for implement control in order to save additional sensors.

The possibilities of the Agricultural Bus System (LBS) conforming to DIN 9684/2-5 are much broader. Based on CAN technology, a maximum of 16 electronic controlled units (ECU) are able to set up a communication network by using prioritized, object-oriented messages. Finished in 1997, this standard forms the basis for ISO 11783, an international standard with 29-bit addressing and structured in the style of ISO Reference Model for OSI. The integration of ISO 11783 into SAE J1939 ensures compatibility with truck and bus manufacturers. Furthermore, the standard offers implements direct access possibilities to the tractor management as well as standardized diagnostics

Intelligent machinery is able to acquire versatile information at a high spatial and temporal resolution. Assuming minimal additional efforts for integration into farm management, this technology provides an important contribution for documentation and traceability. As an interface, the XML standard seems most suitable and integrates agriculture into Microsoft and UNIX worlds.

**Keywords.** Electronic communication, Point-to-point connection, LBS, DIN 9684, CAN, ISOBUS, ISO 11783, SAE J1939, Diagnostics, Documentation, Traceability, XML.

#### 7.1.1 Introduction

Following the adoption of hydraulics and electricity, the use of electronics and computing considerably improved the whole range of agricultural equipment. The

independent subsystems of today become intelligent and can be interconnected. Autonomous process control systems allow new possibilities for use and optimization and therefore a reduced workload. Furthermore, single systems can be integrated in the control loop of the overall agricultural production process. The production becomes more efficient, precise, environmentally friendly, and it will be traceable.

On account of this, land use systems will appear different in the future. A variety of new terms put the possibilities of more *precision* in the foreground:

- precision agriculture with precision crop farming and precision livestock farming,
- precision horticulture,
- precision viticulture, and
- precision forestry.

To realize the potential of these systems, which can be developed widely, information must be recognized as the missing link and needs to be available to the overall system in a clear, fast, reliable, and cheap way. Therefore, the utilization of common and accepted standards is a major premise. This allows the unification of a wide range of technologies and farm specific conditions in open and manufacturer-independent systems.

Despite coming from different initial situations, the use of electronics and information technology (IT) in crop production and livestock husbandry has passed through similar steps.

### Animal Husbandry

The trend of increasing herd sizes resulted in new challenges to provide efficient supply-chain management and to guarantee good climate control for animal husbandry and welfare.

For feeding, the supply unit needs an error-free, reliable identification of a single animal at a readable distance [1]. Active, and later on, passive, transponders (responders), allowed the recognition of all animals at the supply unit and therefore a distinct definition of the feed ration. Clearly defined feed portions make the assignment of individual feed rhythms for each animal possible. This was first used for cows and calves. Control algorithms were developed to incorporate maximum quantities per visit, optimized distribution of portions over the day, or the control of quantities not picked up, resulting in a notification to the operator via the user interface of the process control system.

The next step was the transfer of the feeding equipment from open-loop control to closed-loop control by animal performance. For this, additional sensors for performance measurement (milk meters for diary cattle, weighing units for calves) were added. The combination of the different subsystems and the personal computer (PC) was typically implemented by each manufacturer in an specific proprietary way to provide a closed communication system.

For climate control, various electronic control systems have been developed. Simple systems use operator-predefined, open-loop control by time. More complex systems use additional information from integrated sensors and refined algorithms in closed loop control [2]. Another improvement was the integration of room temperature and humidity as new parameters. Control systems for different compartments were

implemented as integrated and centralized monitoring units, also in a manufacturerspecific way as closed communication systems.

First attempts for the creation of standardized communication interfaces failed because of missing concepts for the overall system, different interests of different manufacturers and fear of losing market share because of widely interchangeable components [3].

# Crop Production

New problems emerged in more and more intensive crop production from higher demands for more precise application of chemicals (fertilizers, pesticides, herbicides), more efficient use of water in field irrigation, reduction of the damaging slippage of tractor tires, supervision tasks with growing sizes of tractor-implement combinations and self-propelled equipment (combine harvester, choppers, sugar beet harvesters).

The use of electronics in agricultural machinery and equipment started with simple monitoring units with sensor, processor, and display. These were followed by closed, self-contained electronic control systems. Two prominent developments were the electronic hitch controller [4] for tractors and the spray computer [5] for chemical plant protection. Both realizations used, in principle, an internal and manufacturer-specific communication system between sensors, controller, actuators, and user interface. This trend continued with all other implement-specific controllers (Figure 1) at the beginning of the 1980s [6].

However, for a smooth interchange of technology from one tractor implement combination to another, a first restriction was clearly visible. A quick and reliable interchange will only be possible if both tractors provide similar basic signals for speed and other values and use subsequently identical interfaces (connectors). To overcome these problems, the first scientific work at that time [7] proposed the use of an open communication system with shared intelligence in *job controllers* and a *centralized user interface* (Figure 2).

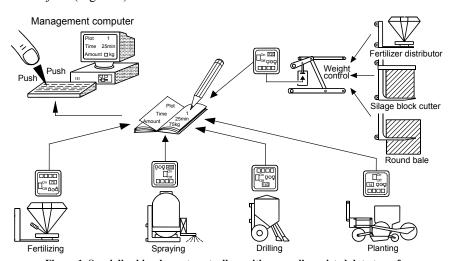


Figure 1. Specialized implement controllers with manually assisted data transfer.

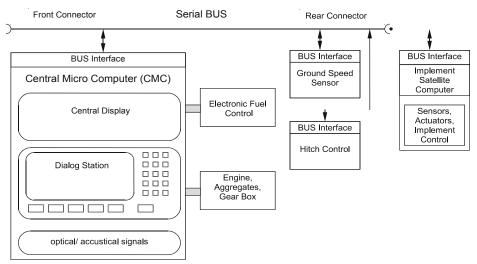


Figure 2. Scheme of a universal information and control system.

As a first implementation, a manufacturer-specific binary unit system (BUS) was developed [8]. It was competed by new families of what became known as *all-purpose mobile farm computers* (Figure 3) with the first smart cards for data transmission to the management computer and with connector-specific implement control [9].

After that, unlike the situation in animal husbandry, large manufacturers of tractors and implements realized the demand and also assisted the standardization of electronic communication systems in crop production.

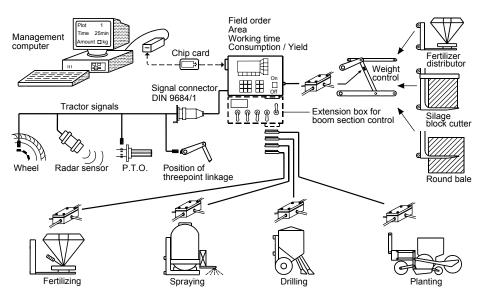


Figure 3. All-purpose mobile farm computer with tractor signals and data transfer media.

# 7.1.2 Tractor-Mounted Sensor Interface for Implement Control

In crop production, tractors are generally used in combination with mounted or trailed implements or transport equipment. In this combination, the tractor provides the required mechanical, electrical, or hydraulic power for drive and cultivation. However, the required and delivered performance parameters of the tractor are not uniform under field conditions. Changing parameters are speed, slippage, engine speed, power take-off speed, and working depth. To achieve a consistent quality of work (e.g., constant output quantity per area), a continuous adjustment of important parameters of the implement is necessary. Hence, each implement should have several sensors for detecting the actual working condition. So, the more economical solution is to make the tractor signals available for the implement via a standardized interface.

A signal outlet that meets these requirements was standardized from 1986 to 1989 as DIN 9684-1 [10] and transferred into ISO 11786 [11] in 1995. Basic information from the tractor is provided by a 7-pole point-to-point-connection to the implement. For each signal, the standard defines voltage limits and slew rates of pulses. For applications, minimum and maximum voltage levels are specified. A complete pin-population is not mandatory, apart from pins 2, 6, and 7 in DIN 9684-1.

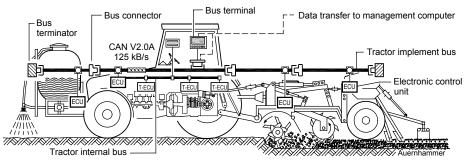
# 7.1.3 Agricultural BUS System (LBS)

Point-to-point-connections according to ISO 11786 only allow data interchange for single tractor-implement combinations and only in one direction. This is a drastic limitation for the development of more complex and intelligent combinations of different equipment that have the potential to reduce the number of tractors, the manpower, the workload of laborers, the energy demand, and the soil compaction (because of fewer applications at ideal work and weather conditions). In sum, this means operation better for the environment at lower cost.

For an optimized adaptation to farm-specific conditions, tractors and implements need to be freely exchangeable and combinable. Therefore, each unit needs to have its own controller with the ability to communicate with each other via standardized interfaces. In a complete system, four bidirectional interfaces are essential for information interchange:

farm management system ↔ tractor implement unit tractor implement unit ↔ operator operator ↔ implement (and tractor) implement ↔ implement

In the period from 1987 till 1997, a first standard was defined at the Landmaschinen- und Ackerschleppervereinigung (LAV) on behalf of the Deutsches Institut für Normung (DIN). A simultaneous research project implemented and verified the standard [12]. The Landwirtschaftliches BUS-System (LBS) developed according to DIN 9684/2-5 (Figure 4) was mainly built upon the needs of central European farms. A fundamental requirement of the agricultural industry was to forbid any direct intervention of the implement-to-tractor functionality.



T-ECU Tractor internal Electronic Control Unit

Figure 4. Mobile electronic communication in an extensive tractor-implement unit by LBS.

# Controller Area Network (CAN)

Instead of developing a new proprietary basic communication system for LBS, it was decided to use a commercially available standard of the automotive industry in order to participate quickly in technical advances, to have economies of scale, and to use the versatile know-how of that branch of industry.

The system of choice in 1989 was CAN version V2.A of the Bosch company [13]. This communication protocol is based on a multi-master architecture with distributed process nodes, serial data transfer and data frames with 11-bit identifiers and 64-bit data fields. The network protocol provides free object-oriented addressability, prioritization, short message frames, and integrated error handling to meet requirements in the most suitable way.

# LBS System Definition

Part 2 of DIN 9684 defines the application range, the basic definitions, the line of action, the configuration, the BUS protocol, the fundamental message format, and the physical layer of the network. This follows the general constitution of CAN according to ISO 11869 [13].

Against the original CAN philosophy, where each node manages only one or just a few sensors or actuators as independent objects, LBS defines nodes by the term *job controller*. Such a job controller provides the whole control and execution activity of a device (the tractor also acts as a device) or the complete functionality of a service (e.g., user interface). Therefore, job controllers are responsible for the complete data communication of a device or a service and their data objects on the BUS.

#### LBS Identifier Structure

Complex tractor implement combinations require the transmission of a great number of data objects under various conditions. This is hardly achievable by 11-bit addressing with a maximum number of 2048 different objects. In consequence, the CAN identifier was specifically modified in LBS (Figure 5).

This allows the identifier to:

- create eight different priority groups, using the most significant three bits;
- assign the highest priority level to system supervision;

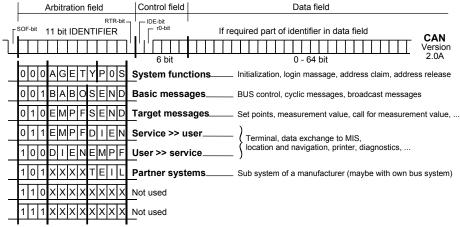


Figure 5. Definition of the 11-bit identifier in LBS.

- assign second highest priority to basic messages;
- assign next priority level to process data;
- offer special services for the BUS system;
- define information in category tables with fixed row and column assignment; and
- shift temporary data definitions into the first 4 data bytes.

# The User Terminal in LBS

A LBS service called the *user station* (BS) was created for human-computer interaction in LBS based networks. The BS acts as a virtual terminal in time-sequential or multi-window mode and provides a means for data representation and interchange with connected devices. For a single job controller, the BS looks like a permanently available terminal. The data interchange is mask-oriented and controlled by each job controller in the LBS network.

At system initialization, the masks for control and data representation of all job controllers need to be uploaded into the data memory of the BS. A menu-driven selection screen allows the operator to choose a certain mask of a certain job controller and to display it in the foreground of the virtual terminal. An interaction via different control elements like soft keys, joysticks or touch screens is allowed due to the fact that the standard only defines the range of functionality but not their technical realization. Direct functions of a machine, like lifting or lowering the boom of a sprayer, can be activated by specially designated keys.

# Data Interchange in LBS

The service task controller in LBS networks was designed for data interchange between the stationary farm management computer and the mobile equipment (Figure 6). This service defines dedicated interfaces for stationary equipment as well as for mobile equipment. Four different ASCII-formatted files are relevant for data interchange with the farm management computer. The syntax corresponds to the Agricultural Data Interchange Syntax (ADIS) of ISO 11787 [13].

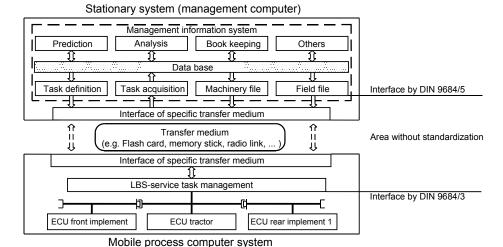


Figure 6. Task controller and data interchange in LBS networks.

For every work assignment, the job file and the feedback file must be redefined and transferred to the mobile equipment. Both other files are identical, as long as the machinery data, the appliances, or the field data are not changing. On the mobile machines, the task controller uses the information of the data files for controlling the implements by means of LBS process data.

The specification of a certain transfer medium was intentionally abandoned. This allows the adaptation of the best available technology in the future.

#### LBS in Practice

After the start of the first test systems in 1993 and the publication of the standard in 1997, the implementation in practice was very reserved. The primary reasons were:

- limited willingness of tractor manufacturers to serialize LBS equipment,
- few available implements with job controllers that conform to LBS,
- incompatibilities,
- insufficient definitions in the standard,
- no direct integration of GPS information,
- problems in coding data transfer between stationary and mobile equipment,
- insufficient definition of the task controller,
- limited economical and ecological pressure for a beneficial utilization, and
- missing test facilities and certification procedures.

By August 2004, about 20,000 tractors in the market were equipped with LBS. Probably only a small percentage of those is in active use (approximately 6,000 LBS terminals).

Nevertheless, a new development has been realized in form of the LBS implement control. For that purpose, the implement manufacturers use the LBS communication and the on-hand LBS terminal in the tractor. Solutions for forage harvesters, fertilizer distributors, seeders, and plows are in practical use.

#### **7.1.4 ISOBUS**

Though a little late, the importance of electronics and information technology in agriculture was also recognized internationally. After the first discussions in 1988, this resulted in the foundation of an appropriate subcommittee (SC 19) in 1991 under the leadership of technical committee 23 (ISO TC 23/SC 19), which is responsible for the whole area of agriculture and forestry. A working group (WG 1) within subcommittee 19 deals exclusively with the application of electronics in mobile machines. In accordance with the participating countries (Denmark, Germany, France, Great Britain, Netherlands, Canada, and the USA) the following definitions were different from LBS DIN 9684:

- using CAN protocol V 2.0B [14] with extended 29 bit identifier,
- increasing the BUS speed to 250 kbit/s,
- structuring the standard in considerably more parts,
- adapting the ISO OSI [15] basic reference model wherever possible,
- defining the standard interoperable to SAE J1939 [16],
- defining the tasks of the tractor ECU for different classes, and
- allowing proprietary message types.

The functionality of DIN 9684 with regard to the agricultural adoption of the information system was adopted almost unmodified by ISO 11783. The new, more complex, standard ISO 11783 [17] received the short term *ISOBUS*.

# Definitions in ISO 11783

The basic concept of the standard follows the OSI reference model, whereas different functionalities are spread across several parts of the standard.

# Part 1: General Standard

This part contains basic definitions of the standard, and characterizes fundamental functions and the organizational structure. At the end of the standardization, this part will be completed and published.

# Part 2: Physical Layer

This part describes the hardware definitions (Figure 7). The main characteristics of the bus structure are:

- BUS length maximum 40 m per segment,
- BUS extension via bridges from segment to segment,
- stub length maximum 1 m,
- ECU distance minimum 0.1 m,
- ECUs maximum 30 per segment,
- ECUs maximum 254 in total, and
- cable (media) twisted quad, unshielded, 75  $\Omega$  impedance, 2 signal wires, 2 BUS-terminator supply wires.

The main characteristics of the connectors are:

- implement BUS: breakaway connector, 9-pole with integrated power supply,
- BUS extension: 9-pole with integrated power supply,
- in-cab plug: 10-pole for terminal connect through,

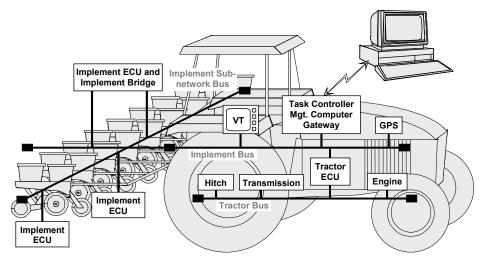


Figure 7. ISOBUS setup of a tractor and a towed implement [18].

- ECU plugs: proprietary from ECU to BUS, and
- diagnostic: 9-pole (note: plug connection due to ISO 11783-2).

This part of the standard also regulates bus failure and fault confinement.

# Part 3: Data Link Layer

The basic protocol is specified as CAN version 2.0 B. To guarantee compatibility between truck and bus applications and off-road (agricultural) applications, the message frame format was harmonized with the standard SAE J1939. All information has to be integrated in *protocol data units* (PDU). A single PDU consists of 7 fields, representing the key for its interpretation. Six fields reside in the identifier: priority, reserved, data page, PDU format, PDU specific, and source address. The seventh field, called data, stays in the 64-bit data field of a single CAN message. The 8672 different possibilities for PDUs are split in two groups, PDU1 and PDU2 (Figure 8).

When using the 480 different PDU1, destination specific and global addressing is allowed. All 8192 PDU2 can only be used to communicate parameter groups as global messages. For the processing of PDUs the procedures, response handling, response times (< 200 ms), or waiting periods (1250 ms) are specified. Five message types are currently supported:

- *Command*—To be a specific or a global destination from a source (e.g., transmission control);
- *Request*—Provides the capability to request information globally or from a specific destination:
- *Broadcast/response*—Can be a broadcast of information or a response to a command or request;
- Acknowledgement—Provides a handshake mechanism; and
- *Group function*—Can be used for groups of special functions (e.g., proprietary functions, multi-packet transport functions, etc.).

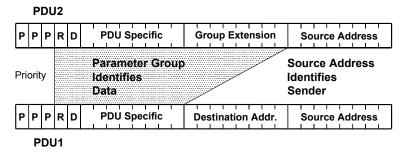


Figure 8. Identifier structure of ISOBUS [18].

#### Part 4: Network Layer

This part of the standard specifies rules, requirements, and features for data interchange within the different segments of the communication network. The main focus lies on the specification of network interconnection units and their tasks like message forwarding, filtering, address translation, message repackaging, or database management. Moreover, the different interconnection unit types, such as repeater, bridge, router, or gateway, and their responsibilities are clarified. A typical example is the scope of the tractor ECU (see Part 9, Tractor ECU).

# Part 5: Network Management

The network management covers the whole area of address, type and name handling for setting up each ECU as a uniquely identifiable member in the network. All important procedures, conventions, and sequences are specified here, like name specifications, name fields, address-management messages and procedures, types of ECU (standard, diagnostic/development tool, network interconnection unit), address types of ECU (non-, service-, command- or self-configurable-address), network initialization procedures, and network functionality.

#### Part 6: Virtual Terminal

A virtual terminal (VT) is an ECU with graphical display and input functions that provides the capability for other ECUs to interact with an operator. It is designed to display information and to retrieve data from a user. In order to reduce communication bus traffic as much as possible, the interface protocol is organized in an object-oriented manner. Apart from some essential exceptions, only the functions, not the design, of the user interface of the VT are defined. The physical size, resolution, orientation, and methods for implementing the graphical display are left to the free decision of the designer.

Information from connected ECUs is shown in display areas that are defined by data masks, alarm masks, and soft key masks. The data of these masks is contained in object definitions, which jointly compose the object pool. Before the first interaction, the object pool must be loaded either via CAN bus or by some other means into the VT. A change between different masks can be initiated by a single message.

This part of the standard corresponds to the definitions in LBS DIN 9684 with extensions in matters of display design, soft keys, navigation methods, editing methods, control methods, and an additional usage methodology.

# Part 7: Implement Messages Application Layer

This part of the standard is closely related to SAE J1939/71 with specific adaptations and extensions for agricultural applications. It also covers all specifications of the signal connector (ISO 11786). A broad range of data types and the appropriate CAN message formats are defined here, such as time and date, speed, distance and direction, key switch state, power maintenance, implement state (park, transport, work), navigation parameters due to NMEA 2000 [19], hitch and power take-off parameters, valve control data, language and units, lighting data, process data, and tractor remote control messages.

# Part 8: Power Train Messages

The power train messages are equal to the definitions in the standard SAE J1939. Therefore, this part has only a cross-reference to the corresponding document of standard SAE J1939. All messages for vehicle-specific data, adjustment and actuation of brakes, dimensions, fuel, on-board electrics, loading, engine power, engine speed, torque, and other parameters are specified here.

# Part 9: Tractor ECU

On a system with an ISOBUS network, the tractor ECU functions as the gateway between the tractor bus and the implement bus. It must appear on the implement bus to be the same as any other ECU in the network and the access to the VT is identical to that of any other implement. As a network interconnection unit, the tractor ECU is also responsible for converting process data and tractor bus messages with appropriate parameters.

Tractor ECUs are divided into classes. A tractor class specifies a minimum set of supported messages, which can be provided to other members in the network. There are three main tractor-implement interface classes:

- Class 1—A tractor ECU with this specification has a simple network-supporting
  interface and provides mainly the basic tractor internal measurements comparable to the signal connector (ISO 11786). It supports power management, stores
  language specific parameters, and allows control of tractor lighting.
- Class 2—Additional to class 1 messages, the main enhancements of class 2 are: time and date, ground and wheel-based distance and direction, rear draft information, full implement lighting message set, and auxiliary valve status. This allows a more sophisticated implement control and security strategy.
- Class 3—This class of tractor ECU accepts commands from the implement bus. In particular, the basic commands for rear hitch, power take-off, and auxiliary valve control must be processed. Hence, the implement is able to control the desired power source and the hitch position.

Special letters shall be appended to the class number when the tractor is able to provide navigational information on the implement bus (N) or when it supports front-mounted implement information (F).

# Part 10: Task Controller and FMIS Data Interchange

This particular standard defines the task management in ISOBUS. In this manner, the term *task* is defined as the execution of work on one field or one customer. The task controller is therefore an ECU on the *mobile implement control system* (MICS), which is responsible for sending, receiving, and logging of process data. Another component, the *farm management information system* (FMIS), consisting of the stationary farm computer with appropriate software, is needed for the generation of tasks.

This standard defines the requirements and services needed for communicating between the task controller and ECUs. Also, the data format to communicate with the farm management computer (ADIS syntax according to ISO 11787), the calculations required for control and the message format sent to the ECU are specified.

The task management has the following workflow:

- 1. planning field tasks and/or operations using the FMIS;
- 2. converting the task data to the data format required on the MICS;
- 3. transferring the task data to the task controller on the MICS;
- 4. the task controller uses the task data for controlling the implement via process data messages;
- 5. the task controller collects simultaneously data of the task execution;
- 6. transferring the collected data to the FMIS computer; and
- 7. converting and evaluating the resulting task data.

In the near future, it is planned to change the syntax for data transfer from ADIS to the broadly based Extensible Markup Language (XML) (see also Section 7.1.5.)

# Part 11: Data Dictionary

The mobile agricultural data element dictionary is a listing of all data objects and their data elements. A data element is an information unit, consisting of the attribute data format, length, digits after decimal point, unit, and direction of communication. For each of 16 implement types that are used in agriculture and forestry, data elements are defined and structured in a table. These 16 implement types correspond to the different device classes (ISO 11783-1, Annex B, Table B4), e.g. tractor, tillage, secondary tillage, planters/seeders, fertilizers, sprayers, harvesters, and so on. Any table consists of 16 rows and 16 columns, therefore there are 256 possible data elements for each device class. A single data element can be identified by a number that is derived from row and column number in the appropriate table. These logically structured tables are a key element for process data communication within the network.

The data dictionary was originally developed and specified for DIN 9684, but also adopted for ISO 11783.

#### Part 12: Diagnostics

The diagnostics part of ISO 11783 is at the development stage. A newly created task force is verifying all needs and requirements in this field. An important item is the specification of a diagnostic connector with extended pin allocation in comparison to the definition in ISO 11783 Part 2. The diagnostic system shall support ISO 11783 and

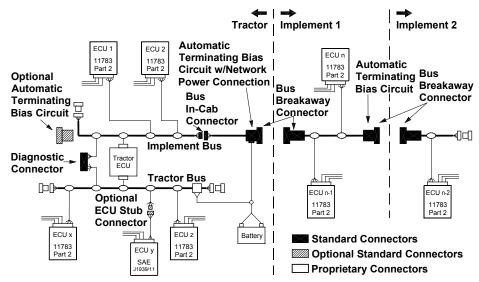


Figure 9. Diagnostic system connected networks.

other standardized networks used in agriculture and forestry (Figure 9). The diagnostic protocols that can be used on the network are KWP 2000 (Keyword Protocol 2000, see ISO 14230 [20]), ISO 157658 [21], and J1939-73 (Application Layer Diagnostics).

#### Part 13: File Server

A file server is a distinct ECU on the mobile implement bus that provides storage for files to all controllers on the network. This enables extended task control while maintaining a single gateway to the management information system. This part of the standard specifies the general message format, the file data format, and the data transmission control. Furthermore, special parameters like command groups, command functions, flags, handles, or file and directory attributes are defined here. A comprehensive message pool enables other network nodes to communicate with the file server and to handle files and their dedicated content.

The file server specifications are under revision and not yet published officially.

#### ISOBUS in Practice

The entire standard ISO 11783 is very large and complex. In comparison with LBS (DIN 9684), it can be expected that the introduction and realization will take considerably more time. After 12 years of standardization work, three stages of development are foreseeable:

- 1. Minimum Standard—ISOBUS will be a worldwide accepted standard in contrast to the German standard LBS (DIN 9684). It is likely that the implementations will only have a minimum level of compatibility concerning their communication abilities. Therefore, ISOBUS-compatible ECUs will:
  - ensure the initialization and network logon procedures;
  - be able to handle basic messages;

- be able to handle a limited set of process data; and
- mostly have no or only a simple interface on the VT.
- 2. Tractor ECU—Depending on the marketing concept, tractor manufacturers will install different equipment versions:
  - implementation level I for standard tractors;
  - implementation level II will be chosen by most full-liners (manufacturers with a complete range of products); and
  - implementation level III for high-end tractors with universal options for the control of implements.
- 3. Proprietary Systems—With the incorporation of proprietary messages in ISO-BUS, full-liners can now realize design concepts where combinations of captive machines and equipment are smarter than in combination with equipment of other manufacturers. It is expected that such solutions will dominate, especially because of the concentration of manufacturers towards a few global players.

#### 7.1.5 Standardized Data Transfer

Electronic process monitoring and control systems must have a channel of information to the management system for monitoring the process status, gathering data from the process, delivering control variables to the process, commanding global tasks, and initiating diagnosis. The required connections are realizable via a broad range of standardized interfaces in on- or off-line mode. Indeed, for a manufacturer-independent information exchange, these interfaces require a clear specification of the transported data and an appropriate syntax.

# Proprietary Protocols

Closed-up control systems almost exclusively use standardized interfaces with proprietary protocols. Point-to-point connections are generally in star topology based on RS232 [22] or RS485 [23]. More complex approaches utilize fieldbus systems.

#### Standardized Protocols

In the 1980s, the Netherlands developed a first agricultural standard for on-farm data transfer. The international standard ISO 11787 "Data interchange between management computer and process computers—Data interchange syntax" defines for point-to-point connections the agricultural data interchange syntax (ADIS) [24]. Here, the transfer for alphanumeric data uses the ASCII file format. The characteristics of the transporting data have to be specified in a data dictionary (DD) [25]:

- ASCII code in 8 bit,
- line format.
- reserved characters, and
- line types with a distinct recognition in the first row.

For each line type, the second row defines a status character. The subsequent characters represent the content, as defined by line type. A closure of each line by the AS-CII characters Carriage Return <CR> and Line Feed <LF> is generally necessary. In practice, the ADIS file format is mainly used in the livestock area with well-defined data dictionaries.

#### XML

Extensible Markup Language (XML) is a successor of ISO 8879 from 1986 [26] and represents a universal concept for data representation. It differs from the initial version by XML being increasingly oriented towards practical usage. It defines documents, logic structures, physical structures, the conformance, and the notation.

To use XML, a software module called an XML processor is necessary in order to read and edit XML documents. As a part of the application, this module embodies the interface to the transported information.

XML is supported in both Microsoft and UNIX. The language HTML also has found a great utilization in the World Wide Web and is therefore of great interest for agricultural applications.

#### 7.1.6 Farm BUS

The use of electronics on farms started in the 1970s in the areas of dairy cattle feeding and climate control in hog and poultry keeping. Independent developments by various small manufacturers has lead to proprietary solutions. Parallel to this, the personal computer (PC) was used for farm management tasks. The worldwide first approach in Germany to standardize data transfer between process controllers and management computers failed because of the undervaluation of information and the rivalry between the system vendors [3]. In contrast to this, efforts in the Netherlands for a data transfer between farmer and livestock breeding organizations with ADIS was crowned with success.

Also, the joint European initiative to standardize animal identification was successful. An urgent need for standardization evolved, driven by the huge number of units, worldwide marketing strategies, and the use in process control, animal monitoring, and veterinary care, as well as the huge interest of the existing breeding organizations. This was addressed and codified 1996 in ISO 11784 [27].

#### Precision Livestock Farming

Today, electronic applications in animal husbandry have reached a very high level. These cover all animal species and range from single animal identification to full commercial operations. With the 1<sup>st</sup> European Conference of Precision Livestock Farming it was possible to bring together common interests [28]. However, that conference also pointed out the deficiencies because of a missing information interconnection

# Network for Livestock Farming

In March 2003, a new working group was established by affirmation of ISO under TC23/SC19/WG2. There is an understanding (Figure 10) about connecting process controllers via a manufacturer-independent BUS system and defining the physical design of the network, about transferring the communication concepts of ADED (Agricultural Data Element Dictionary) in [25] to XML, and working out a comprehensive Data Dictionary. By means of the developing standard ISO 17532 [29], the missing information interconnection should be transferred to a future concept.

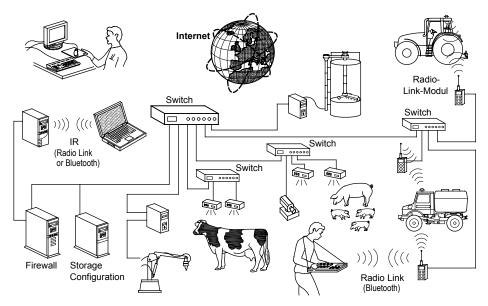


Figure 10. Network for livestock farming (NLF) by ISO 17532.

#### References

- 1. Artmann, R. 1999. Electronic identification systems: State of the art and their further development. *Computers and Electronics in Agriculture* 24(1-2): 5-26.
- 2. Leuschner, P. 1986. Einsatz der Mikroelektronik bei der Stallklimatisierung. *Landtechnik* 41: 485-488.
- 3. Auernhammer, H. ed. 1986. *Schnittstellen für den landwirtschaftlichen Betrieb*. Weihenstephan: Inst. für Landtechnik.
- 4. Hesse, H. 1982. Signalverarbeitung in Pflugregelsystemen. *Grundl. Landtechnik* 32: 54-59.
- 5. Göhlich, H. 1978. Pflanzenschutztechnik 1978. Landtechnik 33: 310-312.
- 6. Auernhammer, H. 1989. *Elektronik in Traktoren und Maschinen*. München: BLV-Verlag.
- 7. Jahns, G., and H. Speckmann. 1984. Agricultural electronics on farm machinery needs standardized data transfer: A concept. ASAE Paper No. 84-1633. St. Joseph, MI: ASAE.
- 8. Peisl, S., H. Auernhammer, and M. Estler. 1994. FuE-Vorhaben Elektronik in der Außenwirtschaft-Ergebnisse der technischen Prüfungen. *Elektronikeinsatz in der Landwirtschaft*. Münster-Hiltrup: Landwirtschaftsverlag, KTBL Arbeitspapier 175.
- Bosselmann, G. 1987. Elektronikeinsatz in der Außenwirtschaft. *Landtechnik* 42: 498-499.

- 10. DIN 9684. 1997+. Landmaschinen und Traktoren-Schnittstellen zur Signal übertragung. Berlin, Germany: Beuth Verlag (parts 1 to 5, different years).
- 11. ISO 11786. 1995. Agricultural tractors and machinery—Tractor-mounted sensor interface—Specifications. Geneva, Switzerland.
- 12. Auernhammer, H., and J. Frisch, eds. 1993. *Mobile Agricultural BUS-System–LBS*. Münster-Hiltrup: Landwirtschaftsverlag, KTBL Arbeitspapier 196.
- 13. ISO 11519. 1994. Road vehicles—Low-speed serial data communication. Geneva, Switzerland.
- 14. ISO 11898. 1993. Road vehicles—Interchange of digital information—Controller area network (CAN) for high speed communication. Geneva, Switzerland.
- 15. SAE J1939. 2003. Recommended practice for a serial control and communications vehicle network. (10 parts, published in different years). Warrendale, PA: SAE.
- 16. ISO/IEC 7498-1. 1994. Information technology–Open systems interconnection–Basic reference model: The basic model. Geneva, Switzerland.
- 17. ISO 11783. partly in print. Tractors, machinery for agriculture and forestry–Serial control and communication data network. Geneva, Switzerland.
- Stone, M. L., D. M. Kee, C. W. Formwalt, and R. K. Benneweis. 1999. ISO 11783: An electronic communications protocol for agricultural equipment. St. Joseph, MI: ASAE.
- 19. NMEA 2000. 2001. Standard for serial-data networking of marine electronic devices. Severna Park, MD: NMEA.
- 20. ISO 14230. 1999. Road vehicles–Diagnostic systems–Keyword Protocol 2000. Geneva.
- 21. ISO 15765. (partly in print). Road vehicles—Diagnostics on controller area networks (CAN). Geneva, Switzerland.
- 22. EIA/TIA 232-E. 1991. Interface between data terminal equipment and data-circuit terminating equipment employing serial binary data interchange. Arlington, VA: EIA/TIA.
- 23. EIA/TIA 485. 1983. Standard for electrical characteristics of generators and receivers for use in balanced digital multipoint systems. Arlington, VA: EIA/TIA.
- 24. ISO 11787. 1995. Machinery for agriculture and forestry–Data interchange between management computer and process computers–Data interchange syntax. Geneva, Switzerland.
- 25. ISO 11788. 1997. Electronic data interchange between information systems in agriculture–Agricultural data element dictionary. Geneva, Switzerland.
- 26. ISO 8879. 1986. Information processing—Text and office systems—Standardized Generalized Markup Language (SGML). Geneva, Switzerland.
- 27. ISO 11784. 1996. Radio frequency identification of animals—Code structure. Geneva, Switzerland.
- 28. Cox, S., ed. 2003. *Precision Livestock Farming*. Wageningen, The Netherlands: Academic Publishers.
- 29. ISO/CD 17532. 2004. Stationary equipment for agricultural and forestry—Communications data network for livestock farming. Geneva, Switzerland.

# 7.2 Internet Use in Agriculture, Remote Service, and Maintenance: E-Commerce, E-Business, E-Consulting, E-Support

K. P. Ferentinos, K. G. Arvanitis, and N. A. Sigrimis

Abstract. This part of Chapter 7 deals with some recent advancements in the agribusiness sector due to the appropriate use of several services available through the Internet. Initially, the general categories of e-business and e-commerce are presented. The possible customer-centered interactions are analyzed and the characteristics of remote service and maintenance developed in agriculture are presented. Subsequently, the development of e-commerce in agriculture is investigated by presenting its characteristics, theoretical benefits, and impact factors. Finally, some representative applications of agribusiness on the Internet are presented and some conclusions on the successful adoption of e-commerce in agriculture are drawn.

Keywords. E-commerce, E-business, Agribusiness, Agri-food chain, Remote service and maintenance.

#### 7.2.1 Introduction

The recent development of information technology (IT) and its penetration into commerce and business transactions has led to the emergence of electronic business and commerce, usually referred to as *e-business* and *e-commerce*. The relevant actions of customer consulting and support have of course followed the same path, leading to the emergence of *e-consulting* and *e-support*.

E-commerce is simply defined as business transactions conducted over the Internet, or more generally through digital communications. Transactions may involve material goods, immaterial services (e.g., consulting and support), or rights and obligations. A variety of high-tech methodologies play an important role in e-commerce, but its backbone is the Internet. Access to Internet communication channels used in e-commerce is often open to everyone but is sometimes restricted and the messages exchanged may be rigidly standardized, as in Electronic Data Interchange (EDI). E-commerce also includes all inter-company and intra-company functions (such as marketing, finance, manufacturing, selling, and negotiation) that enable commerce as well as use of electronic mail, EDI, file transfer, fax, video conferencing, workflow or interaction with a remote computer.

#### 7.2.2 Categories of E-Business and E-Commerce

Transactions performed in e-commerce can be categorized according to the partners involved: consumers, business, and government. Only three of the six possible combinations are presently important: business-to-consumer (B2C), business-to-business (B2B), and consumer-to-consumer (C2C) [1].

- Business-to-consumer (B2C) e-commerce involves retailing products and services to individual shoppers.
- Business-to-business (B2B) e-commerce involves the sales of goods and services among businesses.
- Consumer-to-consumer (C2C) e-commerce involves consumers selling directly to other consumers

Across all industries, global B2B e-commerce was valued at US \$356.7 billion in 2000, which represented around 60% of global e-commerce [2]. According to some analysts, B2B represents close to 80% of all e-commerce transactions and could represent as much as 87% by 2004 [1,3]. That year, worldwide B2B e-commerce revenues could reach nearly US \$2.8 trillion [4]. The contrast in the development of B2B e-commerce as compared to B2C may be linked to the ease and cost effectiveness with which offline B2B processes can be replicated online compared to more complex B2C processes. In addition, B2B commerce can reduce errors in commerce-related documents from 20% down to less than 1% [5]. Recently, C2C e-commerce has gained a large portion of global e-commerce due to the development of web auction sites (eBay, uBid, etc.) which allow people to sell their goods to other consumers by auctioning them off to the highest bidder.

Another way of categorizing e-commerce is in terms of the development approaches that are followed by individual e-businesses. In [6] and [7], three categories are identified, based on the primary drivers in their development:

- *E-marketplaces* are neutral to buyers and sellers, seeking to innovate in the core Internet technology taking into account the interests of both buyers (characteristics and needs) and sellers (margin requirements). With the e-marketplaces, also termed electronic hubs or *e-hubs*, companies can link up to many buyers without having to create point-to-point connections to each and can potentially find new customers. Companies purchasing products don't have to manage several different systems for buying from various suppliers and can save money by comparing prices and purchasing from a wide range of companies [8].
- *E-distribution sites* are designed to serve sellers by removing and replacing the existing distribution chain. They must ensure that current sales are not alienated and also ensure that the best orders are not skimmed off leaving other channels to pick up the smaller and more expensive orders.
- *E-procurement sites* are designed to serve the buyers by aggregating online buyers and using volume to force prices down. They are unattractive to distributors due to price transparency which could potentially erode savings achieved by some cheaper sales channel.

E-marketplaces are expected to emerge as a dominant force in e-commerce, accounting for 56% of the value of all B2B transactions by 2004, compared to 7.5% in 2000 [6]. The advantage of an e-marketplace is its ability to replicate offline behavior. This is achieved by offering a range of applications tailored to meet the needs of both target buyers and sellers [9].

#### 7.2.3 Customer-Centered Interactions

An important trend observed in e-commerce is in the supply of information. The Internet provides companies with new channels of communication and interaction that can create closer yet more cost-effective relationships with customers in sales and marketing as well as in consulting and support. Companies can use the web to provide ongoing information, service, maintenance and support, creating positive interactions

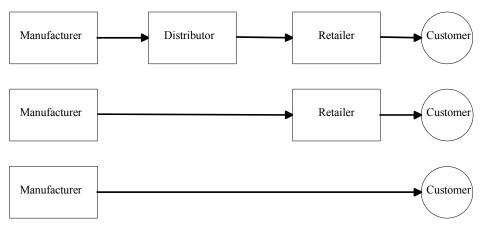


Figure 1. Disintermediation in a value chain.

with customers that can serve as the foundation for long-term relationships and encourage repeat purchases [1]. New approaches to customer service and support have been inspired by the web and other network technologies. Customer questions can be answered with e-mails through the web sites of the companies or by providing helpful information in the web site that can give advice and answers to the customers. In this way, the web is used as a medium through which customers can interact with the companies, at their convenience, and find information on their own that previously required a human customer support expert. According to [1], automated self-service or other web-based responses to customer questions cost one-tenth the price of a live customer service representative on the telephone. In addition, the current development of technology allows the experts of some company to remotely manage and service products (mainly software-based products) via the Internet or direct modem connections.

With the capability of direct sales of the manufacturers' products to retail customers over the Internet, the intermediaries such as distribution or retail outlets are bypassed. Eliminating intermediaries in the distribution chain can significantly reduce purchase transaction costs. To pay for all the steps in a traditional distribution chain, a product may have to be priced as high as 135% of its original cost to manufacture [10]. In addition, with this elimination of intermediaries, companies can achieve higher profits while charging lower prices. The elimination of organizations or business process layers responsible for intermediary steps in a value chain is called *disintermediation* (Figure 1).

#### 7.2.4 Remote Service and Maintenance in E-Commerce and E-Business

Service and maintenance in e-commerce have as targets both the operation of the company and the product that reaches the customer. Businesses that are seriously pursuing e-commerce and e-business need special tools for maintaining their web sites. These tools include web server and e-commerce server software, customer tracking and personalization tools, web content management tools and web site performance

monitoring tools. According to [1], the functions that specialized e-commerce server software must perform for both B2C and B2B e-commerce include the following:

- setting up electronic storefronts and electronic catalogs to display product and pricing information;
- designing electronic shopping carts so consumers can collect the items they wish to purchase;
- making shipping arrangements;
- linking to electronic payment processing systems;
- displaying product availability and tracking shipments;
- connecting to back-office systems where necessary; and
- reporting on both the business transacted through the site and the functioning of the site.

Customer tracking and personalization tools have three main goals [1]:

- collecting and storing data on the behavior of on-line customers and combining that data with data already stored in the company's back-office systems;
- analyzing the data in order better to understand the behavior of on-line customers; and
- identifying developing customer trends.

Web content management tools consist of software that facilitates the collection, assembly and management of content on a web site, intranet or extranet, while web site performance monitoring tools consist of software for overseeing the speed of downloading web pages, performing web transactions, identifying broken links between web pages and for pinpointing other web site problems and bottlenecks [1].

The second main target of remote service and maintenance is the actual product. This includes the customer support via the communication channels between the company and the customer, which have been seriously expanded with the development of the telecommunication technology and the Internet, as explained in the previous section, but it also includes direct product maintenance through the Internet or direct modem connections. Several products, mainly in cases where software is involved, include the capability of remote service and maintenance. The company's expert can directly connect to the product and perform troubleshooting, routine service, software upgrades, etc. Many of these procedures can be done automatically, by connecting to special portals in the company's web site, even without the company's expert.

#### 7.2.5 E-Commerce in Agriculture

During recent years, e-commerce has found its way to agriculture. As the Internet continues to become more popular among people who deal with agricultural businesses of any type, e-commerce finds further applications in agriculture. Participation in e-commerce requires that both buyers and sellers have access to the Internet and that they are able to use the required hardware and software effectively. At the level of business-to-business, common agribusiness B2B transactions such as buying, selling, trading, delivering, and contracting seem to be natural targets for conversion to e-commerce [11]. In agriculture, the B2B and B2C categories of transactions are often

referred to as agribusiness-to-agribusiness (A2A) and agribusiness-to-grower (A2G) [12].

Many theoretical benefits of e-commerce in agriculture can be identified:

- Promotion of information flow [13]—The exchange of information about agricultural products, their characteristics, advantages, disadvantages, etc., can be greatly increased through the information highways of the Internet during ecommerce transactions. In addition, Internet technology provides the opportunity to link individual actors in the food production chain together, irrespective of geographic location. This has the potential to improve market access through online transactions and by reducing geographic obstacles to market reach, such as time and distance.
- Market and price transparency [13,9]—Online access to product and price information allows comparison of products and increases price transparency. Price differentials resulting from geographic location can diminish because of increased competition. This may benefit farmers with regard to input prices but may reduce the prices received for their products. This may be particularly true where many existing products are not differentiated, are required on a regular basis, and where there is a heavy reliance on input supplier expertise in choice of product [9].
- Reduction or elimination of transaction costs [14,15]—With the application of e-commerce, many transactions through the supply chain are either eliminated or simplified. In this way, transaction costs are drastically reduced or even eliminated.
- *Increase in online cooperatives* [9]—The processes of e-commerce transactions give emergence to new opportunities for furthering the concept of cooperatives. From the perspective of farmers, e-cooperatives provide a solution for small businesses to increase critical mass [16].

The Internet plays a role in agribusiness both as a new marketplace and as an information resource. Numerous applications have been developed by different interest groups. Applications in the marketplace can be categorized from a farmer's viewpoint, including services, outputs, and factors of production and inputs [12]. Services, inputs, and production factors are generally purchased via the Internet at a fixed price, whereas outputs are generally traded through an auction. This is perhaps because many of the outputs are perishable and therefore the market price is more sensitive to supply and demand.

The development of e-commerce in agriculture is of course strictly linked with the adoption of the Internet by, mainly, the farmers. Reports show that farmers are slightly behind the general population in adoption [17-19]. In the United States, the proportion of farmers who had access to the Internet had risen from 13% in 1997 to 29% in 1999 [20]. By June 2000, 58% of Australian farms had computer access and 34% had Internet access [21]. In Germany, 78% of commercial farmers with Internet access use it for electronic banking, while 28% use it to purchase goods and 19% use it for selling goods [22]. However, not all farmers use Internet access for productive purposes, and

social and recreational uses may also be important. In general, larger farms tend to adopt both Internet technologies and e-commerce behavior more rapidly [18]. However, the type of farm/non-farm agribusiness appears to be even more important than the size, and accounts for small farm opportunities online. This often relates to niche products and unique market position [23]. When it comes to actual online commerce activity, about 15% of farmers engage in some form of farm-related buying and selling online, but this online activity constituted only 0.3% of total farm commerce in 2000 [24].

Goldman Sachs [25] believes that about 12% of total US agricultural sales could migrate online by 2004. This translates into a \$120 billion by 2004, which is an important increase compared to the \$34 billion in 2000 [26]. However, [27] concludes that despite significant growth in use of the Internet by farmers, it is still far from universal and likely to stay so for some considerable time. The reasons for non-adoption according to the same research have to do with utility, relevance, and affordability of the product, though technical limitations (especially speed of rural telecommunications links) still have an influence [27].

Goldman Sachs also argues that the potential for B2B e-commerce is greater in industries with the following characteristics: a highly diffuse supply chain; pressure to control costs; complex product specifications; processes accounting for at least 20% of total costs; and technological innovation is part of the industry's culture.

The increasing adoption of information technology leads of course to direct production gains and, at the same time, to reductions in operating costs. There are several areas where this might happen, including:

- better retrieval and evaluation of available data for management purposes;
- development of management decision support systems;
- development of processes for quality assurance and external regulatory compliance requirements;
- better links to remote sensing and geographic information systems (GIS) data;
- better links to technical and other information;
- better links to agricultural suppliers;
- more direct feedback from customers and consumers;
- improved supply chain management (see Chapter 8, Section 8.2); and
- opportunities for marketing and other networks to emerge.

According to [28], there are three major factors impacting the development of e-commerce in agriculture:

- *Industry structure*—The structure of agribusiness industries has changed in recent years. Some major consolidations that have taken place at all levels of the value chain can impact agribusiness e-commerce in three ways: (1) reduction of the need to electronically coordinate fragmented marketplaces [29], (2) creation of barriers to the development of transparent electronic marketplaces [30] and (3) development of internal barriers to adoption [28].
- *Product complexity*—The increasing complexity of products being sold in agricultural markets slows down e-commerce development in agriculture. This

complexity is both traditional and end-user driven. Agricultural products, unlike manufactured goods, are partly uniform and their prices depend on several factors. Some quality standards like product grades help simplify product description, but time and location factors also influence the price of a product. In addition, local prices also may vary constantly. With current technology, the cost in time and effort to the buyer of searching for all these attributes may outweigh any advantages the buyer may gain [28]. The end-use driven products that add to the complexity of products have evolved lately due to the focus on consumer demands for healthier, more convenient and more flavorful foods, and also due to the development of trait-specific products valuable to processors along the value chain [28].

• The high-touch nature of transactions—It is undisputable that farms have become increasingly more business-like in recent years and farmers operate in an A2A environment. However, there is great evidence that personal relationships still play a very important role in all transactions in agribusiness. There are some that argue that agriculture is fundamentally driven by relationships [31]. Farm operator evaluation by farm managers relies more on trust and recommendations than production factors [32]. Personal relationships even have an impact on land prices [33].

Statistical results presented in [34] make the impact of e-commerce on the agrifood sector more specific. A slow rate and a small extent to which the adoption of e-commerce is taking place is identified. Similar problematic situations can be resolved by the proposal of specific development directions [35] and the identification of characteristics that are associated with successful e-commerce firms throughout the agrifood chain markets [36]. In addition, the market complexity of the agrifood sector is particularly high, mainly because of the large product variety and different sets of market situations [37]. Thus, specialized systems for electronic trade that identify that complexity of the transaction process are required. Based on that concept, a software prototype is developed in [38] that flexibly addresses the different needs in agrifood markets and explores different process alternatives in transaction processes. The resulted system is evaluated in [39] and its performance shows that electronically supported transaction processes in specific agricultural market applications are more efficient than the respective traditional transaction processes.

Therefore, it is clear that global challenges in the agri-food sector must be met with IT used far more than currently [40]. To realize such a high level of IT use in agribusiness will take cooperation, for example [40]:

- cooperation between different areas of competence such as competence in it or competence in business and market management;
- cooperation between groups of enterprises from all stages of the food supply chain, extension organizations, market organizations, and other related services;
- cooperation between the providers of IT services to secure the technical feasibility and efficiency of digital integration and to adapt technology to content;

- cooperation among small-scale enterprises, especially farmers, to be able to utilize the emerging digital environments and to open opportunities for virtual cooperation;
- cooperation between users, research, extension, and system design to arrive at applications which better fit the operational needs of users; and
- cooperation in the development of system marketing strategies and the development and implementation of training opportunities.

# 7.2.6 Internet Applications in Agribusiness

Numerous Internet applications have been developed in the agribusiness sector by different interest groups. These applications can be divided as market-specific and information-specific. In the first case, that is, in the marketplace, they can be categorized under three broad themes from a farmer's viewpoint [41]:

- Factors of production and inputs—This refers to the Internet sites that trade all possible items that can contribute to agriculture, like land [42], agricultural chemicals [43], machinery and equipment [44], and fertilizers [45].
- Services—This refers to the sites that offer online logistical, transport, and storage services. Some examples are banking facilities such as loans [46], insurance [47], legal services [48], and even some innovative applications like trading services of milk quotas [49].
- Outputs—Applications on the output side are generally in the form of online auctions. Some representative applications include cattle auction sites [50,51], hay [52], fish and fish products [53], specialty market products like walnuts [54] or wine and related items [55].

The information-specific Internet applications in agriculture can be categorized in areas such as information itself, management tools, and links to regulatory bodies [41]:

- Information—This refers to farm magazines that are online, such as Farmers Weekly [56] and Hoard's Dairyman [57], sites that provide market information [54, 58] and analysis [59], online weather reports [60], and specialist advice [61].
- *Management tools*—This refers to online tools that include calculators, databases, information tracking, and analysis tools and electronic forms. Examples of calculators include a beef management profitability calculator [62], a milk quota calculator [63], and a loan calculator [64]. Information tracking and analysis examples can be found on the accountancy packages available at [63] while a representative database application is realized in [58].
- Links to regulatory bodies—Many sites provide links to regulatory bodies for supply of official publications, reports, press releases, and other tools. Examples include the European Commission, Agriculture Directorate-General [65], the Organisation for Economic Co-operation and Development [66], the World Trade Organization [67], and the National Farmers' Union of the UK [68].

#### 7.2.7 Conclusions

The commercialization of the Internet has caused agribusiness firms to rethink their distribution channel. E-commerce provides firms with the ability to reach new customers and old customers in new ways. In addition, e-commerce also allows firms to tap new and old suppliers through new and innovative channels. These possibilities have raised the expectations of improved efficiency and substantial cost savings. The process and function view of the supply chain is used to guide the analysis into Internet/e-commerce adoption by agribusiness firms. The ability of the Internet to reduce transaction costs through improvements in transaction, information, and negotiation functions of the supply chain is associated with higher probabilities of Internet/e-commerce adoption among agribusiness firms [69].

It is evident that, despite the barriers that slow down e-commerce adoption in the agricultural sector, the development of agribusiness e-commerce to date will lead to the successful wide adoption of e-commerce in agriculture. Today's high-tech transactions mean that e-commerce players have to lead individual users toward a new way of conducting business. The nature of adoption of innovation will create opportunities for those who cater to individual needs for learning and training [70]. E-learning and e-training will become in some cases the prerequisite and in other cases the extension of e-consulting and e-support.

# References

- 1. Laudon, K. C., and J. P. Laudon. 2002. *Management Information Systems: Managing the Digital Firm*. Upper Saddle River, NJ: Prentice-Hall Inc.
- 2. Goldman Sachs. 2000. Technology: Internet-Commerce, United States.
- 3. Franck, C., 2000. The New Economy and Agriculture, 2-9. Normandy University.
- 4. Enos, L. 2000. Report: B2B still driving e-commerce. *E-Commerce Times* December 11.
- 5. Kenny, D., and J. F. Marshall. 2000. Contextual marketing. *Harvard Business Review* November-December.
- 6. IDC. 2000. eBusiness trends. IDC Newsletter. Accessed from www.idc.con /newsletter.
- Lynch, J., D. Hickey, T. Grogan, D. Daly, M. Dangerfield, and J. MacNabola. 2000. Irish Farming Online, Agriculture Online: An Industry Analysis. Accessed from www.agriculture.com.
- 8. Dalton, G. 1999. Going, Going, Gone!. Information Week October 4.
- 9. Wilson, P. 2000. An overview of developments and prospects for e-commerce in the agricultural sector. European Commission, Agriculture Directorate—General.
- 10. Mougayar, W. 1998. Operating Digital Markets. New York, NY: McGraw-Hill.
- 11. Shapiro, C., and H. Varian. 1999. *Information Rules*. Boston, MA: Harvard Business School Press.
- 12. Wheatley, W. P., B. Buhr, and D. DiPietre. 2001. E-Commerce in agriculture: Development, strategy, and market implications. Staff Paper P01-6. Minneapolis,

- MN: University of Minnesota, Dept. of Applied Economics. Accessed from agecon.lib.umn.edu.
- 13. Poole, B. 2001. How will agricultural E-Markets evolve? USDA Outlook Forum.
- Porter, M. 2001. Strategy and the Internet. Harvard Business Review 79(2): 63-78.
- 15. Thompson, S. 1996. Potential effects of information technologies on the economic performance of agricultural and food markets. *Symposium on Global Restructuring of Agro-Food Markets: Need for Change in Marketing Policies.*
- 16. Lavouras, S. 2000. Trading on the internet challenges and benefits for buyers. *Proc. of E-Commerce for Agribusiness*.
- 17. Nielsen Media Research. 2001. Various Internet demographic and trend reports. Accessed from www.netratings.com. New York, NY: ACNielsen.
- 18. Ernst, S. 2001. Outlook for IT and E-Commerce in Agriculture. 2002 *Agricultural Policy and Outlook Guide* December, Farm Progress Publications.
- 19. National Agricultural Statistics Service. 1999. Farm Computer Usage and Ownership. Washington, D.C.: USDA.
- 20. Just, D. R., and R. E. Just. 2001. Harnessing the internet for farmers. *Choices* 2<sup>nd</sup> Quarter 2001: 36-40.
- 21. Australian Bureau of Statistics (ABS). 1999-2000. *Use of Information Technology on Farms, Australia*, Cat. No. 8150.0. Canberra, Australia.
- 22. Mueller, R. A. E. 2001. E-Commerce and entrepreneurship in agricultural markets. *American J. Agricultural Economics* 83(5): 1243-1249.
- 23. Ernst, S., and M. Tucker. 2002. "Agri-Culture": A new look at adoption and diffusion on information technology. *International Meeting of Agricultural Communicators in Education*.
- 24. Hopkins, J., and M. Morehart. 2001. Recent findings on farm internet use. Symposium Paper, American Agricultural Economics Association Meeting.
- 25. Goldman Sachs. 1999. B2B: 2B or Not 2B?: An excerpt, September.
- Little, D. 2000. Old MacDonald has a web site. Business Week E.Biz. May 15: EB83-EB88.
- 27. Warren, M. 2003. Farmers and the internet, six years on. *EFITA 2003 Conference*.
- 28. Leroux, N., M. S. Wortman Jr., and E. D. Mathias. 2001. Dominant factors impacting the development of business-to-business (B2B) e-commerce in agriculture. *International Food and Agribusiness Management Review* 4: 205-218.
- 29. Nicolaisen, R. 2001. How will agricultural e-markets evolve? *Proc. of the USDA Outlook Forum.*
- 30. Dolan, R. J., and T. Moon. 1999. Pricing and market making on the Internet. *Harvard Business School Case Study*, No. 9-500-065.
- 31. Moss, L. A. 2001. Who wins and loses and how will e-market affect rural America? *Proc. of the USDA Outlook Forum*.

- 32. Barry, P. J., N. Sotomayor, and L. A. Moss. 1998. Professional farm manager's views on leasing contracts and land control: An Illinois perspective. *J. American Society of Farm Managers and Rural Appraisers* 1998/1999: 15-19.
- 33. Perry G., and L. Robinson. 1999. Personal relationships: Do they influence the sale price of land? *Proc. of the Western Agricultural Economics Association Annual Meeting*.
- 34. Baourakis, G., M. Kourgiantakis, and A. Migdalas. 2002. The impact of e-commerce on agro-food marketing; the case of agricultural cooperatives, firms and consumers in Crete. *British Food J.* 104(8): 580-590.
- 35. Fritz, M., T. Hausen, and G. Schiefer. 2004. Developments and development directions of electronic trade platforms in U.S. and European agri-food markets: Impact on sector organization. *International Food and Agribusiness Management Review* 7(1).
- 36. Montealegre, F., S. Thompson, and J. Eales. 2004. An empirical analysis of the determinants of success of food ad agribusiness e-commerce firms. 2004 International Food and Agribusiness Management, Focus area: Contribution of Technology to Food Chains.
- 37. Vatn, A. 2002. Multifunctional agriculture: Some consequences for international trade regimes. *European Review of Agricultural Economics* 29(3): 309-327.
- 38. Hausen, T., and G. Schiefer. 2003. Electronic trading processes in agrifood markets: Options and experimental experiences. *EFITA 2003 Conference*.
- 39. Hausen, T., and G. Schiefer. 2004. Electronic transaction processes on agrifood markets: Experimental experiences and evaluation. 2004 AFITA/WCCA Joint Congress on IT in Agriculture.
- 40. Schiefer, G. 2003. New technologies and their impact on agriculture, environment and the food industry. *EFITA 2003 Conference*.
- 41. Wilson, P. An overview of developments and prospects for e-commerce in the agricultural sector. European Commission, Agriculture Directorate-General.
- 42. www.LandAndFarm.com
- 43. www.eharvest.com/agchemical
- 44. www.tractorsonline.com
- 45. www.IFI.ie, www.bibbyireland.co.uk, www.gold-dust.com
- 46. www.amconline.com
- 47. www.Rcis.com
- 48. www.foodtrader.com
- 49. www.milkquota.com
- 50. www.breednet.com, www.cattleofferings.com
- 51. www.farms.com
- 52. www.homefarm.com/hfg
- 53. www.pefa.com
- 54. www.agex.com
- 55. www.wineryexchange.com/jsp/index.jsp
- 56. www.fwi.co.uk
- 57. www.hoards.com

- 58. www.agcentralonline.com
- 59. www.forester.com
- 60. www.progressivefarmer.com/weather/default.asp
- 61. www.agrihelp.com
- 62. www.fwi.co.uk/live/markets/mlcfront.html
- 63. www.AgriNet.ie
- 64. www.amconline.co.uk/amc/html/home/frameset.html
- 65. europa.eu.int/comm/dgs/agriculture/index en.htm
- 66. www.oecd.org
- 67. www.wto.org
- 68. www.nfu.org.uk
- 69. Henderson, J., F. Dooley, and J. Akridge. 2000. Adoption of e-commerce strategies for agribusiness firms. *American Agricultural Economics Association Annual Meeting*.
- 70. Judson, B., and K. Kelly. 1999. Hyper wars: Eleven Essential Strategies for Survival and Profit in the Era of Online Business. New York, NY: Touchstone.

# From Production to the User

# 8.1 Food and Raw Materials Storage and Processing

B. Nicolaï, P. Verboven, and C. Franck

Abstract. This chapter outlines how IT is used after harvest and during further processing of agricultural products. Different techniques for advanced planning and scheduling, including MRP2 and more recent techniques based on process models, will be explained and illustrated with an example. Further, it will be shown how process control systems have evolved towards Management Execution Systems (MES), which often also include quality control software covering Hazard Analysis Critical Control Points (HACCP). They are nowadays combined into true Enterprise Resources Planning (ERP) systems that operate at the management level of the company. Advanced simulation software such as Computational Fluid Dynamics (CFD) is also available nowadays for optimizing the design of postharvest and food processes.

**Keywords.** Planning, Scheduling, MRP, ERP, MES, HACCP, CFD, Food, Agricultural product, Processing, Management, IT.

#### 8.1.1 Introduction

Information technology (IT) has become extremely important in industry in general and is implemented both at the factory floor as well as in administration and logistics. The implementation of IT in the agricultural and food processing industry is, however, challenging for many reasons. A first hurdle is the limited storage life of agricultural and food products, so that large inventories are to be avoided and fast stock rotation is mandatory. Also, the large variability of biological materials and the fact that the supply depends on uncontrollable factors (e.g., weather) complicates many processes such as process optimization and scheduling. Further, whereas the processes in most other industries involve simple assembly, in the agro-food industry complicated processing operations that transform the ingredients into a completely different product are often involved. Also, many processes in the agro-food industry are batch processes which, from the IT perspective, are more complicated than continuous processes. Finally, the agro-food industry is quite diverse, and the IT requirements of, for example, an industrial bakery are very different from those of a commercial packinghouse for fresh horticultural products.

Software for food manufacturers aims at coordinating and optimizing all the flows of materials, energy, and financial resources in the company. Friend and Thompson [1] list the following software tools to manage food manufacturing:

- Advanced planning and scheduling (APS) systems assist in optimizing the production with respect to demand forecasts taking into account the limited production resources of the plant;
- Manufacturing execution systems (MES) automate processes and monitoring at the factory floor level;
- Laboratory information management systems (LIMS) automate the dataflow in quality control laboratories;
- Enterprise resource planning (ERP) systems integrate APS and MES systems;
- Supply chain execution (SCE) systems track and manage supply chain activities;
- *Transportation management software* supervises and optimizes the transport of raw materials and final products;
- Product life cycle management tools (PLM) help to develop a marketing concept to a real product; and
- Enterprise application integration (EAI) systems automate the information flow inside the company and to suppliers and customers.

In addition to these production-related systems, much software is currently available for process design, engineering, and optimization.

In this section, an overview of the most important IT components in storage and processing of agricultural products will be given. The emphasis will be on IT solutions for planning, production, and design of food production processes.

# 8.1.2 Advanced Planning and Scheduling

Planning and scheduling involve (1) planning of manufacturing requirements and purchasing of raw materials; (2) planning of the production capacity (availability of processing equipment); (3) detailed production scheduling; (4) recipe management; and (5) production monitoring and, if necessary, rescheduling. Several levels of planning can be distinguished, depending on their scope.

The *strategic plan* involves long-range planning of the activities of the company. It is concerned with translating the business plans and sales plans into long-range production schedules. It enables the planning of resource requirements and material requirements for materials with very long delivery times. It involves high-capital decisions, such as the purchasing of new equipment, hiring new personnel, and construction of new production facilities. Decisions are typically taken by the board of the company based on long-range demand forecasts.

The *tactical plan* usually covers a period of several months and involves the establishment of the *master production schedule* (MPS). The MPS is an overall schedule involving the production of goods, the production capacity, and the purchasing of raw materials.

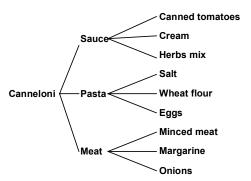


Figure 1. Composition of cannelloni.

The *operational plan* involves the establishment of a detailed production schedule and covers short-time periods, usually in the order of days to weeks. The operational plan includes scheduling of batch operations, physical control of materials such as receiving, shipping, stock management, inventory adaptation to short-term order changes, maintenance planning, purchasing and vendor control, etc.

# Material Requirements and Manufacturing Resources Planning

Independent demand inventory systems are widely in use in the food industry. In such a system, it is assumed that the customer demand is relatively constant. When the inventory has decreased below a certain threshold value, the system automatically places an order at the supplier. Such a system may fail when the inventory drops more sharply than expected and a brief shortage of inventory is caused. For this reason, forecasting procedures have been incorporated in many software packages, but it is clear that the success of an independent-demand inventory system still relies on the predictability of the demand. An even more important drawback of independent-demand inventory systems for food production companies is that it relies on buffer stocks, which can be costly and not always possible, particularly in the case of perishable products.

Most foods are manufactured from food components, which in turn may consist of subcomponents and so on. In Figure 1 the composition of canneloni—a typical Italian pasta product—is shown up to the single ingredient level. Canneloni is assembled from pasta tubes with a meat filling and tomato sauce. The sauce is made of canned tomatoes, cream, and a mix of herbs. This can be continued for every component up to the ingredient level, where ingredients are defined for this purpose as the food components, which are purchased by an external supplier. The complete list of components and subcomponents down to the ingredient level is called the *bill of materials (BOM)*.

MRP (material requirements planning) is a planning method which takes into account the BOM and which does not suffer from the main drawbacks of the independent-demand inventory systems such as the magnitude of the buffers of intermediate products. The MRP algorithm actually relies on a procedure known as explosion of the master production schedule. Suppose that a canneloni order schedule for a period of 8 days is known (Figure 2a), as well as the canneloni stock that is not allocated to a spe-

cific order (Figure 2b). The net canneloni requirements for the 8-day period is then determined from the gross requirements and the available canneloni as shown in Figure 2c. For example, at day 0 the available canneloni (200 + 100 pieces) is sufficient to fulfill the requirements (100 pieces) and there is a remainder of 200 pieces. At day 1 the available canneloni (200 pieces + the 200 remaining pieces from day 0) is still larger than the required amount (300) pieces, and a stock of 100 pieces remains. At day 2, 200 pieces are required, so that the net canneloni requirement is equal to 100 pieces. Because the batch size is 200, one batch must be made available at day 2, and since one batch lasts approximately one day, the production is scheduled for day 1. This in turn implicates that the necessary components (sauce, pasta, cheese) must be available on day 1. For each of the components the above procedure is now accomplished recursively, until the starting dates for the manufacturing of the last subcomponents that are still prepared in the plant are known. Taking into account the supplier lead time, the order dates for the ingredients can now be calculated.

Whereas the original MRP algorithm is essentially limited to material requirements, it was soon recognized that it can readily be extended to planning of production capacity such as labor, machines, capital, purchasing, marketing, and shipping requirements. Such an approach is now called *MRP2* (manufacturing resources planning) to distinguish it from the original MRP. A large variety of packages with MRP2 functionality is now available on the market. Many of them have special features to deal with the particularities of the food industry such as the production of multiple end products from single raw materials (e.g., legs, wings, breast, and liver from chickens), limited keepability of foods, parallel production of components in batches, seasonal availability of raw materials, etc. [2].

#### Scheduling and Logistics

Scheduling involves determining the sequence of operations to take place on each piece of equipment at the plant floor, taking into account its capacity and with the objective of fulfilling the due date constraints specified in the production plan. Factors responsible for the difficulties encountered in scheduling of batch processes in the chemical and food industries include the following [3]:

- cleaning operations required between the manufacture of different products;
- limited raw material resources and production capacity;
- operating preferences for making certain products using certain items of equipment;
- equipment constraints such as size and materials of construction; and
- finding the appropriate balance between inventory level, so as to be able to respond rapidly to customer orders, the storage capacity available, and the cost of storage.

Scheduling software is usually aimed at large food manufacturers, often with more than one plant and many production processes, as the complexity of the scheduling process increases exponentially with the number of different processes and unit operations. Whereas linear integer programming techniques have been used, most modern systems incorporate rule-based systems.

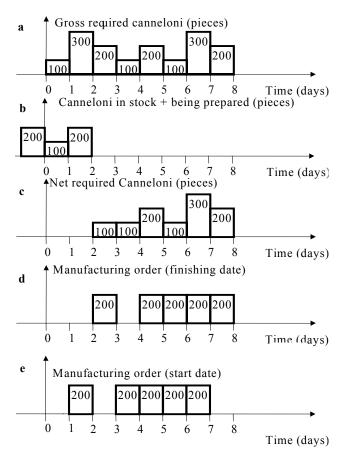


Figure 2. Example of the MRP procedure for determining the production plan of canneloni (after[2]).

The BOM model, on which all MRP2 packages rely, is not always suitable in the agro-food industry. First, the often very short lead times, particularly in areas such as fresh produce handling and processing, require a flexibility to cope with deviations on a very short time horizon, which is often not provided in off-the-shelf MRP2 packages. Further, MRP2 packages often cannot deal with issues such as different grades of ingredients changing the balance of the recipe, or ingredients which must be added after a precise duration of, for example, cooking. Alternative approaches, such as TROPOS (SSI, Basingstoke, UK) rely on a process model for short-term planning, in which inputs and outputs (including materials, catalysts, energy, plant and labour times, machine setpoints, quantities, products, co-products, by-products, recycled material and waste), as well as the quality specifications required by each process, stage,

input and output material, are defined. The process scheduling is then based on this model and stock management is a natural spin off of the production schedule. The system actually plans to the minute.

# 8.1.3 Manufacturing Execution Systems

#### Process Control

Most unit operations in handling and processing of agricultural products and foods are now computer controlled. Unlike many chemical processes, many food processes are batch processes. In 1995 the International Society for Measurement and Control published its S88.01 Batch Control Standard (formerly SP88). The aim of this standard was to define standards and recommended practices for the design and specification of batch control systems as used in the process control industries. Part 1 of the standard covers processes and equipment, batch control concepts including recipes, and batch control activities and functions, which link the equipment and control concepts together. Part 2 further defines the relationships outlined in the control activity model and the data passed between those activities through data modelling. The standard is increasingly being used in the process industry, but in the food process industry only large companies such as Unilever and Nestlé have adapted their batch systems to it [4].

Different approaches are applied when implementing control systems in the food process industry. A *distributed control system* (DCS) provides an integrated software and hardware environment, which runs on a host computer, for process control. Typically, a DCS will have one global database for I/O points that everyone can access. These I/O points are called *tags* and can be mapped to a device on the plant floor or to some internal DCS variable that is then given a meaningful name. In this way the devices can be easily accessed through this name and the operator or engineer does not need to be concerned about the actual details of the device. A wide range of control algorithms, ranging from simple PID control up to advanced adaptive control, are typically provided and the control engineer simply has to define the tuning parameters. Often several operator interfaces are installed both in the control room as well as on the plant floor. Typical DCS suppliers are Emerson Process Management—Fisher Rosemount (Baar, Switzerland) and Honeywell (Morristown, New Jersey, USA).

An alternative approach is that of a combination of *programmable logic controllers* (PLCs) and a user-friendly human interface. A PLC is a hardware controller device which can be directly linked to the production equipment. A variety of control algorithms are usually provided. The advantage of such a system is that PLCs can handle very high-speed I/O operations for discrete manufacturing that could overwhelm some DCS installations. A PLC solution is usually also cheaper than a DCS. However, a typical DCS is usually more stable and reliable, and in general can be much easier integrated in the plant-scale IT systems such as the manufacturing execution system. Typical PLC suppliers are Siemens AG (Munich, Germany) and ABB (Zürich, Switzerland).

PC-based systems are now also often used for the control of food process operations. The system consists of a data acquisition and control package such as Labview

(National Instruments, Austin, Texas, USA) in combination with dedicated acquisition hardware, often a PC card. Such systems are often combined with SCADA (supervisory control and data acquisition) packages, which are aimed at gathering, logging, and reporting information. Libraries for interfacing with PLCs are usually available. Modern SCADA packages such as Scada Portal of ABB (Zürich, Switzersland) provide a broad functionality, including communication with databases systems, PLCs and operators, recipe handling and reporting, and often run in a PC environment. Whereas a PC solution is often cheap in comparison to a DCS or PLC solution, it is to date less reliable and, hence, less often used for critical process control operations.

# **Quality Control Systems**

Quality is very important in the agro-food industry. This industry is one of the few in which the quality of the products may change considerably between production and purchase by the consumer. Also, because of some recent food crises, tracking and field-to-fork traceability is of utmost importance.

Hazard analysis critical control points (HACCP) is a methodology for identifying and controlling product safety hazards. HACCP is internationally accepted and is mandatory in many countries. In the EU it is enforced through Directive 93/43. HACCP is a systematic approach to the identification, evaluation, and control of food safety hazards. Preventing problems from occurring is the paramount goal underlying any HACCP-like system. These systems focus attention on the parts of the process that are most likely to affect the safety of the product [5]. The application of HACCP is normally described in terms of seven principles, which have been formalized by groups such as the Codex Alimentarius of the Food and Agriculture Organisation of the United Nations (www.fao.org):

- conduct a hazard analysis to identify the potential hazard(s) associated with food production at all stages the measures for their control;
- determine the *critical control points (CCPs)*: points, procedures, or operational steps that can be controlled to eliminate the hazards or minimize their likelihood of occurrence;
- establish critical limits, which must be met to ensure the CCP is under control;
- establish a system to monitor control of the CCP;
- establish the corrective action to be taken when monitoring indicates that a particular CCP is not under control;
- establish procedures for verification to confirm that the HACCP system is working effectively; and
- establish documentation concerning all procedures and records appropriate to these principles and their application.

It is estimated that in the US currently more than two-thirds of every industry segment outside of the meat, poultry, and seafood industries have voluntarily implemented HACCP [6].

The implementation of a HACCP plan involves a large administrative effort. Many software packages are now available to reduce this burden. A good example is the

HACCP documentation software of CCFRA (Chipping Campden, UK, www. campden.co.uk).

# Manufacturing Execution Systems (MES)

Manufacturing execution systems (MES) solutions are a next logical step in plant-wide automation schemes [7]. MES provide a level of control between process control and MRP2 and addresses all plant resources including material, equipment, personnel, process instructions, and facilities. An integrated MES typically consists of a relational database management system (RDBMS), computer-aided systems engineering (CASE) tools for application development, a document management system, and interfaces to other IT systems in the plant. Typically scheduling, supervisory monitoring and control, and quality management are provided as well.

# 8.1.4 Laboratory Information Management System (LIMS)

Laboratory information management systems (LIMS) are an example of laboratory informatics, defined as: the specialized application of information technology to optimize and extend laboratory operations. LIMS are specifically designed for analytical laboratories, including R & D and quality assurance labs, to record and organize a large collection of data for rapid search and retrieval [8-9]. LIMS connect the analytical instruments used to collect data to one or more workstations or personal computers, where the data is organized into meaningful information and various report formats. In contrast with databases, which are solely used to archive analyzed data, LIMS contains additional information (metadata) about the early stage of data acquisition, like the biological source, administration (personnel), sample handling, preparation, etc.

Originally, LIMS were developed in-house by organizations wishing to streamline their data acquisition and reporting processes. These in-house LIMS could take considerable time and resources to implement since they were custom-built systems. Parallel to these custom-built LIMS implementations, the first commercial LIMS products were created in the 1980s. Such commercial LIMS were proprietary systems, often developed by analytical instrument manufacturers. Now, many LIMS are highly flexible and take advantage of platforms to offer client/server capabilities and enterprise-wide access to lab information. Web-based LIMS are also available, e.g. the nuclear magnetic resonance group of Kodak has provided worldwide access to spectroscopy-based data by using web technology to implement a platform-independent web-based information management system (WIMS) on their intranet [10].

ThermoLabsystems (UK) is one of the world's leading suppliers of LIMS. Recently, they developed Nautilus LIMS<sup>TM</sup> according to the ISO9001/TickIT standards and exclusively designed for Microsoft Windows® 95/98, NT, and 2000 environments [11]. Nautilus is a ready-to-use package and has been selected by Molkerei Alois Müller—the largest independent yoghurt manufacturer in the UK with over 30% market share and market leader in Europe—to manage quality control data for raw materials and finished dairy desserts [12]. It assists in real-time monitoring of the production processes and plays a pivotal role in ensuring quality control for finished products. By using Nautilus, the amount of paperwork is drastically reduced and testing is speeded

up. The ability to trend all data gives the company the opportunity to make necessary improvements in laboratory productivity.

# 8.1.5 Enterprise Resources Planning (ERP) Systems

A 2001 survey showed that a large number of food manufacturers are devoting nearly the same resources to integrating their manufacturing systems with their IT systems as they are to integrating their different IT systems with each other [13]. The scope of such integration efforts can be limited to the integration of two or more different functionalities. For example, most MRP2 packages now include a scheduling utility as well or can be interfaced with scheduling packages from other suppliers. However, recently there has been a tendency towards full integration of most logistic IT processes in the company. Such systems are called *enterprise resources planning (ERP)* systems. Further, the recent ISA-S95 standard aims at integrating ERP systems with control systems like DCS and SCADA [14]. Also, many vendors of ERP systems are beginning to use web technology for their information exchange through the XML (Extensible Markup Language) standard.

The TROPOS (SSI, Basingstoke, UK) ERP solution is an example of an integrated system which has been implemented in several food companies and which is especially aimed at fast-response manufacturing operations. It is based on a process model rather than a typical BOM-based MRP2 approach. It incorporates components for financial ledgers, demand forecasting, advanced planning and scheduling, inventory optimization, export documentation, and many other business functions. A partnership with a data-logging equipment manufacturer enables live shop-floor data to be exported to planning and control systems, providing major benefits by providing up-to-the-minute information from which accurate delivery promises can be made, and recording highly detailed quality-control data. The same company also has a SCADA software product (SSI-DACS) that can interface with the ERP system.

Several ERP and MES vendors have linked their process control software with HACCP software [15]. The principle is that CCP monitoring is performed by the process control system and that this information is shared with the HACCP software. Often, tight integration with the LIMS software of the quality control laboratory is provided as well. Examples are the CSB-System (Geilenkirchen, Germany) and Nicheware Systems (Birmingham, Atlanta, USA).

# 8.1.6 Process Design Software

# Computer-Aided Food Process Design

Although computer-aided food process design has been advocated for several decades in the scientific literature, only recently it has been applied in practice. This is quite unlike the chemical process industry where computer-aided process design has been used for many years. This is due to several factors. In the chemical process industry most processes are conducted in *continuous* mode and involve large product quantities. The reactions which take place are typically relatively simple and involve only a few components at the same time. Moreover, the reaction kinetics are usually well understood and the physical parameters of the products involved are known. In

the food industry, on the contrary, the *batch* mode is the prime mode of production. The products involved are often complex food materials and involve a large number of ingredients and intermediate processes. The reaction kinetics are very complicated and often unknown and the physical parameters are very variable between different batches and are usually unknown.

The industrial design of food processes has evolved as a highly specialized field based on diverse branches of science including biochemistry, process engineering, microbiology, heat and mass transfer, mathematics, etc. Several attempts have been made in the past towards the development of software to simplify this complex design task. Some important advances in the area of thermal food processes will be described below

The design of thermal sterilization processes is well established and is based on the analysis of the heat penetration in the sterilized food and the kinetics of thermal inactivation of micro-organisms. For conduction-heated foods, the use of a mathematical model (the Fourier equation) to predict the temperature inside the can has been suggested as a flexible alternative to actual temperature measurements. Unfortunately, analytical solutions are known only for relatively simple problems involving generic product geometries (slab, sphere, cylinder), temperature-independent thermophysical properties, and relatively simple boundary conditions. For this reason, Teixeira et al. [16] suggested solving the Fourier equation numerically by means of the finite difference method, and to use the computed center temperature as an input for the calculation of the process lethality by numerical integration. This was implemented in a program that must be considered one of the first examples of computer-aided food process design software. As a further improvement, the use of time-varying retort temperature profiles was considered [17] in order to maximize the retention of thiamin while safeguarding the required process value. This eventually led to the STERILMATE software package for computer-aided design of sterilization processes [18]. More elaborated computer-aided optimization procedures have been described in [19] and [20].

The first attempts to extend this approach to convection heating of cans by means of *computational fluid dynamics (CFD)* were described in [21] and [22]. CFD is the simulation of fluid flow by means of a mathematical model solved on a computer. It serves as a basis to simulate heat and mass transfer in flow systems encountered in other food processes as well, such as continuous sterilization of liquid foods [23,24] and heating of foods in ovens by hot air [25,26]. It is equally applicable to chilling applications, such as air chilling of bulk foods in cold-air rooms [27-31]. An example of a CFD calculation of the airflow in a cool room is displayed in Figure 3. Over the last decades, CFD packages have evolved to a user-friendly, versatile, and interactive software environment that can be run on most computer platforms. CFD has now become the engineering tool for the design of fluid flow processes, reducing costs and design cycles in many fields, from the aerospace to the automobile industry, from

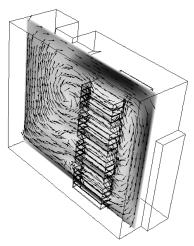


Figure 3. Velocity magnitude (shaded contours) and flow direction (arrows) at plane z = 0.9 m in a loaded cool store. Velocity ranges from 0 to 2 m·s<sup>-1</sup> corresponding to white and black, respectively. Reprinted from *International Journal of Refrigeration*, Nahor et al. [31], CFD model of the airflow, heat and mass transfer in cool stores, p. 376. Copyright 2005, with permission from Elsevier.

meteorology to biomedical engineering, from the chemical to the food industry. The following features render CFD very powerful for complex processes: the geometry of objects can be simulated with high resolution by means of CAD import facilities (with the last generation of computer codes) and exact food shapes can be introduced when computer vision systems are combined with the pre-processing modules of CFD codes.

A wide range of additional physical models is available in CFD packages. In porous media, a detailed geometrical model of air or fluid voids in between particles is often not feasible and the mesh volumes will be larger than the smallest void size. The complex flow in the porous medium then needs to be represented on a superficial scale by an empirical model that relates superficial fluid velocity to the pressure gradient across the bulk of material. It has been applied to bulks of foods (in, e.g., [32,33]). Liquid-solid foods contain particles that affect the fluid flow [34]. The particles then become a separate phase, which can be described using continuum (Eulerian) or discrete (Lagrangian) models. Dispersion of fluid droplets in air has been modelled in [35,36].

Because it is based on fundamental conservation laws, the Navier-Stokes equations, CFD offers a versatile tool for process design and optimization. The more the models contain empirical features, the more the simulations lose their versatility and cannot be directly transferred from one application to another. The more parameters in the models have physical meaning (such as thermophysical fluid properties), the more the model becomes universally applicable over a wide range of processes. It has been mentioned before that physical properties of foods are very variable and not always well documented. On the aspect of uncertainty, food process design software in gen-

eral lacks features, because most of the software originates from other disciplines where uncertainty is much less a factor to be dealt with.

The computational power of computers and fast, accurate numerical solution algorithms lay at the basis of the success of CFD in a wide variety of industries, including food manufacturing and processing [37,38]. Although CFD computer codes offer user-friendly desktop interfaces that allow a step-by-step definition and analysis of the problem, the user should have sufficient skill to judge the accuracy of the approximations and of available models, which may be strongly empirical.

Much more than a calculation of the time-temperature distribution, the food industry is interested in quality and microbial changes. The ChefCad package [39-41] was developed to provide an environment to define food recipes and to simulate the effect of unit operations such as heating and cooling on the microbial as well as the organoleptic quality of foods. It contained finite element routines for the numerical solution of 2D heat conduction problems, an automatic finite element grid generator, routines to calculate the thermophysical properties from the chemical composition of the food, routines to calculate the surface heat transfer coefficient of the heating/cooling fluid, differential equation solvers for the microbial growth/inactivation and texture changes. It also incorporated an expert system for microbial safety diagnosis of the recipe. The expansion of this functionality to more complex heat and mass transfer processes involving convection, radiation and microwave heating is still in its infancy.

# Flow-Sheeting

The main purpose of flow-sheeting packages is to model steady-state material and energy balances during the operation of complicated chemical processes. The model can then be used for process optimization by evaluating the effect of changing some process conditions on the product yield, waste production, and energy consumption of the process. Most flow-sheeting packages actually consist of a language interpreter or compiler, a material properties database, models for different unit operations such as distillation columns, heat exchangers etc., a calculation routine, optimization algorithms, and reporting facilities.

So far the use of flow sheeting in the food industry has been limited. Drown and Petersen [42] described several examples of flow sheeting using the GEMS package. The GEMS package [43] was originally developed for the paper and pulp industry. It incorporates a variety of blocks describing unit operations for the paper and pulp industry, and an executive program to connect the blocks and keep appropriate records. In a first example, the recovery of peel-oil from oranges during juice extraction was simulated. Alternative schemes were compared in order to minimize the operating problems encountered in the original set-up. In a second example, the performance of a potato-blanching process was optimized for controlling sugar content while minimizing energy consumption. In a third example, single and multiple zone drying systems were optimized by minimizing energy consumption, increasing production capacity, and improving product uniformity.

BATCHES from Batch Process Technologies (Purdue, Indiana, USA) is a simulation system designed to allow a user to build models of multi-product batch/semi-

continuous processes as typically found in the biochemical, food, pharmaceutical, and specialty chemical industries. Typical applications include design and scheduling of beverage manufacturing plants, evaluation of energy integration alternatives in dairy processes, etc.

Discrete-event simulation packages are becoming popular for the simulation of production lines where many discrete entities are processed, such as the filling of bottles. One of the most successful is the WITNESS package (Lanner, Redditch, UK). A production line can be assembled intuitively on the computer screen by pointing and clicking. A mixture of discrete/continuous elements can be used and automatic performance reporting is provided. Although most applications are situated in the automotive, electronics, and pharmaceutical industries, WITNESS has been used successfully in the food industry, for example to evaluate the need for installing a new filter and canning line in a brewery. Nahor et al. [44] developed a discrete/continuous time model of controlled atmosphere (CA) cool room storage systems. The model consisted of three interacting sub-models for the prediction of the transient behavior of the processes in the three units, namely, the cool room, the refrigeration system, and the gashandling unit. The modules were implemented in an object-oriented discrete/ continuous time computational environment (EcosimPro, Madrid, Spain). The handling of discrete/continuous events enabled the implementation of practical operational procedures and investigation of their implication directly on the product quality and plant performance/design.

#### 8.1.7 Conclusions

The rapid growth of IT has revolutionized the handling and processing of agricultural products and foods. IT is now introduced at all levels in the company from the factory floor to the management office. Obviously, the complexity of the IT systems have increased correspondingly, and IT is now a major capital investment which has to be planned carefully. It can readily be envisaged that in the future even a more tight integration between all IT functionalities will be realized.

#### References

- 1. Friend, B., and O. Thompson. 2003. The essential guide to manufacturing software. *Food Engineering*, 16/04/2003. Accessed on June 17, 2003, from http://www.foodengineeringmag.com.
- 2. Nicolaï, B. M. 1996. *Computer Integrated Manufacturing in the Food Industry*, ed. G. Mittal. New York, Basel, Hong Kong: Marcel Dekker, Inc.
- 3. Sawyer, P. 1993. *Computer-Controlled Batch Processing*. Rugby, UK: Institution of Chemical Engineers.
- 4. Higgins, K. T. 2003. Better control of batch. *Food Engineering*. Accessed on March 30, 2003, from http://www.foodengineeringmag.com.
- 5. FAO. Accessed on July 25, 2003, from http://www.fao.org/docrep/w8088e/w8088e00.htm.

- 6. Morris, E. 2003. HACCP integrations. *Food Engineering*. Retreived on March 22, 2003, from http://www.foodengineeringmag.com.
- 7. Deuel, A. C. 1994. The benefits of a manufacturing execution system for plant wide automation. *ISA Transactions* 33(2): 113-124.
- 8. Limsource. http://www.limsource.com/intro.html.
- 9. Leung Y. F. http://ihome.cuhk.edu.hk/~b400559/arraysoft\_database.html.
- 10. http://www.elsevier.nl/homepage/saa/trac/wimsarti.htm.
- 11. Thermo LabSystems. http://www.labsystems.com/products/nautilus.
- 12. Laboratorytalk. http://www.laboratorytalk.com/news/the/the103.html.
- 13. Morris, E. 2003. HACCP under the microscopy integrations. *Food Engineering*. Accessed March 30, 2003, from http://www.foodengineeringmag.com.
- 14. Verwater-Lukszo, Z., M. E. van Wissen, and F. Verhofstad. 2003. Improved integration of enterprise and control level with combining ISA batch standards and process models. *Focapo* 2003.
- 15. Morris, C. E. 2003. HACCP Integrations. *Food Engineering*. Accessed on June 17, 2003, from www.foodengineering.com.
- 16. Teixeira, A. A., J. R. Dixon, J. W. Zahradnik, and G. E. Zinsmeister. 1969. Computer optimization of nutrient retention in thermal processing of conduction-heated foods. *Food Technology* 23(6): 137.
- 17. Teixeira, A. A., G. E. Zinsmeister, and J. W. Zahradnik. 1975. Computer simulation of variable retort control and container geometry as a possible means of improving thymine retention in thermally processed foods. *J. Food Science* 40: 656.
- 18. Kim, K. H., A. A. Teixeira, J. Bichier, and M. Tavares. 1993. STERILMATE: Software for designing and evaluating thermal sterilization processes. ASAE Paper No. 93-4051. St. Joseph, MI: ASAE.
- 19. Banga, J. P., J. M. Perez-Martin, J. M. Gallardo, and J. J. Casares. 1991. Optimization of the thermal processing of conduction-heated canned foods: Study of several objective functions. *J. Food Engineering* 14: 25.
- 20. Silva, S. M., M. Hendrickx, F. Oliveira, and P. Tobback. 1992. Critical evaluation of commonly used objective functions to optimize overall quality and nutrient retention of heat-preserved foods. *J. Food Engineering* 17: 241.
- 21. Engelman, M. S., and R. L. Sani. 1983. Finite-element simulation of an in-package pasteurization process. *Numerical Heat Transfer* 6: 41-54.
- 22. Datta, A. K., and A. A. Teixera. 1988. Numerically predicted transient temperature and velocity profiles during natural convection of canned liquid foods. *J. Food Science* 53: 191-195.
- 23. Kumar, A., and M. Bhattacharya. 1991. Transient temperature and velocity profiles in a canned non-Newtonian liquid food in a still-cook retort. *International J. Heat and Mass Transfer* 34: 1083-1096.
- 24. Jung, A., and P. J. Fryer. 1999. Optimising the quality of safe food: Computational modelling of a continuous sterilization process. *Chemical Engineering Science* 54: 717-730.

- 25. Verboven, P., N. Scheerlinck, J. De Baerdemaeker, and B. M. Nicolaï. 2000. Computational fluid dynamics modelling and validation of the temperature distribution in a forced convection oven. *J. Food Engineering* 43(2): 61-73.
- 26. Verboven, P., A. K. Datta, T. A. Nguyen, N. Scheerlinck, and B. M. Nicolaï. 2003. Computation of airflow effects on heat and mass transfer in a microwave oven. *J. Food Engineering* 59(2-3): 181-190.
- 27. Tassou, S. A., and W. Xiang. 1998. Modelling the environment within a wet air-cooled vegetable store. *J. Food Engineering* 38: 169-187.
- 28. Xu, Y., and D. Burfoot. 1999. Simulating the bulk storage of foodstuffs. *J. Food Engineering* 39: 23-29.
- 29. Xu, Y., and D. Burfoot. 1999. Predicting condensation in bulks of foodstuffs. *J. Food Engineering* 40(1-2): 121-127.
- 30. Hoang, M. L., P. Verboven, M. Baelmans, and B. M. Nicolaï. 2003. A continuum model for air flow, heat and mass transfer in bulk of chicory roots. *Trans. ASAE* 46 (6): 1603-1611.
- 31. Nahor, H. B., M. L. Hoang, P. Verboven, M. Baelmans, and B. M. Nicolai. 2005. CFD model of the airflow, heat and mass transfer in cool stores. *International J. Refrigeration* 28: 368-380.
- 32. van der Sman, R. G. M. 2002. Prediction of airflow through a vented box by the Darcy-Forchheimer equation. *J. Food Engineering* 55(1): 49-57.
- 33. Lareo, C., P. J. Freyer, and M. Barigou. 1997. The fluid mechanics of two-phase solid-liquid food flows: A review. *Trans. Institute of Chemical Engineers (Part C)* 75(2): 73-105.
- 34. Reynolds, A. M. 1997. Modelling particle dispersion within a ventilated airspace. *Fluid Dynamics Research* 22(3): 139-152.
- 35. Duchanoy, C. and T. R. G. Jongen. 2003. Efficient simulation of liquid-solid flows with high solids fraction in complex geometries. *Computers and Fluids* 32: 1453-1471.
- 36. Hu, H. H. 1996. Direct simulation of flows of solid-liquid mixtures. *International J. Multiphase Flow* 22(2): 335-352.
- 37. Scott, G., and P. Richardson. 1997. The application of computational fluid dynamics in the food industry. *Trends in Food Science and Technology* 8: 119-124.
- 38. Xia, B., and D. W. Sun. 2002. Applications of computational fluid dynamics (CFD) in the food industry: A review. *Computers and Electronics in Agriculture* 34: 5-24.
- 39. Schellekens, M., T. Martens, T. A. Roberts, B. M. Mackey, B. M. Nicolaï, J. F. Van Impe, and J. De Baerdemaeker. 1994. Computer aided microbial safety design of food processes. *International J. Food Microbiology* 24: 1.
- 40. Nicolaï, B. M., J. F. Van Impe, and M. Schellekens. 1994. Application of expert systems technology to the preparation of minimally processed foods: A case study. J. A Benelux Q. *J. Automatic Control* 35: 50.
- 41. Nicolaï, B. M., W. Obbels, M. Schellekens, B. Verlinden, T. Martens, and J. De Baerdemaeker. 1994. Computational aspects of a computer aided design package

- for the preparation of cook-chill foods. *Proc. of the Food Processing and Automation Conference III*, 190. St. Joseph, MI: ASAE.
- 42. Drown, D. C., and J. N. Petersen. 1985. Computer aided design in the food processing industry. II. Application. *J. Food Technology* 20: 407.
- 43. Edwards, L., and R. Baldus. 1972. GEMS—General Energy and Material Balance System: A modular computer system for pulp and paper applications. Moscow, ID: Idaho Research Foundation Inc. (Cited by Petersen and Drown, 1985.)
- 44. Nahor H. B., N. Scheerlinck, P. Verboven, J. Van Impe, and B. M. Nicolaï. 2004. Continuous/discrete simulation of controlled atmosphere (CA) cool storage systems: Model development and validation using pilot plant CA cool storage. *International J. Refrigeration* 27: 884.

# 8.2 Quality Issues for Agricultural Product Chains

K. P. Ferentinos, I. K. Kookos, K. G. Arvanitis, and N. A. Sigrimis

Abstract. This section presents some theoretical background on supply chains and supply chain management, examines the role of emerging technologies in supply chain management, and provides some information on actions that should be taken to assure environmentally conscious supply chains in agriculture. Then, focus is given to the way supply chain management is applied in agribusiness. Finally, the latest technological advancements on traceability and tagging systems in agricultural processes are presented.

Keywords. Supply chain management, Agribusiness, Traceability, Tagging systems.

#### 8.2.1 Introduction

In the past, organizations have operated independently by taking strategic decisions within a facility, without taking into account component dependencies and interactions along the supply chain. The *supply chain* concept arose from a number of changes in the manufacturing environment such as increased manufacturing cost, shrinking resources, shortened product life cycles, and the globalization of market economies [1]. In fact, the competitive realities of the current marketplace and the impact on the contemporary operations manager are aptly summarized by Skinner [2]: "make an increasing variety of products, on shorter lead times with smaller runs, and flawless quality. Improve the ROI (return on investment) by automating and introducing new technology in processes and materials so that prices can be reduced to meet local and foreign competition. Mechanize—but keep the schedule flexible, inventories low, capital cost minimal, and the work force contented." The advent of *total quality management (TOM)* to achieve consistent and flawless quality, *flexible manufacturing systems* 

(FMS) to achieve quick response and agile manufacturing at reasonable cost, and *sup-ply chain management (SCM)* mechanisms to deliver products quickly with low inventories are regarded as responses to these new competitive pressures [3].

A supply chain, according to Stevens [4], is a system whose constituent parts include material suppliers, production facilities, distribution services, and customers, linked together via the feedforward flow of materials and the feedback flow of information (Figure 1). In other words, it is a network of multiple businesses and relationships. It can be described in terms of five interconnected business systems: (1) engineering, (2) marketing, (3) manufacturing, (4) logistics, and (5) management systems.

According to Beamon [1], the supply chain involves two basic and highly integrated processes: (1) the production planning and inventory control process and (2) the distribution and logistic process. The production planning and inventory control process encompasses the manufacturing and storage processes as well as their interfaces. Production planning describes the design and management of the entire production process while inventory control concerns the design and management of the storage policies and procedures. The distribution and logistic process determines how products are retrieved and transported from the warehouses to distribution facilities and finally to retailers. An integrated supply chain is obtained by the extensive interaction of these two processes.

With the increasing complexity of supply chains today, as global markets continue to open and more firms are becoming multinational, independently managing facilities can result in very poor overall behavior [5]. It is now clear that focusing on a single element in the supply chain cannot assure the effectiveness of the whole system. The quality of data and the high complexity of the supply chain make the overall formulation of the supply chain problem very difficult and the use of an integrated supply chain management completely necessary.

The term *supply chain management* was originally introduced by consultants in the early 1980s [6]. There are several definitions of SCM. A good number appear in [7]. According to the Global Supply Chain Forum, SCM is the integration of key business

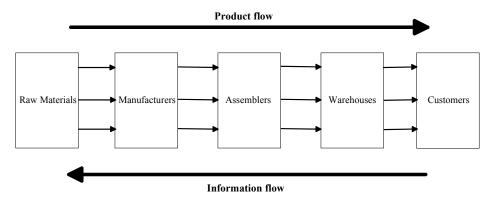


Figure 1. Schematic representation of a generic supply chain.

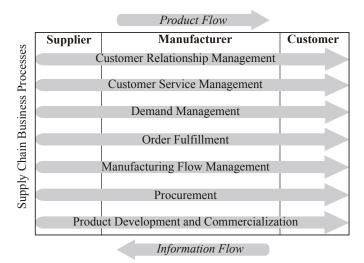


Figure 2. Integrating and managing business processes during SCM across the supply chain.

processes from end user through original suppliers that provides products, services, and information that add value for customers and other stakeholders. To this extent, business processes become supply chain business processes linked across intra- and inter-company boundaries [8] (Figure 2). This means that all activities are managed at the inter-organizational level as well as the departmental level. Instead of focusing on the management of interfirm inventory and transportation capacities, SCM aims to integrate the activities of an entire set of organizations from procurement of material and product components to deliver completed products to the final customer [9].

SCM has evolved in recent years to reflect the fundamental changes that have occurred in the relationship between marketing and corporate strategy. It leads to improvements in channel performance among all channel members and not solely within the focal firm. These improvements are due to the following [10]:

- avoidance of duplication effects by concentrating on core competencies;
- use of inter-organizational standards like *activity based costing (ABC)* and *electronic data interchange (EDI)*; and
- elimination of unnecessary inventory levels by postponing customization towards the end of the supply chain.

The key element of SCM is activity integration. It is a strategy that brings together the application of logistics and its focus on transactions between channel members with that of management and its focus on relationships within the channel. It seeks to achieve a relationship of mutual benefit by defining the organizational structures and contractual relationships between buyer and seller.

#### 8.2.2 Theory and Principles of SCM

A critical literature review analysis on the SCM field [7] concluded that there is a relative lack of theoretical work in the field, compared to empirical studies. A concep-

tual framework of SCM tries to emphasize the interrelated nature of SCM and the need to proceed through several steps to design and successfully manage a supply chain [8]. It consists of three closely interrelated elements: the supply chain network structure, the supply chain business processes, and the supply chain management components. The network structure consists of the member firms and the links between these firms. Business processes are all the activities that produce a specific output of value to the customer. The management components are the managerial variables by which the business processes are integrated and managed across the supply chain.

In a supply chain, all firms, from the raw materials to the ultimate consumer, participate. The degree of the required management of this chain depends on several factors; the most important ones are the complexity of the product, the number of available suppliers, and the availability of raw materials. The dimensional complexity of the structure is given by the length of the supply chain and the number of suppliers and customers at each level. The closeness of the relationship at different points in the supply chain may differ. Management needs to choose the level of partnership appropriate for particular supply chain links [11]. The most appropriate relationships are the ones that best fit the specific set of circumstances [12].

The performance of a supply chain is influenced by the structure of business processes, information systems, and decision support rules, as well as the nature of collaboration between supply chain members. If the supply chain has not been structured properly, as measured by its physical attributes, little can be done to repair the resulting "damage." If the supply chain infrastructure has lengthy and variable lead times, poor understanding of customer demand patterns, poor product quality, or uncertain production capacity, then little competitive advantage can be achieved through more extensive adoption of information systems, decision support tools, or efforts to collaborate with partners. Thus, competitive advantage will exist only if several key elements exist in a supply chain. These include knowing the customer, constructing a lean supply chain organization that eliminates waste, viability and uncertainty, building tightly coupled information infrastructures and business processes, and finally constructing tightly coupled decision support systems.

#### 8.2.3 Major Forms of SCM

Supply chain management systems in use today have evolved over time and often are based on operations research modeling paradigms. The two most popular forms of SCM are summarized below, while modeling approaches are presented in Section 8.2.4

# Efficient Consumer Response (ECR)

One of the quite recent and most important strategies of SCM is *efficient consumer* response (ECR). It is defined as "a grocery industry strategy in which distributors, suppliers and brokers jointly commit to work closely together to bring greater value to the grocery consumer" [13]. This SCM approach aims to meet the goal of better fulfillment of consumer needs via the implementation of a four-part process: efficient

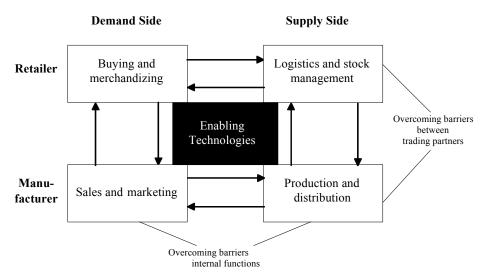


Figure 3. ECR harmonization process by focus areas [14].

replenishment, efficient promotion, efficient store assortment, and efficient product introduction [10]. The realization of ECR strategy requires the use of enabling technologies such as information systems and improved business processes [14]. As a result, harmonization among the various channel members is achieved (Figure 3). The four focus areas (sales and marketing, production and distribution, buying and merchandizing, logistics and stock management) that are shown in Figure 3 are supposed to be installed within the participating companies. They should be interpreted as interorganizational and interdepartmental working groups, thus their implementation suggests the loss of functional and organizational borders within and between firms [10]. In this way, financial and procedural waste from the channel is eliminated and team members are encouraged to work for an increase in performance of the entire channel.

ECR strategy has given positive results in general, even though in various cases more time is needed for the results to be realized. It is true that more empirical evidence is needed on the organizational level and new research should be focused on the evaluation of results from ECR and its influence on the overall performance of the firm

# Just In Time (JIT)

Just in time (JIT) is a management philosophy of continuous improvement that aims to bring certainty and smoothness to the flow of materials through the supply chain by identifying and eliminating wasteful practices and activities, such as holding safety stocks. Businesses hold stocks because of uncertainty, either about the future level of demand or about the lead time to manufacture or replenish stocks. What the JIT approach tries to develop is a network of quality-assured supply partners who can deliver the right quantity to the right place at the right time, every time. The supplies

are delivered against an agreed schedule with absolute certainty on the day they are required, rendering expensive safety stocks redundant.

The American Production and Inventory Control Society (APICS) has the following definition of JIT: "A philosophy of manufacturing based on planned elimination of all waste and continuous improvement of productivity. It encompasses the successful execution of all manufacturing activities required to produce a final product, from design engineering to delivery and including all stages of conversion from raw material onward. The primary elements include having only the required inventory when needed; to improve quality to zero defects; to reduce lead time by reducing setup times, queue lengths and lot sizes; to incrementally revise the operations themselves; and to accomplish these things at minimum cost."

The main purposes of JIT implementation are reducing the cost, improving delivery time, improving quality and performance, adding flexibility, and increasing innovativeness. Some of the prominent types of waste to be eliminated are waste of overproduction and waiting time, transportation, inventory, and processing waste, as well as waste of motion and waste from product defects. When its principles are implemented successfully, significant competitive advantages are realized. However, there are several requirements for the successful implementation of JIT. Among these requirements are:

- improvement of sales forecasts and, where appropriate, production planning so that both purchasing and suppliers can be better informed about requirements;
- set up of effective information systems so suppliers are immediately aware of any changes to programs;
- a quality assurance program under which suppliers are accepting responsibility for quality, monitoring quality during rather than after production, and working towards zero defects; and
- removing non-value-adding activities throughout the whole supply chain, i.e., looking at the total cost picture.

# 8.2.4 Modeling the Supply Chain

The generic supply chain shown in Figure 1 consists of five stages. Each stage can be considered as a source for the stages that follow or as a sink for the stages that precede. Each stage is characterized by its [15]:

- location,
- demands for products and raw materials,
- cost associated with each product or process within the stage,
- customer service requirements,
- technical, legal, or operational constraints,
- capacity constraints (maximum and minimum by product and location),
- maximum inventory investment constraints, and
- order handling requirements and constraints.

Multi-stage models for supply chains design and analysis can be categorized into *descriptive* models or *optimization* models [16]. Descriptive models are developed by modeling practitioners in order to improve their understanding about the functional

relationships in the company and the outside world. Descriptive models include the following [16]:

- forecasting models, to predict demand or raw material cost based on historical data:
- cost relationships, which describe how costs vary as functions of cost drivers;
- resource utilization relationships, which describe how scarce resources are consumed: and
- simulation models, which help evaluate certain circumstances and conditions
  where real data are not available, or used as models for construction of simulated data

Optimization models are developed by modeling practitioners in order to help managers make better decisions. Clearly, the development of optimization models requires data and models as inputs and, consequently, descriptive models are used for this task.

Min and Zhou [17] have classified supply chain models into four major categories: deterministic (non-probabilistic), stochastic (probabilistic), hybrid, and IT-driven. In deterministic models it is assumed that all parameters and data are known while in stochastic models uncertainty is involved in this information. In a recent paper Melachrinoudis and Min [18] present a representative deterministic modeling approach together with an enlightening discussion of the multifaceted nature of the supply chain systems modeling. Lee et al. [19] present a representative stochastic modeling approach by developing a dynamic programming model that aimed to minimize the expected cost of production, inventory holding, and excess demand penalty, subject to production satisfying capacity constraints. Hybrid models involve elements of both stochastic and deterministic models. A representative example is given in [20]. They use a combination of mixed-integer programming models and simulation to determine the number and location of distribution and processing centers as well as the set of market areas covered by each distribution center.

IT-driven models aim to integrate and coordinate various phases of supply chain planning on a real-time basis using application software. The aim is to enhance visibility throughout the supply chain. Models in this category include, among others, warehouse management systems (WMS), transportation management systems (TMS), distribution resource planning (DRP), and geographic information systems (GIS). A representative example is given in [21].

# Supply Chain Performance Measures

Apart from the need for a model that describes the behavior of the supply chain there is also the need to evaluate its performance. Performance measures are used to determine the efficiency and the effectiveness of an existing system, and more importantly to compare competing alternative systems in order to design proposed systems by determining the values of the decision variables. The most commonly used measures of performance can be categorized in *qualitative* and *quantitative* performance measures

Beamon [1] defines qualitative performance measures to be the measures for which there is no single direct numerical measurement. *Customer satisfaction* (the degree to

which the customers are satisfied with the product) is among the most important qualitative performance measures and is comprised of three elements: pre-transaction satisfaction (prior to product purchase), transaction satisfaction (services related to the physical distribution of products), and post-transaction satisfaction (related to the use of products). Supply chain *flexibility* (degree of response to random fluctuation in the demand) and *risk management* (minimizing the effects of the risk inherent to all relationships within the supply chain) are also among the most important qualitative performance measures.

Most of the quantitative performance measures are directly related to the cost or profit of the supply chain. The minimization of cost, the minimization of inventory investment, the maximization of sales or the maximization of profit (revenues – cost) are the performance measures related directly to the cost or profit. Other quantitative measures are measures based on customer responsiveness. Maximizing the fraction of customer orders filled on time and minimizing the time between the placement of an order and the delivery of the product are among the more commonly used performance measures not directly related to cost or profit.

# Design Variables in Supply Chain Modeling

The models and the performance criteria discussed above can be integrated in order to help obtain decisions related to the structure and the operation of the supply chain system. These decisions are expressed in terms of the design variables of the supply chain and are related to the activities of supply chain management that can be classified into three levels [22]: operation level, tactical level, and strategic level. The design variables can be classified as [23] design variables at configuration level or design variables at operational management and control level. Variables that belong to the configuration level are normally related to the tactical and strategic level activities and determine the configuration (topology) of the supply chain with regard to the parties involved, roles to be performed, manners of co-operation, constraints to executing roles, and the IT or physical infrastructure used. Variables that belong to the operational management and control level are normally related to the tactical and operational level activities and determine, for a given topology, the co-operation and integrated planning of operations.

The most important decision variables that are functionally related to the performance of a supply chain are related to:

- *location decisions*—the number, size and physical location of the production units, warehouses and distribution centers;
- *production decisions*—the allocation of suppliers to plants, products produced at each plant, etc.;
- inventory decisions—the management of the inventory levels; and
- transportation decisions—the type of transportation means, size of shipments, etc.

#### **Optimization Models and Their Solution**

A model can be used as a means for describing the essential features of the problem under consideration. The completeness and precision that a mathematical model offers, as well as the understanding involved in its development, make models (and modeling development) invaluable tools in improving our understanding of complex, interacting, and highly integrated systems such as the supply chain. The greatest value of supply chain models comes from the fact that a proper model can be used in order to assist in the decision-making process.

In the past, the use of optimization models was determined by the availability of large and expensive computers. As a result their use and value was limited and many users were forced to settle for heuristic or simulation approaches in order to determine the optimum values of the decision variables. Currently, computer power is available at acceptably low cost and, combined with the availability of cost-effective data processing systems, it allows the extensive use of optimization techniques for determining the optimum values of the design variables related to the structure, control, and operation of a supply chain.

The mathematical models used in the optimization of supply chains involve binary variables together with continues variables and they can be stated in the following general form:

$$\min_{\mathbf{x},\mathbf{y}} J(\mathbf{x},\mathbf{y})$$

subject to f(x, y) = 0 and g(x, y)

and 
$$\mathbf{g}(\mathbf{x}, \mathbf{y}) \le \mathbf{0}, l_i \le x_i \le u_i, \forall i, y_j \in \{0,1\}, \forall j$$
 (1)

where  $\mathbf{x}$  = the vector of continuous variables,

y = the vector of binary optimization variables

J = a single or multi-objective quantitative performance criterion

 $\mathbf{f}$  and  $\mathbf{g}$  = the vector-valued functions that describe the behavior and the constraints of the supply chain

This general form of the mathematical programming problem is known as mixed-integer, non-linear programming problem (MINLP, [24]). When the objective function and the constraints are linear then Equation 1 is simplified to:

$$\min_{\mathbf{x},\mathbf{y}} \mathbf{c}^{\mathrm{T}} \mathbf{x} + \mathbf{d}^{\mathrm{T}} \mathbf{y}$$

subject to Ax + By = 0

and 
$$Dx + Ey \le 0$$
,  $l_i \le x_i \le u_i, \forall i, y_j \in \{0,1\}, \forall j$  (2)

This form is known as a mixed-integer, linear programming problem (MILP) and is considerably easier to solve compared to the MINLP problem. However, due to the combinatorial nature of the domain of y variables any attempt to enumerate completely all alternative solutions (there are  $2^n$  alternative combinations for n binary variables) is deemed to fail. Furthermore, MILP problems belong to the class of NP-complete problems. Accessible sources on modeling using integer variables and the solution of MIP problems (mixed integer, linear or non-linear programming problems) are the books by Floudas [24] and Williams [25], while Wolsey [26] offers a more advanced (yet readable) presentation.

Algorithmic approaches for solving MILP problems can be classified as *branch* and bound methods, *cutting plane* methods, *decomposition* methods, *logic-based* methods, or *branch* and cut, which is a combination of the branch and bound and cutting planes methods.

A very important idea in solving MIP problems is the idea of *relaxation* [26]. A problem (RP)  $\mathbf{z}^{RP} = \min\{\phi(\mathbf{z}): \mathbf{z} \in T \subseteq \mathbb{R}^n\}$  is a relaxation of (IP)  $\mathbf{z}^{IP} = \min\{\theta(\mathbf{z}): \mathbf{z} \in W \subseteq \mathbb{R}^n\}$  if:  $W \subseteq T$  and  $\phi(\mathbf{z}) \leq \theta(\mathbf{z})$ ,  $\forall \mathbf{z} \in W$ . If RP is a relaxation of IP, then  $\mathbf{z}^{RP} \leq \mathbf{z}^{IP}$ . The most well-known relaxation is the relaxation of the integrality constraints (when the constraints  $y_j \in \{0,1\}$  are relaxed to  $y_j \in [0,1]$ ) denoted as linear programming relaxation. In this case the resulting problem is a linear programming problem (LP), a problem much easier to solve compared to the MILP problem. By solving the LP problem we obtain a lower bound of the solution of the initial MILP problem. An upper bound of the solution of the MILP problem can be obtained by fixing arbitrarily the binary variables (*restriction*).

The branch and bound methodology for solving MIP problems is based on the idea of performing an "intelligent" enumeration of the alternatives without examining all combinations of the y variables. A key element in such an enumeration is the representation of alternatives via a binary tree. At each node of this tree a linear programming relaxation is solved. Then, based on branching criteria, children nodes are generated. A node is not branched (fathomed) if the linear programming relaxation is infeasible, if the solution of the linear programming relaxation is an integer solution, or if the solution of the linear programming relaxation is worse that the current best integer solution.

Details of many of the recent mathematical programming systems as well as modeling languages are given in the web page maintained by the Institution for Operations Research and the Management Sciences (see for example www.lionhrtpub.com/orms/surveys/LP/LP-survey.html for the most recent survey). AMPL (www.ampl.com), GAMS (www.gams.com) and XPRESS-MP (www.dash.co.uk) are among the most well-known languages, while CPLEX (www.ilog.com), OSL (www.research.ibm.com/osl) and XPRESS (www.dash.co.uk) are among the most successful general MIP solvers.

#### 8.2.5 The Role of Emerging Technologies in SCM

SCM used to be simple compared to what it is today. Manufacturers sold to whole-salers or directly to retailers. Salespeople called on their supply chain customers and wrote orders. Or, retailers called in their orders or sent them by mail. This low-tech supply chain started to die out in the 1980s and was almost extinct by the mid-1990s. Supply chains are changing dramatically as the world economy becomes networked and the Internet and other emerging technologies are playing an important role in the interactions between members of the supply chain.

Figure 4, taken from [27], shows the extended supply chain where net-works/technology/and the Internet are at the nexus of the vendor/retailer/customer triangle. The traditional domain of logistics and supply chain management has been

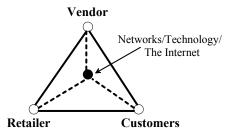


Figure 4. A networked, economy-induced supply chain [27].

the vendor-to-retailer link, which included topics like inventory control systems, category management, channel coordination, channel partnerships, and retailer networks. In the retailer-to-customer link, several new problems and research questions have arisen, which are managed with processes like data warehouses and loyalty programs, multiple selling channels, assortment planning, third party logistics intermediaries, and reverse logistics (which is the process of getting merchandise back through the channel). Finally, the growth of the Internet has expanded the research opportunities in the vendor-to-customer link and in supply chains in general, especially for cost reduction and service improvements [28]. Disintermediation, resulting from manufacturers selling directly to the ultimate consumer, has occurred with increased frequency and intensity as a result of the Internet. Also, manufacturers have an increased interest in and ability to strengthen their customer orientation [27].

#### 8.2.6 Environmentally Conscious SCM

The supply chain contains the extraction of raw materials, manufacturing, distribution, and use of goods. Waste generated in each component on the supply chain is collected at the end of the chain. The emissions and waste of the supply chain are transported and transformed and result in water, air, and soil pollution with damaging effects to the environment. The amount of waste and emissions can be reduced with certain actions and changes within the supply chain. That is, new decisions are necessary to decrease emissions and waste flows. Legal requirements and changing consumer preferences increasingly make suppliers and manufacturers responsible for their products, even beyond their sale and delivery [29].

Figure 5 presents potential environmental actions in a supply chain. The first actions, such as waste treatment, have been effect-directed. Somewhat more integrated are waste-directed and emission-directed adaptations in technology, such as reuse of materials and packaging and recovery of products. The most integrated approach is source-directed and deals with adaptation of raw materials, product redesign, and process changes [29].

It is a fact that there is a growing attention for environmental issues in SCM during the last years, in concordance with the general shift from corrective policies towards prevention. This development has led to a close interaction between SCM and environmental management. In the majority of supply chain cases, environmental

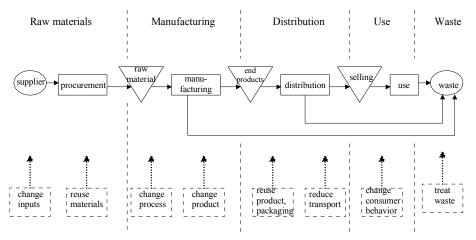


Figure 5. Environmentally conscious supply chain [29].

management is embedded into the general SCM scheme with appropriate actions that lead to environmentally conscious decisions and results. Moreover, these environmental actions are not taken at the end of the chain as correction measures, like some years before; they are now taken throughout the entire supply chain process, mostly as prevention measures.

# 8.2.7 SCM in Agribusiness

The first scholarly effort in the disciplines of marketing and logistics, which are parts of SCM, seems to be a report on the distribution of farm products [30] (cited in [27]). The design of supply chains of agricultural commodities like dairy products, fruits, and flowers can be complicated because in each link of the supply chain the quality of a product is influenced intentionally and unintentionally [31]. Agri-food supply chain managers must be concerned with control of food quality and safety and with the potential for weather-related supply variability. These concerns, unique to the food sector, may justify a different approach to supply chain management than the product-based approach suggested by general management theory [32]. Perishable products like food require a time-efficient supply chain, even if rapid delivery is costly.

The general members of a supply chain of a product are shown in Figure 6. Examples of members are factories, farmers, auctions, transporters, wholesalers, and retailers. In non-agri chains, the actions of each member of the chain modify the product characteristics in such a way that the product reaches the end user according to target specifications. During transportation and storage, basically nothing happens to the product states. In supply chains for agricultural products, however, a very important feature is product quality, which is continuously liable to changes. This continuous process is referred to as *quality development*, which can be slowed down or accelerated. Generally, changes are irreversible. Quality development of agricultural com-

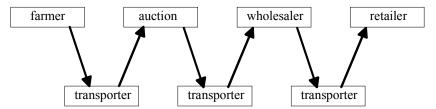


Figure 6. Example of members of an agri chain.

modities largely based on biological, physical, and chemical processes [33]. The following factors influence quality development [34]:

- process conditions, which are ambient conditions influencing product characteristics, such as temperature, relative humidity, light intensity, concentration of gasses and physical forces on the product;
- throughput time in a link during which the product is exposed to the process conditions; and
- appearance state of the product, such as packing and particle size.

The following three types of actions can be distinguished in agri chains [31]:

- *handling*, which are actions that intentionally alter or modify the appearance states of a product, e.g., wrapping, cutting, and labeling;
- *processing*, which are actions that intentionally alter or modify the quality states of a product, e.g., cooling and drying; and
- *transportation and storage*, which are actions that intentionally and unintentionally alter the quality states of a product.

The control of conditions during these processes can be costly. An additional difficulty is that in some cases, an action cannot be easily defined as one of the categories above. If there are models that describe the physical, chemical, and biological changes of the products due to specific process conditions and the associated costs, it is possible to optimize step-wise decision problems of agri chains, as it was done for example in [31], using dynamic programming.

In general, agribusiness research evolves along two parallel levels of analysis: the study of coordination between vertical and horizontal participants within the agri chain, known as agribusiness economics, and the study of decision-making within the alternative agri chain governance, known as agribusiness SCM [35].

As the agribusiness supply chain has become more complex, mainly by the increased liberalization of market policies and the globalization phenomena during the 1990s, its management requires advanced methodologies and strategies, even in developing countries. (For a collection of recent papers on agri-product SCM in developing countries, see [36].) If one also considers that in addition to economizing on transaction costs, other objectives like quality, screening, animal safety, traceability, and community development are emerging, then it is a consequence that new frameworks, such as networking models, system simulations, ecological footprinting, and reverse logistics, are necessary for the integrated management of agri chains.

## Emerging Technologies in Agribusiness SCM

A variety of emerging technologies have transformed the way SCM operates in agribusiness. Information technology (IT) can provide information about a wide range of product attributes and is an effective and important way to coordinate activities in the supply chain [37]. Examples of IT methodologies are the electronic data interchange (EDI), such as EDI-pigs and EDI-flowers in the Netherlands, the identification and recording (I & R) systems of produce, and the quality assurance systems to assure quality of produce and, by doing so, improve competitiveness.

It has been suggested [38] that characteristics of agricultural products should be considered in decisions about IT-based coordination of the supply chain. The nature of the product being sold is the main characteristic of interest here, and more specifically, understanding whether the product is "functional" or "innovative" [32]. *Functional* products are those that have predictable demand. *Innovative* products are differentiated, have many varieties, and usually exhibit short lifecycles.

According to Fisher [38], the primary objective of SCM for a functional product should be the reduction of costs of the physical functions along the chain. Typical examples of IT for physical functions include the automation of ordering processes and payment mechanisms, scheduling of warehousing and delivery and control systems for quality assurance in production [32]. The SCM of innovative products should focus less on costs and more on delivering the attributes that consumers desire. The consumer demand is the primary factor that has to be analyzed and the choice of suppliers should be based on speed and flexibility. The problem with innovative products is that the companies that introduce them cannot know *a priori* their consumption rates, thus the ideal SCM is responsive. Major tools for this responsive operation are JIT production and systems that link orders to sales, in order to limit stockouts or overstocks, while IT systems that are usually used are scanner data collection and customer loyalty cards, which enable food retailers to understand and predict consumer desires.

Several IT systems are applied in agribusiness SCM cases in order to offer solutions in several food-specific issues. Food safety issues, for example, are very important in agri chains. Fast detection and response to food safety problems require the ability to trace back small lots, from retail to processor or even to the farm [32]. The solution is given by IT systems involving bar-coded products at all stages of processing. Another food-specific issue that can be managed by the implementation of IT systems in SCM is the supply variation due to biological cycles and weather conditions. This variation causes food input costs to vary, because raw material costs are not easily controllable and perhaps not even predictable. This supply unpredictability makes necessary for SCM to better understand commodity markets, using a variety of information sources [39]. Finally, the seasonality of agricultural production, particularly for crops, can affect SCM approaches. However, no amount of IT investment could eliminate the need for extensive storage of products [32].

E-commerce is a quite recent IT methodology in agriculture. It gained a great expansion with the development of the Internet. Common agribusiness transactions such as buying, selling, trading, delivering, and contracting seem to be natural targets for conversion to e-commerce [40]. The main theoretical benefits of e-commerce in agri-

business include promotion of information flow, market transparency and price discovery, facilitation of industry coordination, and reduction or elimination of transaction costs [41]. Internet-based e-commerce is a very recent phenomenon. Several Internet-based e-commerce business models exist: auctions, exchanges, and catalogs. Sachs [42] discussed the general barriers cited by businesses to Internet-based e-commerce adoption. Those include unclear return on investments, lack of stakeholder support, and complicated technology.

In summary, competitive advantages to the entire supply chain can be brought by the information generated by the use of appropriate IT systems in agri-food SCM, to the extent that information is shared. IT is a valuable tool for managing agribusiness supply chains that are capable of rapid response but of course it has its limits. However, the rapid development of IT and the Internet promises the minimization of those limits in the near future.

# 8.2.8 Traceability and Tagging Systems in Agriculture

The rapid development of agriculture during the last years, together with the dynamics of the global food system, have resulted in high demand for capturing and sharing information within the agri-food supply chain. This is achieved with the development of appropriate traceability and tagging systems. *Traceability* is the ability to track a product batch and its history through the whole production chain (or part of it) from harvest through transport, storage, processing, distribution, and sales *(chain traceability)* or internally *(internal traceability)* in one of the steps in the chain, for example the production step [43]. Traceability is a general concept and its fundamentals are independent of the type of product, production, and control system it serves [44].

# Principles of Traceability Systems

Simple traceability systems existed from early times in the food industry. With the increasing implementation of Good Manufacturing Practice [45] and ISO 9000 quality management in food manufacture, traceability systems have become more advanced. They now cover more information and more steps in the production chain. The four fundamental features of traceability, according to [43], are:

- 1. *product*, which may relate materials, their origin, processing history and their distribution and location after delivery;
- 2. data, which relates calculations and data generated throughout the quality loop;
- 3. *calibration*, which relates measuring equipment to national or international standards, basic physical constants or properties, or reference materials; and
- 4. *IT and programming,* which relates design and implementation back to the requirements for a system.

According to [46], an integrated supply chain traceability system in agriculture and food business, must encompass the following features: product traceability, process traceability, inputs traceability, disease traceability, genetic traceability, and measurement traceability. The main concept of product traceability is the ability to identify products uniquely. This identification can be made by physical marking on the product or its package or by use of records [47]. The use of computers and IT enables a large

amount of data to be handled, therefore traceability systems with very detailed information about both the product and its processing history can be developed.

Traceability can be distinguished in chain traceability and internal traceability. Chain traceability can be applied by managing information in two main different ways: (1) information can be stored locally in each of the steps in the chain sending only product identification information along with the product, and (2) information can follow the product all the way through the chain, something that is necessary if it is desired to bring information from early steps in the chain to the consumer or to advertise special features of a product. In practice, most information is stored locally and little follows the product. The advantages of chain traceability are that it [43]:

- establishes the basis for efficient recall procedures to minimize losses;
- provides information about the raw material that can be used for better quality and process control;
- avoids unnecessary repetition of measurements in two or more successive steps.
- improves incentive for maintaining inherent quality or raw materials;
- makes possible the marketing of special raw material or product features; and
- meets current and future government requirements.

On the other hand, internal traceability is performed within a step in the chain. Some advantages that can be drawn with this type of traceability are [43]:

- the possibility for improved process control;
- cause-and-effect indications when product does not conform to standards;
- the possibility of correlating product data with raw material characteristics and processing data;
- better planning to optimize the use of raw material for each product type;
- the avoidance of uneconomic mixing of high- and low-quality raw materials;
- the ease of information retrieval in quality management audits; and
- better grounds for implementing IT solutions to control and management systems, e.g., computer-based quality management systems, etc.

The establishment of internal traceability can be easy enough for pure batch processing, but it can be very difficult for continuous or semi-continuous processing.

# Product Identification and Tagging Systems

Advances in geospatial science and technology [48] such as *remote sensing (RS)*, geographic information systems (GIS), and global positioning systems (GPS), can be used to collect site-specific data on individual animals, plants, soil properties, maturity, yield, and quality, as well as environmental and climatic data and to monitor animal movement and disease epidemiology. Grains and fresh products are generally handled in batches that may contain materials from different farms, particularly when the volume supplied by each individual farm is not sufficient to warrant a separate supply chain. Even so, it is possible to label each product or bag of grain so that it can be traced back to the origin [46].

Bar codes are the most common technologies for identifying raw food materials or finished products. In the livestock industry, ear tags are attached to the animal and the label on the tag may contain numbers or their combination with alphabets, which together contain information such as breed, date of birth, farm/paddock, movements, vaccinations, etc. When the animal is slaughtered, additional information may be added to the label, such as date slaughtered, time put in storage, and environmental conditions in storage. In addition to the information contained in the label, the farm record should also document the type and source of feed and other treatments and inputs used to raise the animals. Electronic identification tags are now available which can store large amount of data and also enable data on tags to be read automatically, without contact, using electronic card readers. These cards may be connected to a computer or downloaded later outside the measurement site.

A product can be identified by reading a tag attached to it, by recognizing it using image-processing techniques, or by looking it up from its position. These methods are described below.

# Tags

If a product can be sufficiently close to a reader, then reading an identification tag attached to is often an efficient method. There are several tag technologies [49]:

- Infrared tags—The tag may be a low-cost, low-power infrared "active badge" worn by a person [50], or a "beacon" attached to a thing. These devices emit the identifier over IrDA, for reception by infrared-equipped devices including PDAs.
- Optically sensed tags—These include standard bar codes, already found on many everyday items, and symbols specially designed for easy capture with a digital camera, including "cybercodes" [51] and "glyphs" [52]. Bar codes can be printed and inexpensive readers are becoming available.
- *RFID tags*—Radio frequency identification (RFID) tags can be read at a distance and, since they operate by induction, require no power source of their own [53].
- Contact tags—iButtons (Dallas Semiconductor, http://www.ibutton.com) are read by electrical contact with their casing. Like RFID tags, they do not require their own power.

These tagging technologies have relative advantages and disadvantages in terms of cost and suitability for different physical environments. For example, sometimes reading at a distance is desirable (IR, RFID, glyph recognition); in other situations it is preferable that the user should bring the reader up close to make a definite and unambiguous identification, such as by scanning a bar code.

# Computer Vision

Stereo computer vision—object recognition—may be used instead of tagging [54]. This has the advantage of eliminating the logistics of tagging but it has the disadvantage of requiring relatively powerful computing resources. Considerable work is needed before this method could be used routinely.

#### Positioning

In cases where objects move rarely or not at all, or are automatically tracked, a third means of identification is to use a positioning sensor to determine the product's coordinates, and so look them up in a database. By adding an electronic compass, we can also identify remote objects by pointing at them. The GPS is a widely available

positioning system, but only outdoors. Short-range radio frequency (RF) triangulation may be deployed indoors or out. Ultrasound techniques can be incorporated for more fine-grained positioning, down to a few centimeters [55].

As the demand for high quality in agricultural products increases, traceability has become an important factor in global agribusiness trade, primarily because of the rising incidence of new regulatory food standards and the concern for environmental sustainability. Traceability enables full backward and forward identification of the product with respect to time and location in the supply chain and thereby facilitates cost-effective withdrawal of products when defects occur and when safety standards are violated [46]. Traceable chains are technology-intensive and information-driven. Current advances in computers and electronics (mechatronics) in agriculture and progress in geospatial science and technology tools for precision agriculture will facilitate successful integration of traceability into existing agricultural mechanization systems.

# References

- 1. Beamon, B. M. 1998. Supply chain design and analysis: Models and methods. *International J. Production Economics* 55: 281-294.
- 2. Skinner, W. 1985. *Manufacturing: The Formidable Competitive Weapon*. New York, NY: Wiley.
- 3. Erenguc, S. S., N. C. Simpson, and A. J. Vakharia. 1999. Integrated production/planning in supply chains: An invited review. *European J. Operational Research* 115: 219-236.
- 4. Stevens, G. C. 1989. Integrating the supply chain. *International J. Physical Distribution and Materials Management* 19: 3-8.
- 5. Thomas, D. J., and P. M. Griffin. 1996. Coordinated supply chain management. *European J. Operational Research* 94: 1-15.
- 6. Oliver, R. K., and M. D. Webber. 1982. Supply-chain management: Logistics catches up with strategy. *Outlook; 1992. Logistics, The Strategic Issue*, ed. M. G. Christopher. London, UK: Chapman and Hall.
- 7. Croom, S., P. Romano, and M. Giannakis. 2000. Supply chain management: An analytical framework for critical literature review. *European J. Purchasing and Supply Management* 6: 67-83.
- 8. Lambert, D. M., and M. C. Cooper. 2000. Issues in supply chain management. *Industrial Marketing Management* 29: 65-83.
- 9. Schary, P., and T. Skjtt-Larsen. 1995. *Managing the Global Supply Chain*. Copenhagen, Denmark: Handelshojskolen Forlag.
- 10. Alvarado, U. Y., and H. Kotzab. 2001. Supply chain management; The integration of logistics in marketing. *Industrial Marketing Management* 30: 183-198.
- 11. Lambert, D. M., M. A. Emmelhainz, and J. T. Gardner. 1996. Developing and implementing supply chain partnership. *The International J. Logistics Management* 7(2): 1-17.

- 12. Cooper, M. C., and J. T. Gardner. 1993. Good business relationships: More than just partnerships or strategic alliances. *International J. Physical Distribution and Logistics Management* 23(6): 14-20.
- 13. Salmon, K. 1993. Efficient Consumer Response: Enhancing Consumer Value in the Grocery Industry. Washington, DC: Food Marketing Institute.
- 14. Efficient Consumer Response Europe. 1996. European value chain analysis. Final Report.
- 15. Gattorna, J. L., and D. W. Walters. 1996. *Managing the Supply Chain: A Strategic Perspective*. London, UK: MacMillan Press Ltd.
- 16. Shapiro, J. F. 2001. *Modeling the Supply Chain*. Duxbury, CA: Thomson Learning.
- 17. Min, H., and G. Zhou. 2002. Supply chain modeling: Past, present and future. *Computers and Industrial Engineering* 43: 231-249.
- 18. Melachrinoudis, E., and H. Min. 2000. The dynamic relocation and phase-out of a hybrid, two-echelon plant/warehouse facility: A multiple objective approach. *European J. Operational Research* 123: 1-15.
- 19. Lee, H. L., V. Padmanabhan, and S. Whang. 1997. Information distortion in a supply chain: The bullwhip effect. *Management Science* 43: 546-558.
- 20. Karabakal, N., A. Gunal, and W. Ritchie. 2000. Supply-chain analysis at Volkswagen of America. *Interfaces* 30(4): 46-55.
- 21. Al-Mashari, M., and M. Zairi. 2000. Supply chain re-engineering using enterprise resource planning (ERP) software of a SAP R/3 implementation case. *International J. Physical Distribution and Logistics Management* 30: 296-313.
- 22. Lee, Y. H., M. K. Cho, S. J. Kim, and Y. B. Kim. 2002. Supply chain simulation with discrete continuous combined modelling. *Computers and Industrial Engineering* 43: 375-392.
- 23. van der Vorst, J. G. A. J., A. J. M. Beulens, and P. van Beek. 2000. Modeling and simulating multi-echelon food systems. *European J. Operational Research* 122: 354-366.
- 24. Floudas, C. 1995. *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. New York, NY: Oxford University Press.
- 25. Williams, H. P. 1993. *Model Building in Mathematical Programming*. 3<sup>rd</sup> ed. Chichester, UK: John Wiley and Sons.
- 26. Wolsey, L. A. 1998. *Integer Programming*. New York, NY: John Wiley and Sons.
- 27. Levy, M., and D. Grewal. 2000. Guest editor's overview of the issue: Supply chain management in a networked economy. *J. Retailing* 76: 415-429.
- 28. Lancioni, R. A., M. F. Smith, and T. A. Oliva. 2000. The role of the Internet in supply chain management. *Industrial Marketing Management* 29: 45-56.
- 29. Bloemhof-Ruwaard, J. M., P. van Beek, L. Hordijk, and L. N. van Wassenhove. 1995. Interactions between operational research and environmental management. *European J. Operational Research* 85: 229-243.
- 30. Crowell, R. 1901. Report of the Industrial Commission on the Distribution of Farm Products. Washington DC: Government Printing Office.

- 31. Gigler, J. K., E. M. T. Hendrix, R. A. Heesen, V. G. W. van den Hazelkamp, and G. Meerdink. 2002. On optimisation of agri chains by dynamic programming. *European J. Operational Research* 139: 613-625.
- 32. Salin, V. 1998. Information technology in agri-food supply chains. *International Food and Agribusiness Management Review* 1: 329-334.
- 33. Zwietering, M. 1993. Modeling of the microbial quality of food. PhD diss. Wageningen, the Netherlands: Wageningen Agricultural University.
- 34. Zwietering, M., and K. van't Riet. 1994. Management of Agrichains. *Managements Studies and the Agribusiness*, ed. G. Hagelaar, 108-117.
- 35. Cook, M. L., and F. R. Chaddad. 2000. Agroindustrialization of the global agrifood economy: Bridging development economics and agribusiness research. *Agricultural Economics* 23: 207-218.
- 36. Johnson, G. I., and P. J. Hofman, eds. 2004. Agriproduct supply-chain management in developing countries. *Proc. of a Workshop by the Australian Centre for International Agricultural Research*. Canberra, Australia: Australian Centre for International Agricultural Research.
- 37. Hannus, T., O. Poignee, and G. Schiefer. 2003. The implementation of a webbased supply chain information system—Experiences with a regional quality grain program. *Presented at the EFITA 2003 Conference*.
- 38. Fisher, M. L. 1997. What is the right supply chain for your product. *Harvard Business Review* Mar-Apr: 105-116.
- 39. Champion, S. C., and A. P. Fearne. 2001. Supply chain management: A "First Principles" consideration of its application to wool marketing. *Wool Tech. Sheep Breed* 49: 222-236.
- 40. Shapiro, C., and H. Varian. 1999. *Information Rules*. Boston, MA: Harvard Business School Press.
- 41. Leroux, N., M. S. Wortman Jr., and E. D. Mathias. 2001. Dominant factors impacting the development of business-to-business (B2B) e-commerce in agriculture. *Intl. Food and Agribusiness Management Review* 4: 205-218.
- 42. Sachs, G. 2000. B2B: 2B or not 2B, the second survey. Accessed from http://www.gs.com/ high-tech/research/b2b-second-survey.pdf.
- 43. Moe, T. 1998. Perspectives on traceability in food manufacture. *Trends in Food Science and Technology* 9: 211-214.
- 44. Kim, H. M., M. S. Fox, and M. Gruninger. 1995. Ontology of quality for enterprise modeling. *Proc. of WET-ICE, IEEE*, 105-116.
- 45. Anderson, K. G. 1990. Regulations and factory enforcement arising from the UK Food Safety Bill. *Food Control* 1: 199-203.
- Opara, L. U. 2002. Engineering and technical outlook on traceability of agricultural production and products. *Agricultural Engineering International:* The CIGR J. Scientific Research and Development. Vol. IV, Invited Article No. 18.
- 47. Early, R. 1995. *Guide to Quality Management Systems for the Food Industry*. Blackie Academic and Professional.

- 48. Bossler, J. D. 2001. *Manual of Geospatial Science and Technology*. Philadelphia, PA: Taylor and Francis Group.
- 49. Want, R., and D. M. Russell. Ubiquitous electronic tagging. *Distributed Systems Online, IEEE*. http://www.computer.org/dsonline/articles/ds2wan.htm.
- 50. Want, R., A. Hopper, V. Falcao, and J. Gibbons. 1992. The active badge location system. *ACM Transaction on Information Systems* 10: 91-102.
- 51. Rekimoto, J., and Y. Ayatsuka. 2000. CyberCode: Designing augmented reality environments with visual tags. *Proc. of Designing Augmented Reality Environments*. ACM.
- 52. Hecht, D. L. 1994. Embedded data glyph technology for hardcopy digital documents. *Proc. Society of Photo-Optical Instrumentation Engineers Symposium on Electronic Imaging, Science and Technology* 2171: 341-352.
- 53. Sigrimis, N. A., N. R. Scott, and C. Czarniecki, 1985. A passive electronic identification system for livestock. *Trans. ASAE* 28: 622-629.
- 54. Brumitt, B., B. Meyers, J. Krumm, A. Kern, and S. Shafer. 2000. EasyLiving: Technologies for intelligent environments. *Proc. of the 2<sup>nd</sup> Intl. Symposium on Handheld and Ubiquitous Computing*. 12-29.
- 55. Harter, A., A. Hopper, P. Steggles, A. Ward, and P. Webster. 1999. The Anatomy of a context-aware application. *Proc.* 5<sup>th</sup> Annual ACM/IEEE Intl. Conference on Mobile Computing and Networking, 59-68.

# **9** Low Cost IT for Developing Countries

H. K. Purwadaria, I W. Budiastra, Suroso, I W. Astika, and D. R. Heldman

Abstract. In developing countries, a low cost information system must be capable of reaching farmers in the villages. The cost must always be balanced against the benefits. Radio and TV are two ways to spread information. Examples of IT systems are given for many developing countries. In Indonesia, CD-ROMs and video cassettes are used to disseminate appropriate technology for small and medium enterprises. Topics covered include agricultural machinery, cropping systems, aquaculture, animal husbandry, water management and sanitation, and food processing technologies. Another example is computer software for optimum scheduling of tractor and implement operation.

Examples of IT development in agriculture are a control system for nutrient solutions, product quality evaluation using low cost image analysis, near infrared and ultrasonic systems for evaluating the internal quality of agricultural products, neural networks for forecasting rainfall runoff, controlling production and post-harvest handling, fuzzy control of tea processing, genetic algorithms for controlling fermentation processes, and watershed optimization.

All these examples are designed to provide the fullest benefit for the most people and thus to accelerate the economic growth and human welfare in developing countries.

**Keywords.** Low cost IT, Developing countries, Information system, Industrialized agriculture, E-business.

#### 9 1 Introduction

H. K. Purwadaria

Information and communication technology has penetrated the community life of people in the developing countries and has become a key factor in their economic growth, as well as in the developed countries. Even though the developing countries are struggling to improve their economic welfare, some (such as India and China) have made a significant thrust in their IT industries. Low cost IT is a priority since it can reach most people and provide benefits for them. Hence, the current use of low

cost IT will be discussed first, then research to support the industrialization of agriculture will be illustrated, followed by possible uses of advanced IT in the future.

It should be realized that the criteria of low cost IT are viewed on relative basis since the cost should always be balanced against the benefit, and price does change from region to region and time to time, especially with the speed of new-found technologies. While many computers and chips have been manufactured in India, China, and some other developing countries, it is also worthwhile to note that various IT hardware manufactured in the developed countries have found markets and have been adopted in the developing countries. Low cost IT from the developed countries that may be transferred to support the developing countries are also briefly discussed here.

# 9.2 Current Uses of Information Systems for Agriculture

I W. Astika and H. K. Purwadaria

In the developing countries, a low cost information system must be capable of covering a wide area reaching farmers in the villages. Radio is one of the alternatives to spread information to farmers, along with TV and telephone. Both of the latter are being used more and more recently. Radio communication is cheaper since the cost of production and broadcasting are inexpensive, and the coverage is wide. Radio can reach an isolated community and be understood by illiterate people. However, the effectiveness of radio as a medium to spread information is not fully reliable. Most radio listeners seek entertainment programs, as it is true with TV. Only a few look for serious information.

Computer technology has the potential to provide the community information, including agricultural information, in a system that can be accessed at any time and which can be multiplied at a relatively low cost. By the multimedia facility built into computers, information can be developed and presented in text, pictures, and voices, so it will be similar to books, radio, or TV. It is true that most farmers in the developing countries do not own a personal computer (PC), however the problem can be solved by building telecommunication shops where farmers can make use of PC and Internet services at relatively low cost.

The *multipurpose telecenter* (MPTC) or Internet shop developed in Bangladesh [1] has become popular in many developing countries and offers integrated information and communication technology (ICT) services for rural communities with accessibility to the net. These centers are called by various names throughout the world: *Information Kiosks* in India, *Thai RuralNet* in Thailand, *Telecenters* in Brazil, *Warnet* in Indonesia, and *Public Information Centers* in Albania. In Indonesia, for example, the *Warnet* has grown commercially and reached the cities in the subdistricts with a current cost of about Rp 3000/hr (US\$ 0.33/hr). In this case, the governments of developing countries could be the Internet providers and make the telephone networks available and accessible to the remote areas.

The government of India, along with private enterprises and local and international NGOs, have sponsored hundreds of ICT projects in order to spread information closer

to people in the country. Many projects establish telecenters, while others develop e-government, applications software, video and TV programs, and implement related training. One of the telecenter projects is aimed at empowering farmers with information on agriculture via the Internet and videoconferencing. Each communication center covers surrounding villages within a radius of 25 km, and is able to train farmers using interactive technical CDs, visual presentations, and VCDs. Farmers may also access global and national market information, meteorological data, disaster management techniques, and pest and disease control information. Educated unemployed rural youth are encouraged to set up their own kiosks in their own villages in order to earn a living. The kiosk owners will be allowed to charge farmers for printing land and revenue records, and pest control and diagnostic services. The owner of the kiosk is bound to provide free of charge services such as display of market rates, offline training on agriculture by using materials like CDs and VCDs, and providing information on government schemes [2].

Another project in India is to make agri-information available to farmers across several states in their regional languages. The information is displayed in the web portal called www.jfarmindia.com. There are 12 crop categories covering the total 92 crops. The farmers can explore information ranging from crop production, farm machinery, crop protection, and processing technologies [3].

A project run by a private enterprise aims at answering specific needs of the rural community is set in the web site www.indiagriline.com. In the future, the web site is expected to provide other services online: trading in agricommodities and other industrial products, contract management platforms, online banking, online retailing, latest news, special transaction, telecom infrastructure, and online consulting [4].

The Thai RuralNet has been developed with assistance from leading local and international organizations. It has successfully developed and tested a graphic-based web portal specifically designed to meet the needs of rural citizens. This enables people to find relevant rural-oriented content. A target group called the Ratburi group has already been able to gain a real (0.07 to 0.12 US\$) increase in their ginger price resulting from this initiative. Training in basic computer skills has been undertaken for youth volunteers in rural communities in three district provinces [5].

The Agricultural Information Network (AIN) is an Internet portal set up at hundreds of offices of the Thai Bank for Agriculture and Agricultural Cooperatives (BAAC), enabling farmers, field officers, policy-makers, and government officials to communicate and access relevant and useful agricultural information. AIN was conceived by BAAC and delivers information and services to over 5 million farm households throughout Thailand. This information includes prices of agricultural goods in various locations, information related to agricultural technologies, and best practices adopted by other successful farmers. To better plan their crops, farmers may want (for example) to know the locations of the best soil types for growing maize, or what areas in their district have been affected by drought in the last five years. In addition to farmers, the potential users of the network include BAAC staff, members, and clients, agricultural extension agents, government, and other financial institutions [6].

# 9.2.1 Agro-Information Distribution Using Radio Broadcasts

In many developing countries, where a sophisticated information network is not available in villages, radio is still the only media for information distribution on both general and agricultural subjects. Government relay stations have assisted in developing agricultural extension programs for the farmers, and also family planning programs for the community. A case study in Honduras revealed that radio programs, training, and visits were successful in motivating progressive farmers in adopting new technology [7]. Training programs and personal visits were considered to be more effective than radio, but they are more costly and attained less coverage area in the same time period.

In Indonesia, an agricultural extension program by radio broadcast is commonly listened to by the farmer groups under the supervision of a local agricultural extension officer. The radio station may provide a resource person during the program who can respond to questions put by the farmers, anywhere telephone is accessible. Thus, the radio program can be made interactive to increase its effectiveness and benefit.

The market information of the agricultural commodities is monitored by the Department of Industry and Trade. The information generally covers the daily market price of food products such as rice, maize, soybean, cassava, fruits and vegetables, and

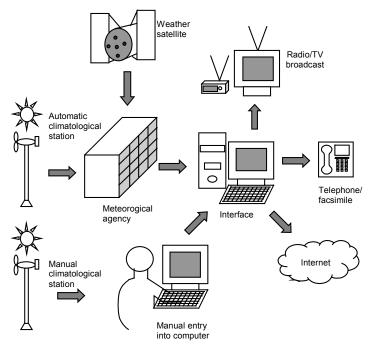


Figure 1. Semi-automatic weather information system.

the plantation crops such as tea, coffee, cacao, rubber sheet, and crude palm oil. The price is specific per location, and the minimum and maximum prices per commodity are also announced.

Weather information, originally recorded by the meteorology station, is also transferred to the radio station for broadcasting. The weather information system can be improved by a semi-automatic system as illustrated by Figure 1. The daily weather data can be recorded digitally by the climatological stations, or alternatively, data from the weather instruments is inputted to the computer manually. Data from the stations is then pooled by the National Agency of Meteorology (NAM) through the computer network and integrated with the data from satellites. NAM makes the weather prediction for the days to come, and releases the actual and predicted weather to the radio station to be broadcast. The weather data is also made available on the web site and through a telephone line.

The government of India runs a wireless-telephone project (using radio waves) to make the telephone accessible to the people at remote areas in the country. The handsets, operated on the wireless in local loop (WLL), are carried by postmen while they are out for delivering letters and money orders. The handsets are operable in an area 5 km from the nearest tower. People have to pay for the telephone services [8].

# 9.2.2 Computerized Agricultural Information Systems

#### CD-ROM and Video Cassette

In Indonesia, the Ministry of Research and Technology has produced a CD-ROM containing the appropriate technology for small and medium enterprises (SMEs). The CD-ROM is one of the information packages distributed via the technology information shops (*warintek* is the Indonesian term) program. The *warinteks* are located in regional libraries, universities, and the offices of the subdistrict government, where they can be accessed by the community, students, lecturers, and professionals.

The CD-ROM covers agricultural machinery, cropping systems for various commodities, aquaculture for salt water and fresh water, animal husbandry, water management and sanitation, and food processing technologies. The CD is easily operated under Microsoft Windows<sup>®</sup> using HTML or PDF, which allows the users to print the necessary information. The same information is currently available on the web.

Similar to using CD-ROMs, the National Science and Technology Board of India uses videocassettes to store training materials. The board distributes a series of video programs for entrepreneurship training. The videos are visually exciting and easy to understand. The materials are based on realistic case studies and have been field tested on sample groups. The presentations are a mixture of graphics and audiovisual. Each topic is of 25-30 minutes in the English language [9].

# An Interactive Information System Accessed through Telephone Lines

The voice response system in currently produced PCs gives the users the ability to access information through the telephone line. The agro-information system has been implemented in the system in a way similar to the information provided by the telephone company for users requesting their monthly bill information. The users dial a

given number, and the reply comes from the computer giving necessary instructions to the users to dial more numbers until certain specific information is located. Some examples in Indonesia are the mango postharvest handling system [10] and the rice postharvest handling system [11]. The mango postharvest handling system covers the technology of harvesting, sorting, packaging, storage, postharvest diseases, transportation, and marketing for various mango species. The rice postharvest handling system describes the optimum harvest time, agricultural machinery for harvesting, threshing, drying, storage, and milling.

With UNDP funding, the Philippines Rice Research Institute (PhilRice) conducted research on the development of an SMS-based information system on the rice seed-stock inventory for farmers. With mobile phones, the seed growers and seed centers were linked together and shared a real-time seed-stock inventory system. The system had an SMS server to handle incoming data and queries, and a database to process data, directory, and report generation [12].

# A Web Site for Marketing Agricultural Products

Internet technology has been phenomenal for the growth of web sites for agricultural commodities in the developing countries. The web sites are commonly developed either by private companies or by the government. Mostly, the web sites aim to facilitate transactions for commodities such as being used in e-business, but some information could be added such as pre- and postharvest technologies, news, and discussion forums such as in e-marketplaces. Users are not charged when accessing the web sites, since the web sites gain income from advertising.

Even though buying products through the Internet has its own constraints such as the quality assurance of the products, a few small and medium entrepreneurs (SMEs) in Indonesia have set up e-businesses. However, like many other developing countries, an integrated business and technology information web site still needs to be developed.

The Department of Agriculture in Indonesia has launched a web site to promote agricultural commodities: www.fintrac.com/indoag/promotion. The main menu contains price information, pre- and postharvest technologies, marketing information, production statistics, on-line transactions, how and where to make a good investment, product standardization, and list of agribusiness importers and exporters.

The State Government of India has also initiated a project called the Dairy Information System Kiosk (DISK). The project aims at providing data analysis and decision support to help rural milk collection societies in improving their milk productivity. The project consists of two components: an application running at the rural milk collection society, and a portal at the district level serving transactional and information needs of all members. The portal has illustrative content in Gujarati and English [13].

In China, an estimated more than 2000 agriculturally oriented web sites are spreading all over the country. However, only a few web sites are expected to survive for long, since most of the web sites lack a viable marketing mechanism. Based on the experience to date, a UNDP-funded project has been started, targeting Qingdao, Shandong Province, China. The project, called *all-round ICT services*, is presented at various levels of government. Regional, county, district, and village are combined together

to help the farmers in marketing activities. At each level, different information and application modules are developed. In addition to the Internet as the backbone, the county-run television station, calling center, telephone, mobile phone, and village-run broadcasts will be used to meet the needs of farmers [14].

### 9.2.3 Decision Support Systems (DSS) for Agriculture

### Optimum Scheduling of Tractors and Implements for Sugarcane Plantations

Computer software to optimize the scheduling of tractor and implement operation on sugarcane plantations [15] has been developed. The optimization program covers various alternatives for combining tractors, implements, operators, land area, and time for specific jobs in the plantation. The optimum schedule is characterized by minimizing the tractor down time and the completion of the whole job in a timely fashion.

The software is developed interactively by Visual Basic  $4.0^{\circ}$ . The program also takes into account the previous weather conditions, the number of requested jobs, the availability of operators, tractors, and implements, and the job sequences. The optimization is done using random search methods.

### An Internet-Based Expert System for Tractor Troubleshooting

An Internet-based expert system has been designed to diagnose problems with hand tractors [16]. The trouble identification is based on a tree diagram provided by the hand tractor manufacturer. The items described by the tree diagram are converted into a chart containing various fields: (1) identification for each corresponding branch; (2) codes to make inquiries, to reply, or to draw a conclusion; (3) the title of the branch; and (4) a users' manual. To track and to identify the trouble, the database is called by a routine program. The expert system has been installed at a commercial web host that supports active server pages (ASP).

#### A Commodity-Specific Database System for Agribusiness Development

A DSS (decision support system) [17] has been developed having two main functions: to provide data required for the agribusiness development, and to support the decision in determining a prioritized commodity for a specific location. The database structure is presented in Figure 2, and the procedure to determine the prioritized commodity is illustrated in Figure 3.

The data can be tracked interactively by choosing the proper menus. The user may put in a specified commodity to obtain its botanical description, plant characteristics, and the postharvest handling system. Fifty-five commodities have been programmed into the database. The system includes agroecology data from a district in Java: land use, type of soil, soil structure, topography, and weather conditions, which are organized in a spatial database in raster form.

To determine suitable commodities for a specific location, the system is completed by a method called the *multi-criteria decision analysis*. When a location is entered by the user, the system will respond by suggesting 10 potential commodities suited to the given location. The criteria used in the analysis are the pricing index, the production trend, the profit index, the availability of labor, and the probability for industrial usage. The system then provides the final prioritized commodities as a recommendation.

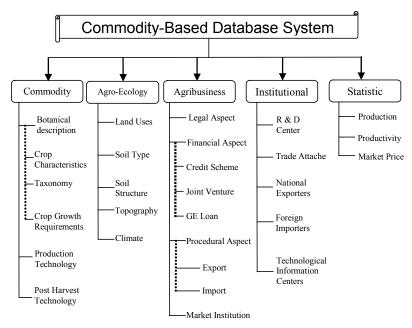


Figure 2. Database structure to support the development of an agribusiness [17].

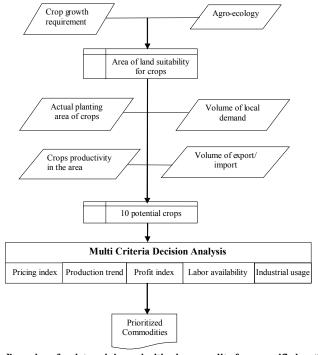


Figure 3. Procedure for determining prioritized commodity for a specific location [17].

# 9.3. IT Development for Industrialized Agriculture

I W. Budiastra, Suroso, and D. R. Heldman

Basically, a low cost IT hardware system is likely to result in lower performance. However, with rapid development of manufacturing technology and high demand for technology, many products including information technology are less expensive than before. Many high performance IT components such as personal computers or some transducers and interfaces are inexpensive now.

# 9.3.1 IT Applications for Agricultural Production

# A pH Control System for Nutrient Solutions

A fuzzy-based control system for maintaining a pH level of nutrient solution in ebb-and-flow hydroponic cultures (Figure 4) was developed and used in a greenhouse [18]. The culture vessel consisted of six blocks, each of which contained four potted flowers. The nutrient solution flowed into and filled the cultivation bench to 5 to 10 cm from the pot base for 10 minutes, then flowed back into the tank and flowed into the next block. The flow rate of the nutrition used in this experiment was 2.4 l/min and the measuring apparatus was a Hanna pH meter (HI8710E). The solidity control was by flowing 0.3 M H<sub>3</sub>PO<sub>4</sub> and 0.4 M KOH from Marriot tube with constant debit. Calibration of the pH meter was done on voltage basis using a PCL-812PG interface. Integrating with a fuzzy logic controller, the system could maintain pH at the setpoint level.

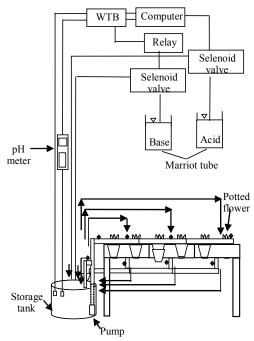


Figure 4. Schematic diagram of ebb and flow system with pH control system for nutrient solution [18].

### Applications of Image Analysis

Applications of image processing and computer vision in agriculture include remote sensing, predicting environmental conditions for bioproduction, precision farming, and quality evaluation of agricultural and food products.

Image analysis has been used extensively in the area of remote sensing to identify and determine the area of forest, estate, crops, and fishing zones. For example, it has been used to determine irrigated areas in the Zayandeh River Basin, Isfahan, Iran [19], and land use and land cover classification in the Amazonian area, Brazil [20]. Image processing, based on satellite data analysis, was used to forecast fishing zones in India [21]. Image analysis has been applied in precision farming, as in a guidance system for agricultural vehicles [22]. Applications of image processing include distance diagnosis in pest management in Greece [23]. Image processing for pest management (IPPM) is the automation of the diagnostic process based on interactions of growers and experts via the Internet, to enable pest diagnosis in crops at the earliest possible stage, using photographs produced by growers coupled with text message exchanges. An image processing system and neural network was used to successfully distinguished weeds from sunflower plants for controlling the application of herbicide in Turkey [24].

A similar system was used to estimate successfully measure crop cover percentage of wheat [25], leaf area of grapevines (26), and Internet growth in greenhouse [27].

# 9.3.2 IT Applications for Quality Evaluation of Agricultural Products

#### Image Processing Systems

A typical low cost image processing system for evaluating quality of agricultural products is illustrated in Figure 5. The hardware system consists of an illuminating source (tungsten lamp, Philips), a CCD camera (JVC TK 870 E) to record the light reflectance from the product surface, and an image processing board (Video Blaster Co.) to transform the analog signal into digital for further image processing in a PC 486. The CCD sensor had the capability to record images to  $512 \times 512$  pixels with 256 lighting intensity levels, while the image processing board had the capability to process images to  $640 \times 480$  pixels. The system had been used successfully to determine maturity and size of mangoes [28]. The average sensing time required for image processing was 5 seconds/product.

Image processing has been applied to evaluate the grades of the orchids *Dendrobium sonia* and *Dendrobium* 'Miss Singapore' based on the stalk length, number of total inflorescences, and number of blooming inflorescences [29]. It can also detect the abnormalities of the orchid petals caused by physical defects and discoloration spots of pesticide residue.

The applications of image analysis for evaluation of agricultural and food products have created opportunities for on-line detection systems and research instruments to provide new and improved understanding of the sensory attributes of foods. A thermal camera and image processing algorithm was developed to estimate the number and diameter of apple fruits in an orchard during the growing season in Slovenia [30]. In detecting weeds in a green lawn field, an image recognition system has been developed

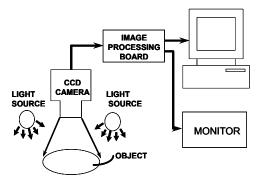


Figure 5. A typical low cost image processing system [19].

using gray-scale uniformity, which is not affected by color and thus overcomes the problem of the similar green color of the weeds and the lawn field [31]. The system also allows the analysis of the morphological image to find the exact location of the center of gravity to pinpoint the weeds. A high spatial resolution (0.5 to 1.0 mm) hyperspectral imaging system was used as a tool for selecting better multispectral methods to detect apple surface defects and contamination [32]. Machine vision systems for online sorting of potatoes, and apples have been demonstrated by Tao et al. [33,34]. Arham et al. [35] and Damiri et al. [36] have evaluated the maturity and ripeness of two kind of citrus (Citrus aurantifolia Swingle and Citrus medica) by image processing and developed the algorithm using an artificial neural network. An analysis of texture features, based on co-occurrence matrices (COMs), was conducted to determine the performance of dual-wavelength imaging for discriminating unwholesome poultry carcasses from wholesome carcasses [37]. The non-destructive, rapid, and costeffective methods for characterization of meat tenderness have been described by Tan [38]. Additional applications of image analysis for evaluation of sensory attributes of geometric, mechanical, and sensory properties of extruded foods were described by Gao et al. [39,40].

#### Near Infrared Systems

Figure 6 illustrates a typical NIR system (700 to 2500 nm) to measure the chemical composition of agricultural products. The hardware system was assembled from a halogen 1amp, light chopper, grating monochromator, integrating sphere with 60-mm diameter, PbS sensor, filter, lock-in amplifier, and AD converter (PCL 812, PC Lab Card). The optical system was manufactured by Shimadzu Co., Japan. Wavelength scanning was done by a stepping motor controlled by the computerized pulse motor [41-43]. The system successfully measured the fructose and malic acid content of apples [41], and sucrose and malic acid of mangoes [42,43]. The average sensing time required for NIR was 10 seconds/product.

A similar NIR system was used to detect *Rhizopus* spores on red tomatoes. Spore-free and infected tomatoes were classified using discriminant analysis with an accuracy of 81% and 75%, respectively. About 96% of the infected tomatoes were properly detected by the neural network method [44].

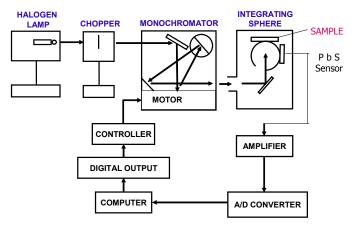


Figure 6. NIR system for evaluating the quality of agricultural products [41-43].

#### Ultrasonic Systems

A low cost ultrasonic system for evaluating the internal quality of agricultural products is illustrated in Figure 7. The apparatus was designed to measure the transmission of ultrasonic waves in durian pulp and for whole durian fruit. The components of the assembly were an ultrasonic tester (composed of a timing current, pulse generator, pulse amplifier, and voltage amplifier); a transmitter transducer T (Panametrics A 309S); a receiver transducer R (Panametric V302); a digital oscilloscope (HP 54201 A, Hewlett Packard); a general purpose interface bus (GPIB) using PCL 848A (PCLab Card); and a personal computer. The function of the ultrasonic tester was to send the ultrasonic wave to the transmitter transducer and to receive it back from the receiver transducer. This system could detect the firmness and defects in durian nondestructively based on ultrasonic wave transmission characteristics at 50 kHz [45,46].

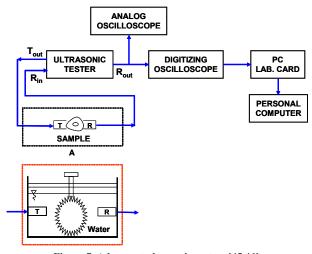


Figure 7. A low cost ultrasonic system [45,46].

Ultrasonic sensing combined with a radiometer was applied to estimate tiller density and leaf area index of winter wheat [47]. By using the combination of the coefficient of variation of normalized difference vegetation index (NDVI) and a new form of compound vegetation index derived from radiometer and ultrasonic sensor height output measurements, the tiller numbers and leaf area index (LAI) were estimated accurately.

#### Electrical Property Measurement Systems

An electrical property measurement system called a *Q meter*, basically a variable frequency oscillator used to excite an LRC circuit, has been developed. It consists of three basic functional blocks: source, resonating component (constant inductivity and variable capacitance in parallel to the product probe), and a voltmeter. With the source of a constant voltage and an appropriate frequency supplied by the oscillator, the capacitor is tuned for resonance of the LRC circuit. The Q meter can measure conductivity, power factor, and dielectric constant of agricultural products [48]. The system was applied to detect the moisture content of grain, seed, and pepper.

# 9.3.3 Artificial Neural Networks (ANNs) for Controlled Production and Post-Harvest Handling

### ANNs for Controlled Production

A study on the application of ANNs to control the growth of chili was conducted in 1999 [27]. The ANN was used to determine the water and nutrient status in the chili plant, which then were used to calculate the water and nutrients that should be provided to the plant. The inputs to the ANN were the image parameter of the chili plant (height, width, area, R, G, B) and the age of the chili plant. The outputs from the ANN were water and nutrient status.

#### ANNs for Controlled Post-Harvest Handling

ANNs have also been used in the post-harvest handling of tropical agricultural products. Rejo [49] developed a model for classifying fruits from durian plants (*Durio zibethinus* Murr.) based on maturity and ripeness using a multi-layer ANN. The inputs to the ANN were nondestructive characteristics of the durian fruits, i.e., density and zero moment power. The outputs were the level of maturity and ripeness of the durian fruit. The zero moment power was an acoustic characteristic, which could be measured by ultrasonic wave. Suyantohadi [50] developed a model to identify the maturity level of mango (cv. *arumanis*) using an ANN. The inputs to the ANN were two color variables of the mango, fruit density, length, and thickness. The output from the ANN was formed into binary digits representing groups in the maturity levels of mango.

An ANN has been applied to predict seed weight inside lanzone fruits [51]. The fruit, from the tropical plant lanzone (*Lanszium domesticum* Corr.), consists of 4 to 6 segments and has a white pulp and sweet-sour taste. However, a fruit commonly contains 2 to 3 segments with a bitter seed that bothers consumers if the seeds are not carefully removed prior to or during the consumption. The inputs to the ANN were fruit weight, diameter, and transmittance intensity ratio (the ratio of the transmittance

value of the visible light through the fruit to the transmittance value of the reference). The output was the weight of seeds inside the fruit.

A neural network classifier was applied to differentiate between 2- to 3-week old sunflower plants and common cocklebur weeds of similar size, shape, and color [52]. Color images were obtained by a digital camera in natural sunlight. A specific objective was to minimize the subsequent image processing operations needed to enhance the images and to extract the features needed by a back-propagation neural network classifier. Neural network structures with different numbers of hidden layers and neurons in them were tested to find the optimal classifier. The maximum number of correctly recognized images in distinguishing weeds from sunflower plants was 71 (out of 86), while it was 82 and 74 in separating sunflower and weed images from bare soil images, respectively.

An artificial neural network model was presented for the prediction of viscosity of fruit juice as a function of concentration and temperature [53]. The fruit juices were orange, peach, pear, *Malus floribunda*, and black current. The viscosity data of juices (1.53 to 3300 mPa s) were obtained from the literature for a wide range of concentration (5° to 70°Brix) and temperature (30.7° to 71.7°C). The model was able to predict viscosity with a mean absolute error of 3.78 mPa s. Predicting viscosity using the ANN has proved to be a simple, convenient, and accurate method. The model could be incorporated in the heat transfer calculations during fruit processing, where concentration and temperature-dependent viscosity values were required. This might also be useful in mass transfer calculations during the filtration of juice using membranes for clarification.

### ANNs for Runoff Forecasting

A specific dynamic neural network called a *state space neural network* (SSNN) has been modified to perform short-term rainfall runoff forecasts [54]. A new learning method developed from the interchange of the roles of the network states and the weight matrix was applied to train the SSNN and helped the network to evolve into a time-variant model while forecasting the rainfall runoff process. A study case was implemented in Taiwan's Wu-Tu watershed, where the runoff path lines were short and steep. Forty-seven events from 1966 to 1997 were forecasted via the SSNN, and the results were validated via various criteria. The convergence of the new learning algorithm was shown during the model training process. Performances of the SSNN for short term rainfall runoff forecasting revealed that the specific dynamic recurrent neural network was appropriate for hydrological forecasts.

### 9.3.4 Fuzzy Control for the Processing Industry

### Fuzzy Control for the Tea Withering Process

The quality of tea is dependent on the experience of the operators of tea processing machines. Fuzzy logic control can be applied for controlling the tea processing. Basically, black tea processing consists of a series of operations in the following order: partial removal of moisture (withering); leaf disintegration into small pieces (cutting,

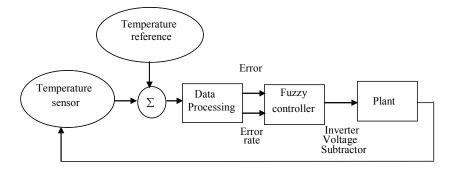


Figure 8. Fuzzy logic control system for temperature of tea withering [55].

rolling); quality development by exposure to air (oxidation or fermentation); completion of moisture removal (drying or firing); and sieving into size fractions with fiber removal (sorting). The complexity of the temperature control processes in the withering trough, however, makes it difficult to apply and implement analytical control design techniques that rely on a mathematical model of the plant to be controlled.

The fuzzy logic approach to temperature control in withering process has been implemented in the Tea Research Centre, Gambung Area, Indonesia [55]. The input of the developed control system consists of temperature error (the difference between the temperature of the withering chamber and the setpoint) and the temperature error rate of variation. The output of the controller consists of the reduction number, i.e., the number subtracted from inverter voltage at discrete time n-l to obtain inverter voltage at discrete time n. Major components for the tea withering control system included the sensor, actuator (inverter), ADC (analog-digital converter), DAC (digital-analog converter) and the plant. The plant consisted of a withering chamber, heater, and blower. A block diagram of control system is presented in Figure 8.

### Fuzzy Control for the Tea Rolling Process

Leaf disintegration into small pieces in black tea processing is described in a number of ways, e.g., rolling, cutting, crushing, tearing, but the basic requirements are size reduction with a degree of cell disruption to allow the exposure of the new surfaces to air in the subsequent fermentation stage. A type of machinery used at the rolling process is called the *orthodox machine*.

The orthodox machine is a batch machine. A weighed batch of withered leaf is placed in a cylindrical hopper positioned above a larger circular table, which has a series of ridges or battens across the surface. The hopper and table move in an eccentric fashion with regard to one another, causing the leaf at the bottom of the cylindrical hopper to be rolled, bruised, and broken up into fragments. The top of the inside of the cylindrical hopper is fitted with a cap, which is used to apply downward pressure. The rolling period is generally in the range 20 to 30 min. The quality of black tea produced by the orthodox machine relies on the temperature of leaf during the rolling process, which depends on the pressure and speed of rolling.

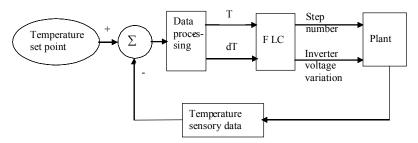


Figure 9. Block diagram of a fuzzy logic controller (FLC) for the tea rolling process [56].

A fuzzy logic controller (FLC) was applied to achieve the desired temperature in the press cap roller machine [56]. The FLC was implemented to perform as a multiple-input, multiple-output (MIMO) controller with temperature error and its rate as inputs and inverter voltage changes and variations of step number in the stepper as outputs. Changes in voltage provided a control signal to produce speed variation in the rolling motor, whereas variations in step number of stepper motor generated a control signal to adjust the pressing cylinder position. The block diagram is shown in Figure 9.

### 9.3.5 Genetic Algorithms for Bioresources and Bioprocesses

Plankton has been identified by feature-based soft computing consisting of an image processing and genetic algorithm [57]. The image of the plankton was segmented using the Sobel method for obtaining feature boundaries. The result was an  $n \times n$  matrix containing values of 0 or 1 for each matrix element representing a feature of plankton. The plankton feature was evaluated based on shape. Three geometrical shapes—circular, elliptical, rectangular—were used as the standard shapes for the plankton. The distances between the edge diagonally to the shape were unique for each geometry. There were four distances for each shape. The standard geometric function was calculated as follows:

$$F(s) = \sum d(i) \tag{1}$$

where F(s) = standard geometric function

d(i) = distance between the edge diagonally to the shape of a standard geometry Let F(m) represent the function of a plankton being classified, thus,

$$F(m) = \sum dp(i) \tag{2}$$

where dp(i) = distance between the edge diagonally to the shape of a classified object. In order to classify, F(m) must be compared with F(s). The difference, called error, was used to search the closest shape.

$$Er = |F(s) - F(m)| \tag{3}$$

The genetic algorithm to identify the plankton shape was divided into a series of cyclic steps as follows: (1) initialization, in which an initial potential solution is generated randomly; (2) evaluation, in which each individual of the population was evaluated to minimize error, Min(Er); (3) selection and crossover; and (4) mutation.

### Genetic Algorithms for Fermentation Processes

The growth of *Saccharomyces cerevisiae* in a batch-fed process indicates nonlinearities, describing the complex and time-varying nature of the fermentation process. A non-linear mathematical model for the growth of *S. cerevisiae* has been developed consisting of four variables, two inputs, and five parameters [58]. The stated variables were biomass concentration, glucose concentration, product concentration, and broth volume. The inputs were feed rate and feed concentration. The parameters in the model were  $\mu_m$ , specific growth rate; ks saturation constant/half velocity constant;  $Y_p$  product yield;  $Y_k$  biomass yield; and  $q_p$  product formation rate. These five parameters were optimized using the genetic algorithm. In this optimization, the performance criteria were given by an objective function F (error criteria) as follows:

$$F = \sum (Y_m - Y_d)^2 \tag{4}$$

where  $Y_m$  is the output of the model, and  $Y_d$  is the output of experimental data.

The genetic algorithm was coded in MATHLAB® with the followings steps:

- 1. The initial population is generated randomly with certain number and length;
- 2. The fitness of each string of the population is evaluated to the objective function *F* in Equation 4;
- 3. The strings are reproduced with a probabilistic method;
- 4. Crossover between reproduced strings with certain crossover probability; and
- 5. Repeat step 4 until the defined limit is met.

### Genetic Algorithms for Watershed Optimization

Genetic algorithms have been employed for the estimation of model parameters and function optimization in the Kashinagar watershed of the Vamsadhara River basin, Orissa, India [59]. Daily rainfall and runoff data collected for the years 1984-1995 were used for the development and optimization of the runoff prediction model. The prediction model was developed by considering the process as non-linear and dynamic in nature. The predicted values by the non-linear rainfall runoff prediction model were 1 to 2 days ahead of the measured values. This time problem was corrected by applying a backward shift operator technique, which improved the correlation coefficient by about 15%.

### 9.4 Possible Uses of Advanced IT in the Future

H. K. Purwadaria

The use of GIS (geographic information systems), remote sensing, radar, and their derivatives has been explored in some areas of agriculture. Many researchers have used this system accessing crop yield to support improvement of farming efficiency in anticipation of future limitation of suitable land, water resources, and crop productivity. For example, monitoring of cassava production has been assessed in Nakhon Ratchasima Province, Thailand, using remote sensing and GIS [60]. Factors influencing cassava productivity, among others, include soil moisture, drainage, organic matter

content, and fertility level. Factors were evaluated for land suitability classification to identify potential land productivity.

Monitoring of growth stages of wetland paddy has been investigated using spaceborne multidata SAR (synthetic aperture radar) images in Kedah State, Malaysia, and was found to be useful for delineating and differentiating various growth stages of paddy [61].

The ability of RADARSAT, equipped by SAR, was assessed in understanding the backscatter response as a function of the rice crop growth in Zhaoqing, China [62]. Early results indicated that the backscatter change for the paddy was variable, attributed to non-homogeneous growing conditions and the variability of the vegetative phase. More work is required to make the RADARSAT imagery usable for forecasting the yield of rice crop in a given area.

Researchers from the Taiwan Agricultural Research Institution, Wufeng, reported experiment modeling of rice growth (leaf numbers, plant height, leaf area index, leaf dry weight, and aboveground dry weight) from characteristics of reflectance spectra. The rice growth was reasonably assessed and monitored, either from the ground or satellites, when proper wavebands were selected [63].

GIS and remote sensing was also utilized to predict tea yield in Sri Lanka [64] by correlating tea LAI (leaf area index) with optical remote sensing parameters, expressed in normalized difference vegetation index (NDVI). A suitable model of tea yield was developed by selecting weights of considered variables such as existing spatial, meteorological, and agronomic variables, and LAI values.

Some development also improved the use of web GIS in providing environmental, social, economic, and geographic information through the Internet [65]. A GIS application when developed on ArcView could easily be converted to XML (Extensible Markup Language) and deployed on any web server producing any thematic maps that interactively changed their pattern based on a user query on the browser window. Examples of thematic maps produced by the system were available nitrogen and mineral status in the soil. This system was developed and on trial in Mahaboob Nagar district, India

#### References

- Akbar, S. U. 2004. ICT and social transformation in rural Bangladesh. Accessed on February 15, 2005, from http://www.iimahd.ernet.in/egov/ifip/dec2004/ dec2004.htm.
- Government of Maharashtra, India. 2005. Integrated wireless in local loop project: Maharashtra Tele-Centre. Accessed on February 15, 2005, from http:// www.mssrf.org.
- 3. Tractor and Farmer Equipments India Ltd. 2005. J-FarmIndia.Com Agricultural Information and Services. Accessed on February 15, 2005, from http://www.jfarmindia.com.
- 4. EID Parry Enterprise. 2005. India Agriline: Agricultural Information. Accessed on February 15, 2005, from http://www.agriline.com.

- 5. The World Bank and UN ESCAP. 2005. About Thai Rural Net project. Accessed on February 15, 2005, from http://www.ThairuralNet.org.
- 6. The ICT Development Group. 2005. Agricultural Information Network. Accessed on February 15, 2005, from http://www.ictdevgroup.com/ case\_study.php?story=3.
- 7. Taylor, M. M. J. 1995. Evaluation of multimedia extension program in Honduras. *Economic Development and Cultural Changes* 43(4): 821-834.
- 8. Government of India. 2005. Grameen Sanchar Sewak Project. Accessed on February 15, 2005, from http://www.itforchange.net/ict4d/display/66.
- 9. National Science and Technology Board of India. 2005. Information Dissemination. Accessed on February 15, 2005, from http://www.insteb.com.
- Tafzi, F. 1999. Pengembangan sistem informasi untuk penanganan pasca panen mangga (Development of information system for postharvest handling of mango). Master's thesis. Indonesia: Post Harvest Technology Study Program, Bogor Agricultural University.
- 11. Sahrial. 1999. Analisis dan rancangan sistem informasi untuk penanganan pascapanen padi (Analysis and design of information system for postharvest handling of paddy). Master's thesis. Indonesia: Post Harvest Technology Study Program, Bogor Agricultural University.
- Asia Pacific Development Information Program of UNDP. 2005. Achieving Universal Access: Developing a Philippine Business Model and Government Interventions Strategies for Viable Community Telecenters in Rural Areas. Accessed on February 15, 2005, from http://www.apdip.net/projects/ictrnd/ 2004/L62-ph/.
- 13. National Dairy Development Board, India. 2005. Dairy Information System Kiosk Project (DISK). Accessed on February 15, 2005, from http://www.iimahd.ernet.in/ egov/disk.htm.
- Asia Pacific Development Information Program of UNDP. 2005. ICT Serving for Agriculture and Rural Development in an Prototype County. Accessed on February 15, 2005, from http://hephaestus.apdip.net/projects/ictrnd/2002/china/.
- 15. Astika, I W., R. Cahyoutomo, and F. X. Mulyantara. 2001. Development of computer software for optimum scheduling of farm machinery operations in a sugar cane plantation. *Proc. of 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications*.
- Solahudin, M., K. B. Seminar, and Y. Suharnoto. 2000. Internet-based expert system for diagnosing hand tractor troubles – Sistem pakar berbasis Internet untuk diagnosa kerusakan traktor tangan. *Buletin Keteknikan Pertanian* 14(2): 118-127.
- 17. Pertiwi, S., C. Sutanto, M. Djojomartono, and A. Permana. 2000. Commodity based database system for agriculture and agribussiness development. *Proc.* 2<sup>nd</sup> *Asian Conference for Information Technology in Agriculture*.
- 18. Suhardiyanto, H., K. B. Seminar, Y. Chadirin, and B. I. Setiawan. 2001. Development of a pH control system for nutrient solution in ebb and flow

- hydroponic culture based on fuzzy logic. *Proc.* 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications.
- 19. Toomanian, N., A. S. M. Gieske, and M. Akbary. 2004. Irrigated area determination by NOAA-Landsat upscaling techniques, Zayandeh River Basin, Isfahan, Iran. *International J. Remote Sensing* 25(22): 4945-4960.
- 20. Lu, D., M. Batistella, E. Moran, and P. Mausel. 2004. Application of spectral mixture analysis to Amazonian land-use and land-cover classification. *International J. Remote Sensing* 25(23): 5345-5358.
- 21. Solanki, H. U., R. M. Dwivedi, S. R. Nayak, V. S. Somvanshi, D. K. Gulati, and S. K. Patnayak. 2003. Fishery forecast using OCM chlorophyll concentration and AVHRR SST: Validation results off Gujarat coast, India. *International J. Remote Sensing* 24(18): 3691-3700.
- 22. Han, S., Q. Zhang, B. Ni, and J. F. Reid. 2004. A guidance directrix approach to vision-based vehicle guidance systems. *Computers and Electronics in Agriculture* 43(3): 179-186.
- 23 Koumporous, Y., Y. Mahaman, M. Maliappis, H. C. Passam, A. D. Sideridis, and V. Zorkadis. 2004. Image processing for distance diagnosis in pest management. Computers and Electronics in Agriculture 44(2): 121-132.
- 24. Kavdir, I. 2004. Discrimination of sunflower, weed and soil by artificial neural networks. *Computers and Electronics in Agriculture* 44(2): 153-161.
- 25. Reyniers, M., E. Vrindts, and J. De Baerdemaeker. 2004. Optical measurement of crop cover for yield prediction of wheat. *Biosystems Engineering* 89(4): 383-395.
- 26. William III, L, and T. E. Martinson. 2004. Nondestructive leaf area estimation of 'Niagara' and 'DeChaunac' grapevines. *Scientia Horticulturae* 98(4): 493-499.
- 27. Suroso, I., D. M. Subrata, K. Soelistiadji, and J. Pitoyo. 2001. Development of chili growth control system using artificial neural network. *Proc.* 2<sup>nd</sup> *IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications*.
- 28. Budiastra, I W., D. Saputra, and H. K. Purwadaria. 1995. Assessment on physical characteristics of mango cultivar *gedong* using image processing. *Proc. The Role of Agricultural Engineering to Support Agricultural Industrialization Era.*
- 29. Wahyono, A., H. K. Purwadaria, Suroso, and Irawati. 2003. Development of grading algorithm for Dendrobium orchid using image processing. *Proc. of the* 21<sup>st</sup> ASEAN/3<sup>rd</sup> APEC Seminar on Postharvest Technology.
- 30. Stanko, D., M. Lakota, and M. Hoevar. 2004. Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging. *Computers and Electronics in Agriculture* 42(1): 31-43.
- 31. Ahmad, U., N. Kondo, S. Arima, M. Monta, and K. Mohri. 1998. Weed detection in lawn field based on gray scale uniformity. *Proc.of the 3<sup>rd</sup> International Conference on Advanced Mechatronics*.
- 32. Mehl, P. M., Y. R. Chen, M. S. Kim, and D. E. Chan. 2004. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *J. Food Engineering* 61(1): 67-72.

- 33. Tao, Y., C. T. Morrow, P. Heinemann, and H. J. Sommer. 1995. Fourier-based separation technique for shape grading of potatoes using machine vision. *Trans. ASAE* 38: 949-957.
- 34. Tao, Y., Z. Verghese, C. T. Morrow, J. H. Sommer, and P. Heinemann. 1995. Machine vision for color inspection of potatoes and apples. *Trans ASAE* 38: 1555-1561.
- 35. Arham, Z., U. Ahmad, and Suroso 2004. The evaluation of maturity and ripeness of the limonia using image processing and artificial neural network. *Jurnal Keteknikan Pertanian* 18(3): 210-218.
- 36. Damiri, D. J., U. Ahmad, and Suroso. 2004. Identification of maturity and ripeness of the lime using image processing and artificial neural network. *Jurnal Keteknikan Pertanian* 18(1): 48-60.
- 37. Park, B., K. C. Lawrence, W. R. Windham, and Y. R. Chen. 2002. Discriminant analysis of dual-wavelength spectral images for classifying poultry carcasses. *Computers and Electronics in Agriculture* 33(3): 219-232.
- 38. Tan, J. 2004. Meat quality evaluation by computer vision. *J. Food Eng.* 61: 27-35.
- 39. Gao, X. 1996. Analysis of expanded-food texture by image processing. Part. I. Geometric properties. Part II. Mechanical properties. *J. of Food Process Eng.* 19: 425-456.
- 40. Gao, X., J. Tan, P. Shatadal, and H. Heymann. 1999. Evaluating expanded-food sensory properties by image analysis. *J. Texture Studies* 30: 291-304.
- 41. Ikeda, Y., I W. Budiastra, and T. Nishizu. 1992. On predicting concentrations of individual sugars and malic acid of the fruits by near-infrared reflectance spectrometry. *Proc. of Advances on Agricultural Engineering and Technology*, Vol. II.
- 42. Purwadaria, H. K., I W. Budiastra, and D. Saputra. 1995. Near infrared reflectance testing to predict sucrose and malic acid concentrations of mangoes. *Proc. of Control Applications in Post-Harvest and Processing Technology*.
- 43. Saputra, D., I W. Budiastra, and H. K. Purwadaria. 1995. Classification of mango by near infrared diffuse reflectance. *Proc. of Food Processing Automation IV*.
- 44. Hahn, F., I. Lopez, and G. Hernandez. 2004. Spectral detection and neural network discrimination of *Rhizopus stolonifer* spores on red tomatoes. *Biosystems Engineering* 89(1): 93-100.
- 45. Budiastra, I W., A. Trisnobudi, and H. K. Purwadaria. 1999. Ultrasound system for automation of internal quality evaluation of durian. *Proc. of the 14<sup>th</sup> World Congress IFAC*, Vol. K.
- 46. Haryanto, B., I W. Budiastra, H. K. Purwadaria, and A. Trisnobudi. 2001. Determination of acoustics properties of durian fruits. *Proc. of the 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications*.
- 47. Scotford, I. M., and P. C. H. Miller. 2004. Estimating tiller density and leaf area index of winter wheat using spectral reflectance and ultrasonic sensing techniques. *Biosystems Engineering* 89(4): 395-409.

- 48. Harmen, and A. H. Tambunan. 2005. *Penentuan sifat dielektrik pertanian* (Determination of dielectric properties for agricultural products). *Jurnal Keteknikan Pertanian* 18(3): 153-160.
- 49. Rejo, A., Suroso, I W. Budiastra, H. K. Purwadaria, S. Susanto, and Y. Y. Nazaruddin. 2001. Model for predicting and classifying durian fruit based on maturity and ripeness using neural network. *Proc. of 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications*.
- 50. Suyantohadi, A., G. T. Mulyati, W. Supartono, and T. F. Djafar. 2001. Identification of the maturity level of mango "Arumanis" using artificial neural network. *Proc. of 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications*.
- 51. Hendri, Suroso, H. K. Purwadaria, S. E. Widodo, and I W. Budiastra. 2001. Neural network application on non-destructive evaluation of lanzone using visible light. *Proc. of 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications*.
- 52. Kavdır, I. 2004. Discrimination of sunflower, weed and soil by artificial neural networks (abstract). *Computers and Electronics in Agriculture* 44 (2): 153-160. Accessed on February 16, 2005 from http://www.sciencedirect.com/.
- 53. Rai, P., G. Majumdar, S. DasGupta, and S. De. 2004. Prediction of the viscosity of clarified fruit juice using artificial neural network: a combined effect of concentration and temperature (abstract). *J. Food Engineering* 68(4): 527-533. Accessed on February 16, 2005 from http://www.sciencedirect.com/.
- 54. Pan, T. Y., and R. Y. Wang. 2004. State space neural networks for short term rainfall-runoff forecasting (abstract). *J. Hydrology* 297(1-4): 34-50. Accessed on February 16, 2005 from http://www.sciencedirect.com/.
- 55. Bambang, R., Fatih, and T. Abas. 2001. Intelligent temperature control in tea withering process using fuzzy logic: An experiment study. *Proc. of 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications*.
- 56. Bambang, R., S. F. Mahbub, and T. Abas. 2001. Application of fuzzy logic control to tea rolling process. *Proc. of 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications*.
- 57. Purnomo, M. H., W. T. Sesulihatin, and O. Puspitorini. 2001. Plankton feature identification by genetic algorithm. *Proc. of 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control for Agricultural Applications*.
- 58. Joelianto, E., R. Dwihandayani, and L. Ling. 2001. Parameter estimation using genetic algorithm for fed-batch growth of *Saccharomyces cerevisiae*. *Proc. of 2<sup>nd</sup> IFAC-CIGR Workshop on Intelligent Control on Agricultural Applications*.
- 59. Agrawal, R. K., and J. K. Singh. 2003. Application of a genetic algorithm in the development and optimisation of a non-linear dynamic runoff model. *Biosystems Engineering* 86(1): 87-95.
- 60. Eiumnoh, A., R. P. Shrestha, S. Baimoung, P. Kesawapitak, and A. Noomharm. 2005. Field prediction and potential monitoring of cassava production in Nakon Ratchasima. Accessed on February 18, 2005, from http://www.GISdevelopment.net/agriculture/.

- 61. Abubakar, S. B. B., and A. T. B. Shaari. 2005. SAR Backscatter response of various growth stages of wetland rice paddy in Malaysia. Accessed on February 18, 2005, from http://www.GISdevelopment.net/agriculture/.
- 62. Staple, G. C., and J. Hurley. 2005. Rice crop monitoring in China using RADARSAT SAR-Initial Results. Accessed on February 18, 2005, from http://www.GISdevelopment.net/agriculture/.
- 63. Yang, C. M., and M. R. Su. 2005. Modelling rice growth from characteristics of reflectance spectra. Accessed on February 18, 2005, from http://www.GISdevelopment.net/agriculture/.
- 64. Rajapakse, R. M. S. S., N. K. Tripathi, and K. Honda. 2005. Modelling tea (*Camellia* (L) O. Kuntze) yield using satellite derived (a), land use and meteorological data. Accessed on February 18, 2005, from http://www.GISdevelopment.net/agriculture/
- 65. Veeranna, B., J. Jayachandra, and S. Gopi. 2005. Inter-active online mapping with XML for agriculture information system at district level. Accessed on February 18, 2005, from http://www.GISdevelopment.net/agriculture/.