Woodhead Publishing in Food Science and Technology

Food process modelling

Edited by L. M. M. Tijskens, M. L. A. T. M. Hertog and B. M. Nicolaï

WOODHEAD PUBLISHING LIMITED

Food process modelling

Edited by
L M M Tijskens, M L A T M Hertog and
B M Nicolaï

CRC Press
Boca Raton Boston New York Washington, DC

WOODHEAD PUBLISHING LIMITED
Cambridge England

Published by Woodhead Publishing Limited
Abington Hall, Abington
Cambridge CB1 6AH
England
www.woodhead-publishing.com

Published in North and South America by CRC Press LLC
2000 Corporate Blvd, NW
Boca Raton FL 33431
USA

First published 2001, Woodhead Publishing Limited and CRC Press LLC

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

Woodhead Publishing Limited ISBN 1 85573 565 2
CRC Press ISBN 0-8493-1224-8
CRC Press order number: WP1224

Cover design by The ColourStudio
Project managed by Macfarlane Production Services, Markyate, Hertfordshire (macfarl@aol.com)
Typeset by MHL Typesetting Limited, Coventry, Warwickshire
Printed by T J International, Padstow, Cornwall, England

Related titles from Woodhead's food science, technology and nutrition list:

Instrumentation and sensors for the food industry (ISBN: 1 85573 560 1)

The first edition of this book quickly established itself as a standard work in its field, providing an authoritative and practical guide to the range of instrumentation and sensors available to the food industry professional. This new edition has been comprehensively revised to include new developments and techniques.

Food processing technology: Principles and practice (ISBN: 1 85573 533 4)

The first edition of *Food processing technology* was quickly adopted as the standard text by many food science and technology courses. The publication of this completely revised new edition is set to confirm the position of this textbook as the best single-volume introduction to food manufacturing technologies available. New chapters include computer control of processing, novel 'minimal' technologies including processing using high pressures or pulsed electric fields, ohmic heating and an extended chapter on modified atmosphere packaging.

Foodborne pathogens (ISBN: 1 85573 454 0)

A practical guide to identifying, understanding and controlling foodborne pathogens. This book relates current research to practical strategies for risk analysis and control and is designed for both microbiologists and non-specialists, particularly those concerned directly with food processing operations. The first part of the book examines specific microbiological hazards. This is followed by an examination of risk assessment and the concluding section provides a guide to controlling pathogens throughout the supply chain from farmer to consumer.

Details of these books and a complete list of Woodhead's food science, technology and nutrition titles can be obtained by:

- visiting our web site at www.woodhead-publishing.com
- contacting Customer Services (e-mail: sales@woodhead-publishing.com; fax: +44 (0) 1223 893694; tel.: +44 (0) 1223 891358 ext.30; address: Woodhead Publishing Limited, Abington Hall, Abington, Cambridge CB1 6AH, England)

If you would like to receive information on forthcoming titles in this area, please send your address details to: Francis Dodds (address, tel. and fax as above; e-mail: francisd@woodhead-publishing.com). Please confirm which subject areas you are interested in.

Contents

Preface

Ever since man walked on the face of this earth, he has consumed food to survive, to maintain his body and mind, and later for pleasure too. In the early days, he had to gather his food wherever he found it. Gradually man started to understand how to grow and cultivate his own food, and he gained more and more control over his environment. All this time, man was curious about what he observed and the mechanisms behind it, longing to understand what happened, what happens to himself, to his food and to his environment.

In the current age of highly developed science and technology, the modern aid to analyse, interpret and understand our surrounding world is modelling. Modelling can be regarded as bringing together the concepts developed by product and process experts into a coherent and consistent entity. By subsequently translating such conceptual models into their mathematical equivalents and by implementing these into computer programs, such models can be used for quantitative analyses and, ultimately, for making predictions. Sometimes expert and modeller are united in one person. In that situation, dedicated models will emerge. When they are different people, the model will often include views that are more generic in nature.

Modelling can be conducted on an almost infinite number of levels, reflecting different degrees of real life, ranging from the purely theoretical to the completely empirical. Nowadays modelling is used in almost every discipline by an ever-increasing number of people. Few of these modellers had a dedicated education in this area, and most of them had to learn the trade the hard way, in daily practice by trial and error, without being aware of the pitfalls of this powerful tool.

In this book, we have attempted to bring some order and rules to the jungle of techniques available for modelling the processes and phenomena that play a part

in our daily food. We need to realise that what we taste and perceive in our food and that what we like in our food, is the result both of processes that occur naturally in our food and processes that we apply to it. On top of that, the techniques used for modelling are applied processes as well. The title of this book, *Food Process Modelling*, has, therefore, to be understood in terms of those three types of processes:

- Processes occurring in food
- Processes applied to food
- Processes applied to model food behaviour.

This book is subdivided into five major parts, each covering a selected area from theory or practice, either involving modelling techniques, or involving particular food processes to which the models are applied.

In Part I the principles and procedures of fundamental, deductive approaches of modelling are explained and discussed. The essence of deductive modelling is the conversion of theories and concepts into mathematical and computer formulations, virtually without applying information contained in the measured data. Data are only used to calibrate and validate the developed models.

In Part II the principles and procedures of empirical, inductive approaches of modelling are explained and discussed. The essence of inductive modelling is to extract as much useful information contained in the measured data as possible, without *a priori* knowledge of the processes involved. Data, for this type of modelling, are the only source of information available. Of course, reality is never that black or white, so models are not purely black or white either. In each of the chapters in Parts I and II some combination of the two types of modelling approaches will be found.

In Parts III and IV of this book, practical examples are provided in the area of production, processing and storage of fresh foods. Models discussed range from almost purely deductive to almost purely inductive. Part III covers the agricultural production, from fruits and vegetables to the dairy and meat sectors. Part IV looks at a range of processing technologies.

The last part of the book discusses aspects of quality and safety. It is not dedicated to one product or process, but to the conglomerate of actions the food industry has to take to bring food to consumers. The emphasis lies on quality and safety throughout the entire food chain, from production, through storage, transport and handling to retail, and finally to the consumer's place to maintain his body and mind, and hopefully to be enjoyed as well.

We would like to thank the staff of Woodhead Publishing Ltd for all their help and guidance.

Pol Tijskens
Maarten Hertog
Bart Nicolaï

Contributors

Chapter 1

Dr B. P. Hills
Institute of Food Research
Norwich Laboratory
Norwich Research Park
Colney
Norwich NR4 7UA
England

Fax: +44 1603 507723
E-mail: brian.hills@bbsrc.ac.uk

Chapter 2

Dr M. Sloof
Everest BV
Reitscheweg 55
5232 BX 's-Hertogenbosch
The Netherlands

Fax: +31 73 645 09 20
E-mail: m.sloof@everest.nl

Chapter 3

Professor Dr M. A. J. S. van Boekel
Wageningen Agricultural University
Department of Food Science
PO Box 8129
6700 EV Wageningen
The Netherlands

Fax: +31 317 484893
E-mail: tiny.vanboekel@ift.fdsci.wau.nl

L. M. M. Tijskens
ATO-BV
PO Box 17
6700 AA Wageningen
The Netherlands

Fax: +31 317 475 347
E-mail: l.m.m.tijskens@ato.wag-ur.nl

Chapter 4

Professor Dr B. M. Nicolaï
P. Verboven
N. Scheerlinck
Department of Agro-Engineering and Economics

Katholieke Universiteit Leuven
de Croylaan 42
B-3001 Heverlee
Belgium

Fax: +32 16 32 29 55
E-mail: bart.nicolai@agr.kuleuven.ac.be

Chapter 5

R. W. Sierenberg
Herbarenerf 5
2743 HA Waddinxveen
The Netherlands

E-mail: r.w.sierenberg@12move.nl

Chapter 6

F. Verdenius
Production and Control Systems
ATO BV
PO Box 17
6700 AA Wageningen
The Netherlands

Fax: +31 317 475 347
E-mail: f.verdenius@ato.wag-ur.nl

Dr L. Hunter
University of Colorado
School of Medicine
UCHSC Box C236
4200 E. 9th Avenue
Denver CO80262
USA

Fax: +1 303 315 1098
E-mail: larry.hunter@uchsc.edu

Chapter 7

Dr G. Holmes
Dr T. C. Smith
University of Waikato
Department of Computer Science
G Block
Hillcrest Road
Hamilton
New Zealand

Fax: +64 7 838 4155
E-mail: tcs@cs.waikato.ac.nz

Chapter 8

Professor Dr J. F. M. van Impe
K. Bernaerts
A. H. Geeraerd
F. Poschet
K. J. Versyck
Department of Food and Microbial Technology
Katholieke Universiteit Leuven
Kasteelpark Arenberg 23
B-3001 Heverlee
Belgium

Fax: +32 16 32 19 60
E-mail:
jan.vanimpe@agr.kuleuven.ac.be

Chapter 9

Dr L. F. M. Marcelis
Head of Cluster Cropping Systems
Plant Research International
PO Box 16
6700 AA Wageningen
The Netherlands

Fax: +31 317 42 31 10
E-mail: l.f.m.marcelis@plant.wag-ur.nl

Chapter 10

Mr C. Gary
Dr M. Tchamitchian
INRA
Domaine Saint-Paul
Site Agroparc
84914 Avignon cedex 9
France

Fax: +33 4 32 72 22 82
E-mail: gary@avignon.inra.fr

Chapter 11

Dr C. F. E. Topp
The Scottish Agricultural College
Agriculture and Food Economics Dept
Management Division
Auchincruive
Ayr KA6 5HW
Scotland

Fax: +44 1292 525 052
E-mail: k.topp@au.sac.ac.uk

Chapter 12

Dr K. G. Rickert
Director of Research
Faculty of Natural Resources,
Agriculture and Veterinary Science
The University of Queensland
Gatton Campus
Gatton
Q4343
Australia

Fax: +61 7 5460 1324
E-mail: krickert@uqg.uq.edu.au

Chapter 13

Dr J. Hofmeester
DSM-Gist Australia Pty. Ltd
PO Box 83
Moorebank NSW 2170
Australia

Fax: +61 2 9821 1442
E-mail:
joep.hofmeester@dsm-group.com

Chapter 14

M. L. A. T. M. Hertog
Food Technology Research Centre
Massey University
Private Bag 11 222
Palmerston North
New Zealand

Fax: +64 6 350 5610
E-mail: m.l.hertog@massey.ac.nz

Dr N. H. Banks
Zespri Innovation
PO Box 4043
Mt Maunganui South
New Zealand
Fax: +64 7 575 1645
E-mail: banksn@kiwi.co.nz

Chapter 15

Dr Q. Tuan Pham
School of Chemical Engineering and
Industrial Chemistry
The University of New South Wales
Sydney 2052
Australia

Fax: +61 2 9385 5966
E-mail: tuan.pham@unsw.edu.au

Chapter 16

Dr S. Bakalis
Dr P. W. Cox
Professor Dr P. J. Fryer
School of Chemical Engineering
University of Birmingham
Edgbaston
Birmingham B15 2TT
England

Fax: +44 121 414 5377
E-mail: p.j.fryer@bham.ac.uk

Chapter 17

E. C. Wilkinson
L. M. M. Tijskens
ATO-BV
PO Box 17
6700 AA Wageningen
The Netherlands

Fax: +31 317 47 53 47
E-mail: l.m.m.tijskens@ato.wag-ur.nl

Chapter 18

Dr J. Baranyi
Institute of Food Research
Norwich Research Park
Colney
Norwich NR4 7UA
England

Fax: +44 1603 507723
E-mail: jozsef.baranyi@bbsrc.ac.uk

Dr C. Pin
Departamento de Higiene y Tecnologia de los Alimentos
(Nutricion y Bromatologia III)
Facultad de Veterinaria
Universidad Complutense de Madrid
28040 Madrid
Spain

Fax: +34 91 3943743
E-mail: carmenpi@eucmax.sim.ucm.es

Chapter 19

Dr P. S. Taoukis
National Technical University of Athens
Department of Chemical Engineering
Div IV
Laboratory of Food Chemistry and Technology
5 Iroon Polytechniou
Zografou Campus
15780 Athens
Greece

Fax: +30 1 7723163
E-mail: taoukis@chemeng.ntua.gr

Chapter 20

Dr R. A. C. M. Broekmeulen
Department of Technology Management
Technische Universiteit Eindhoven
PO Box 513
5600 MB Eindhoven
The Netherlands

Fax: +31 402 46 45 31
E-mail: r.a.c.m.broekmeulen@tm.tue.nl

Chapter 21

L. M. M. Tijskens
A. C. Koster
J. M. E. Jonker
ATO-BV
PO Box 17
6700 AA Wageningen
The Netherlands

Fax: +31 317 47 53 47
E-mail: l.m.m.tijskens@ato.wag-ur.nl

Part I

The principles of modelling: fundamental approaches

Introduction

In modern science, all simple problems have been solved by now. What remains are problems that are so complex, one cannot unravel their mysteries by simple experiments and deductions. To tackle these complex problems, we have to decompose them into their constituting processes and each of these processes has to be studied, analysed and modelled separately. As a consequence, experimental research and analysis of resulting data relies more and more on theoretical, fundamental and generic models. These types of models are deduced from views and theories currently available. These deductive models possess the intrinsic opportunity of reusability and parameter transfer. This means that, for the same product, the same model and the same model structure can be used over and over again, in different situations for different problems while keeping the attached parameters and the parameter values actually the same. At the same time, this approach ensures that the models built and applied are firmly rooted into the contemporary views and theories of chemistry, biochemistry, physics and physiology as applied in agriculture.

So, it is a logical choice to start this book on food process modelling with an overview of the fundamental techniques of modelling.

In Chapter 1 an overview is given of the powers and pitfalls of deductive modelling based on three examples of processes frequently encountered in our food systems. This chapter describes the problems one encounters in making models useful for practical application, in terms of generic modelling and parameter estimation and parameter transfer.

Chapter 2 devotes its attention to the difficulties and rules of decomposing problems into their constituting processes. In the realm of artificial intelligence,

this technique has been known for a long time. In this discipline the processes are, based on the subject at hand, more or less defined by the researchers themselves. In food and agricultural research, the choice of processes is not free at all, and the main problem reduces to finding the processes occurring in nature, and to finding appropriate simplifications for these problems. One should include all processes necessary to describe adequately the problem but definitely not more. This chapter covers the translation of the rules and techniques, developed in artificial intelligence, to make them applicable for food and agricultural research.

In Chapter 3, the basic and fundamental rules of chemical and biochemical kinetics are described and made available for modelling food behaviour in the pre- and postharvest period. Many of the changes occurring in our food are of a chemical nature. Application of these rules and techniques are therefore of utmost importance for developing generic and applicable models for food and agricultural research.

In Chapter 4, the attention is devoted to modelling physical processes that frequently occur in our food or that are frequently applied to it. Heat and mass transfer are the processes most commonly occurring. Since heat and mass transfer are three-dimensional problems, describing and modelling the distribution in space and time inherently results in three-dimensional solutions. This constitutes an additional problem for fundamental modelling and results in most cases in very complex models and time consuming simulations.

Chapter 5 is a warm plea for the recognition of parallelism. Everything in daily life is the result of interacting parallel processes. This interaction between parallel processes typically results in discrete time events. As long as modellers limit themselves to model single continuous processes, discrete events can often be avoided. Other cases, like logistic problems, can be largely defined in terms of discrete events, neglecting minor continuous processes. Due to former technical problems in coping with continuous processes and discrete time events at the same time, the two modelling approaches developed separately for years. Using his experience as modeller and developer of modelling software, the author shows how the barrier between discrete and continuous modelling techniques can be removed to model real world systems fully utilising parallelism.

Pol Tijskens

1

The power and pitfalls of deductive modelling

B. P. Hills, Institute of Food Research, Norwich

1.1 Introduction

The dictionary defines the word ‘deduction’ as ‘inference from the general to the particular’, as opposed to ‘induction’, which is the ‘inference of the general law from particular instances’. In its idealized form, deductive modelling of food processing therefore starts with the general laws of (chemical-)physics, and uses them to build realistic mathematical models of food processing operations. The power and pitfalls of this deductive modelling approach are best illustrated with a real-world example that is familiar to most food scientists, namely the commercial drying of pasta. During the manufacture of pasta, such as spaghetti, the extrusion step is followed by drying in a hot airstream so that the spaghetti can be packed and stored as the familiar bundles of straight, hard spaghetti cylinders. The drying step has to be optimized very carefully because drying at too fast a rate sets up severe moisture gradients in the pasta, which causes differential shrinkage, bending and stress cracking. On the other hand, drying too slowly in a hot, humid atmosphere permits the growth of spoilage microorganisms. One way of optimizing the drying conditions is to develop mechanistic deductive models describing the radial mass transport of water through the spaghetti cylinders as a function of air velocity, humidity and temperature. To do this ‘deductively’ we could start by referring to a classic physics textbook such as Crank’s *The Mathematics of Diffusion*, and hypothesize that the drying process can be modelled as radial isothermal Fickian diffusion of water through the cylindrical pasta matrix together with a constant water diffusion coefficient. Crank presents many analytical solutions for this problem corresponding to various boundary conditions. Unfortunately, none of these solutions would be of real value for optimizing pasta drying

because they fail to take account of important aspects of the phenomenon. In particular no account is taken of

- the dependence of the local water diffusivity on the local water content
- the non-uniform radial shrinkage as water is lost
- the stress fields set up by the non-uniform shrinkage
- the fact that the surface boundary conditions will be time dependent because the surface dries and
- it ignores the possibility that the drying may not be truly isothermal so that the coupled equations of heat and mass transport would need to be solved.

These are formidable theoretical problems that have not yet been fully solved and they illustrate nicely some of the strengths and weaknesses of the mechanistic deductive modelling approach.

The strengths of deductive modelling include the following.

1. The models are based on established physical principles, in this example the equations describing mass and heat transfer. The models are therefore internally consistent and unphysical values for fitting parameters can be eliminated quickly.
2. Because the models are based on established physical principles, additional complexities such as stress cracking can, in principle, be incorporated systematically.
3. The effects of changing ingredients, such as from hard to soft wheat varieties, and processing conditions can, in principle, be given a rational explanation in terms of altered parameters such as transport coefficients.
4. Deductive modelling drives us ever deeper into fundamental science as we seek answers to questions such as how transport coefficients depend on ingredient modification, microstructure and composition.
5. Once a model has been shown to be satisfactory on a simple system it is merely a computational exercise to adapt it to different processing conditions and more complex sample geometry.

The last point emphasizes that the ‘universality’ of a deductive model is a meaningful test of its usefulness. If a model is deduced from established, fundamental laws then it should be widely applicable within the space of variables and systems described by the model. This is a major advantage of deductive modelling over empirical or inductive modelling.

On the other hand, there are some obvious difficulties in the deductive modelling approach. For example, consider the following points.

1. To be solvable as analytic solutions or, in some cases, to avoid numerical instabilities, the model sometimes has to be simplified by neglecting complications such as stress fields, non-isothermal diffusion and multi-component diffusion, non-linear effects and the like. The question then arises as to whether the model is sufficiently realistic to be useful.
2. Removing these simplifications greatly increases the number of unknown parameters that need to be independently measured or estimated and

increases the computing time. This, in turn, endangers the 'universality' and hence usefulness of the model.

3. Limits to our knowledge and/or resources usually place a limit on the depth to which deductive modelling can be applied in real food processing problems. It is then usual to introduce empirical parameters or parameters that are adjusted by comparison with experiment. For example, in pasta drying one could always probe ever deeper into fundamental aspects by developing models predicting the dependence of the moisture diffusion coefficient on pasta composition and microstructure. This would be important if the optimization included, for example, choosing the right combination of hard and soft wheat varieties in the extrusion step. Usually the price paid for ever deeper deductive modelling is an increasing number of molecular and microstructural parameters such as pore sizes, starch granule sizes and amylopectin branching structures. More usually an arbitrary cut-off point is chosen where empirical relationships are introduced into the model. In the case of pasta drying this could be an empirical relationship describing the dependence of the moisture diffusion coefficient on moisture content, composition and temperature. Of course, this often destroys the fundamental character of the deductive model so it is always better to postpone the introduction of empirical relationships as long as possible. In the limiting case where parameters have to be measured or determined for every new system or set of conditions the model has essentially zero predictive value.

It is also worth bearing in mind that spatially dependent deductive models need to be tested with spatially resolved experimental data. For example, it is quite possible to obtain reasonable fits to the drying curve for pasta (i.e. the total mass loss versus drying time) with a simple Fickian diffusion model with a moisture-dependent diffusion coefficient. But this apparent agreement does not show that the model is valid because important factors, such as shrinkage and stress, are not being modelled. In this example the space-time resolved moisture profiles obtained with MRI (Magnetic Resonance Imaging) provide a far more stringent test of the drying models, and highlight their inadequacy. In fact MRI is proving to be an invaluable technique in food processing science as it is capable of monitoring non-invasively the space-time evolution of mass transport, phase transitions, temperature and many important quality factors.[1]

Having listed some of the strengths and weaknesses of deductive modelling we now need to consider the more general issue of the role of this type of modelling in food manufacturing.

1.2 Deductive modelling and process optimization

The possibility of optimizing a food manufacturing operation by developing deductive models and computer simulations of all aspects of the production,

packaging, storage, cooking and consumption stages is a pipe-dream of every food engineer. While certain aspects of production and storage can be modelled realistically, many cannot, either because we are still ignorant of the underlying physico-chemical processes or because of the shear complexity of the computing problem. It is also important that we take a unified approach to the problem because there is little point in optimizing one aspect, such as a processing step, if the final product has an unacceptably short microbial 'keeping time' (or, more loosely, 'shelf-life') or if some other quality factor, such as its flavour or texture, is altered detrimentally. Yet there are few, if any, theoretical models which attempt to optimize more than one quality or processing aspect at a time, and developing an integrated approach must remain high on any future modelling agenda.

To illustrate the importance of this integrated approach we will, in what follows, discuss three separate mechanistic deductive models concerned with differing aspects of the whole food manufacturing operation. In the last section the shortcomings of this piecemeal approach to modelling will be discussed. The first model involves optimizing a processing operation such as baking, extrusion, drying or freezing and involves minimizing energy expenditure while retaining acceptable quality of the end-product. This requires detailed knowledge of the mass and heat transfer as well as the space-time evolution of any associated reactions and/or phase changes affecting food quality. As mentioned, this information can be supplied by non-invasive techniques such as magnetic resonance imaging (MRI), but modelling the MRI data with conventional heat and mass transport models presents several challenges and these will be discussed.

The second aspect involves predicting the 'keeping-quality' of the product of the processing operation. At least three interrelated processes need to be considered in keeping-quality and shelf-life prediction, namely microbial spoilage, spoilage by slow chemical reactions and spoilage by mass transport and phase changes. Several new deductive probabilistic models for the survival and growth of microorganisms in spatially heterogeneous food matrices will be presented and future developments outlined.

The third aspect to be considered concerns the eating stage, namely the challenge of modelling quality factors such as flavour release in the mouth. What is required here is the ability to predict the effect of changing a food formulation on the time-intensity flavour release profiles in the mouth. Each of these aspects will be considered in turn beginning with the processing stage.

Figure 1.1 shows a schematic of a typical industrial process, in this case an oven arrangement typical of a roasting, baking, drying or toasting operation. In this and all other types of food processing the manufacturing operation has to be optimized to produce the desired food quality with minimum cost. Process design and analysis attempts to do this by combining a theoretical simulation of the whole process with selective measurement of parameter values. Parameters in the simulation enter at various levels that are indicated in Fig. 1.1. There are those associated with the equipment design, such as the size of the air nozzles

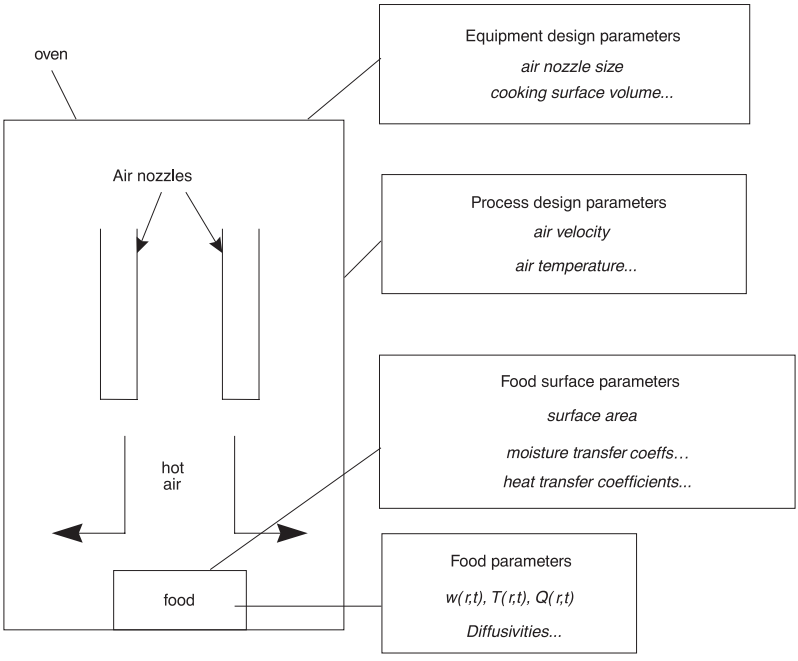


Fig. 1.1 A schematic of a dry cooking process in an industrial oven. The various types of parameter needed to simulate the cooking process are listed.

and the oven area to volume ratio. Others characterize the process design, such as air velocity, temperature and relative humidity. Then there are parameters characterizing the food surface such as the moisture and heat transfer coefficients. Finally there are parameters characterizing the macroscopic state inside the food, including the time dependent moisture content distribution, $w(\mathbf{r}, t)$, temperature distribution, $T(\mathbf{r}, t)$ and food quality factors, $Q_i(\mathbf{r}, t)$ as well as their associated transport parameters such as the moisture diffusivity, thermal diffusivity, solute diffusivities and chemical reaction rates. These parameters are functions of time, t, and of the position $\mathbf{r}$ (which denotes the coordinate x, y, z) within the food sample.

If one starts with the hypothesis that the heat and mass transport obey diffusion kinetics, then finite element simulations of the coupled heat and mass transport are possible. Ignoring the complications of stress cracking (called 'checking' in the case of biscuits) we can write, from first principles,

$$\partial W(\mathbf{r}, t)/\partial t = \nabla.\{D(\mathbf{r}, t)\nabla W(\mathbf{r}, t)\} \tag{1.1}$$

$$\partial T(\mathbf{r}, t)/\partial t = \nabla.\{K(\mathbf{r}, t)\nabla T(\mathbf{r}, t)\} - C_p(\partial W(\mathbf{r}, t)/\partial t) \tag{1.2}$$

Here C_p is the heat of vaporization and the last term in equation (1.2) is the heat required to evaporate the water, which couples the two equations together. Of course, the diffusion coefficient, $D(\mathbf{r}, t)$ and the thermal diffusion coefficient,

$K(\mathbf{r},t)$, also couple the equations because both are dependent on the local values of $W(\mathbf{r},t)$ and $T(\mathbf{r},t)$:

$$D(\mathbf{r},t) = F_1\{W(\mathbf{r},t), T(\mathbf{r},t)\} \tag{1.3}$$

$$K(\mathbf{r},t) = F_2\{W(\mathbf{r},t), T(\mathbf{r},t)\} \tag{1.4}$$

Unfortunately the form of the functions $F_1\{\ldots\}$ and $F_2\{\ldots\}$ is not, in general, known and have to be determined either independently or by fitting experimental data. The alternative is to go deeper in the deductive mode and work at the molecular and microstructural levels trying to derive the functional forms F_1 and F_2 from the basic physics of molecular diffusion including aspects such as restriction by microstructural obstructions, multicompartment diffusion, percolation theory and the like. These are non-trivial problems that would deter most researchers interested in process optimization and would mark the limit of the deductive mode of reasoning.

Of course, equations (1.1) to (1.2) need to be solved with appropriate surface boundary conditions involving changing surface moisture content, surface temperature and surface topology if there is non-uniform shrinkage. Unfortunately, in many real foods, even this scheme is inadequate because mass transfer does not necessarily proceed by diffusion. Indeed, the heterogeneous and porous structure of many real foods means that capillary forces can dominate liquid transport and fast transport via the vapour phase cannot be neglected.

Experimental input from non-invasive techniques such as MRI can help us identify the transport mechanism by providing real-time measurements of moisture distributions $W(\mathbf{r},t)$ and temperature distributions $T(\mathbf{r},t)$ and, in some cases, the changes in certain quality factors, $Q_i(\mathbf{r},t)$, both in the bulk food and at the food surface.[1] As more MRI data becomes available, it is apparent that many of our current transport models are failing to describe the real space-time behaviour of the processing operation. Moreover, the MRI data can often be acquired on a range of distance scales from the microscopic to macroscopic and here also the data highlights our inability to predict macroscopic transport behaviour from underlying changes in food microstructure and in the microscopic distribution of phase changes and mass distribution.

1.3 Modelling the keeping-quality and shelf-life of foods

Optimizing a processing operation such as drying or freezing is only the first of many aspects to be considered in food manufacture. If the end-product of the process has an unacceptably short shelf-life for whatever reason, then we can hardly say that the manufacturing process has been ‘optimized’. Space does not permit a detailed analysis of all the physico-chemical aspects of food shelf-life, such as retrogradation and Maillard reactions. Instead we will focus on the microbiological aspects of food stability. The empirical modelling approach,

such as the popular food micromodel[11] is useful for first estimating the survival and growth behaviour of a food-borne pathogen in a food of known pH, water activity and temperature. However, being empirical, the predictions are merely intrapolations of measured survival, lag and doubling times, and cannot be extrapolated outside the experimental parameter space. Deductive approaches, such as that described in Hills *et al.*[2] and Hills and Wright,[3] do not suffer from this shortcoming, and are capable of including probabilistic aspects of survival and growth.

Probabilistic aspects are of profound importance when attempting to model the microbial stability of foods. Most food matrices are sufficiently viscous or rigid to immobilize bacterial cells. If a cell exits the lag or resuscitation phase and enters the exponential growth phase it will form an isolated (micro-)colony in the food matrix. The observation that a gelatine gel growth medium inoculated with a stationary phase culture gives rise to microcolonies differing in size is therefore indicative of a distribution of lag times and/or doubling times within the population. In these circumstances it is not the average lag time that determines the microbial shelf-life of the food, but rather the cell with the shortest lag time. It is therefore essential to develop probabilistic models of the survival and growth of food-borne pathogens in heterogeneous foods. One approach to the development of probabilistic deductive models, which is being researched at the author's laboratory, is to define various sub-populations of bacterial cells within a culture and assume a Poisson random process characterized by a constant transition probability per unit time for transitions between the sub-populations. The following outlines this development for the case of lagged exponential growth. Generalizations to injury and resuscitation are also being developed.

Figure 1.2 is a schematic of microcolonies in a food matrix, in this case, a gelatine gel containing the synthetic growth medium, MOPS. The total number of cells in the food, $N(t)$, can be regarded as the sum of two sub-populations, the first are the single cells that are still in the lag phase, $n_{\text{lag}}(t)$, the second are all the cells in the exponential growth phase, $n_{\text{exp}}(t)$,

$$N(t) = n_{\text{lag}}(t) + n_{\text{exp}}(t) \tag{1.5}$$

where, initially, $n_{\text{lag}}(0) = N(0)$ and $n_{\text{exp}}(0) = 0$. We can derive a lagged growth expression for $N(t)$ by postulating a Poisson random process such that the probability that a particular cell has a lag time τ is given as:

$$\begin{aligned} p(\tau) &= e^{[-\tau/<\tau>]}/<\tau>, \quad 0 \leq \tau \leq \infty \\ &= 0 \qquad\qquad\qquad\quad \text{otherwise} \end{aligned} \tag{1.6}$$

Here $<\tau>$ is the mean lag time. Clearly the probability, $p_{\text{lag}}(t)$, that a cell is still in the lag phase after time t is,

$$p_{\text{lag}}(t) = \int_t^{\infty} p(\tau)d\tau = e^{[-t/<\tau>]} \tag{1.7}$$

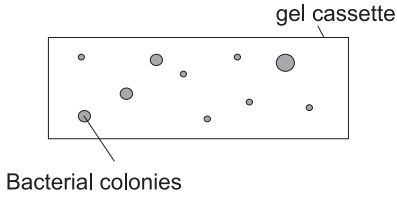


Fig. 1.2 A schematic showing the existence of a colony size distribution in a gelled growth medium.

$$n_{\text{lag}}(t) = N(0)p_{\text{lag}}(t) = N(0)e^{[-t/<\tau>]} \tag{1.8}$$

$$dn_{\text{lag}}/dt = -kn_{\text{lag}} \quad \text{where } k = 1/ <\tau> \tag{1.9}$$

An equation for $n_{\text{exp}}(t)$ can be derived in an analogous way. Let A be the probability per unit time that a cell in the exponential phase divides; then

$$dn_{\text{exp}}/dt = (kn_{\text{lag}} + An_{\text{exp}}). \tag{1.10}$$

The equation of evolution of $N(t)$ is obtained by solving these equations with the result,

$$N(t) = n_{\text{exp}}(t) + n_{\text{lag}}(t) \tag{1.11}$$

$$N(t) = N(0)\{ke^{At} + Ae^{-kt}\}/(k + A) \tag{1.12}$$

where $t_d = 1\text{n}\,2/A$ is the doubling time. Equation (1.12) is remarkable because it was first derived from kinetic considerations of the cell cycle in Hills and Wright.[3] It has the typical form of a lagged exponential shown schematically in Fig. 1.3.

In the limit of $t >> k^{-1}$ only exponential growth occurs, so that

$$N(t) = N(0)e^{A(t-t_{\text{eff}})} \tag{1.13}$$

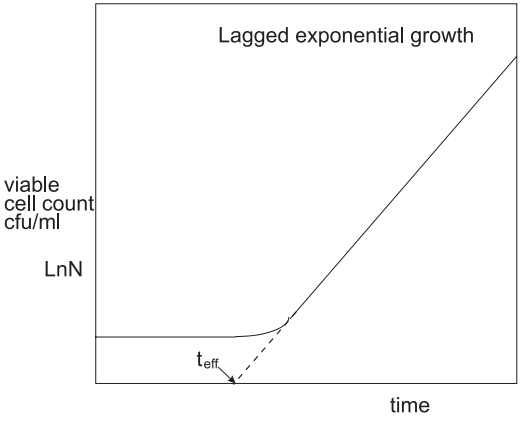


Fig. 1.3 A schematic of the lagged exponential predicted by equation (1.12) showing the effective population lag time, t_{eff}.

where

$$t_{\text{eff}} = (1/A)\ln[1 + A < \tau >] \tag{1.14}$$

Here t_{eff} is the effective lag time measured for the whole population; whereas $< \tau >$ is the average lag time per cell. Clearly these two quantities are not, in general, the same. In fact equation (1.14) permits us to derive a novel relationship between the effective lag time and the mean doubling time:

$$t_{\text{eff}}/t_{\text{d}} = \ln[1 + A < \tau >]/\ln 2 \tag{1.15}$$

The same probabilistic formulation permits us to derive an expression for the colony size distribution in the food matrix. The derivation is somewhat lengthy so we merely quote the result. If $p\{N_i|t\}$ is the probablity that there is a colony of N_i cells in the ith colony at time t, then,

$$p\{N_i|t\} \quad = e^{[-t/<\tau>]} N_i^{[1/A<\tau>-1]} \quad N_i > 1 \tag{1.16}$$

(in the growth phase)

$$= e^{[-t/<\tau>]} \qquad N_i = 1 \tag{1.17}$$

(in the lag phase)

This distribution provides an important experimental test of the theory, because the colony size distribution is a readily accessible quantity. These tests are now under way using a laser-gel cassette scanner.[2]

By defining appropriate sub-populations of cells the same probabilistic approach can be used to derive the time evolution for injury, sublethal injury, resuscitation and preadaption:

$$N(t) = n_{\text{lag}}(t) + n_{\text{exp}}(t) + n_{\text{inj}}(t) + n_{\text{res}}(t) + \ldots \tag{1.18}$$

where there is a probability for transition between the physiological states. For example, the simple case of injury and death is treated by postulating that the probability of a given cell in the inoculum suffering an injury after time t_{inj} is a Poisson random process:

$$p(t_{\text{inj}}) = (1/ < t_{\text{inj}} >)e^{[-t_{\text{inj}}/<t_{\text{inj}}>]} \tag{1.19}$$

where $< t_{\text{inj}} >$ is the mean injury time. It follows that

$$d\ln N/dt = -1/ < t_{\text{inj}} >= -D/\ln 10 \tag{1.20}$$

where D is the conventional D number for death.

The effect of the changing environment surrounding the cells can be included by defining the dependence of the transition probabilities for lag, growth, injury and resuscitation on those variables, such that,

$$A = f_1\{[\text{O}], \text{pH}, a_{\text{w}}, T, [C], [N] \ldots\} \tag{1.21}$$

$$< \tau > = f_2\{[\text{O}], \text{pH}, a_{\text{w}}, T, [C], [N] \ldots\} \tag{1.22}$$

$$< \tau_{\text{inj}} > = f_3\{[\text{O}], \text{pH}, a_w, T, [C], [N] \ldots\} \tag{1.23}$$

$$< \tau_{\text{res}} > = f_4\{[\text{O}], \text{pH}, a_w, T, [C], [N] \ldots\} \tag{1.24}$$

One of the outstanding challenges is to discover the appropriate functional forms of $f_1, f_2 \ldots f_4$. Several functional forms have been proposed in Hills and Wright[3] but all involve a number of unknown parameters that can only be determined by fitting experimental data. In Barker and Grimson[4] the changes in $N(t)$, pH, glucose and oxygen concentrations during the batch fermentation of *Salmonella typhimurium* were used to determine these parameters, but the important question as to the universality of these parameter values remains to be explored.

This example highlights again one of the essential criteria for the value of a deductive, mechanistic model. The functional forms, $f_i\{\ldots\}$ and their associated parameter values should, ideally, be universally valid for all types of microorganisms, physiological states and medium compositions. The less this is true, the weaker the model. In the opposite extreme, the model will be essentially of zero predictive value if experiments must first be done for every situation in order to determine parameter values or functional forms in the model, which is always the case for completely empirical models.

The growth model outlined above makes no explicit mention of the physiological state of the cells. Yet this aspect is known to have a powerful influence on survival and growth. For example, the lag time of many microorganisms when subjected to a stress situation, such as low water activity, can be shortened by 'pre-stress adaption', whereby the microorganism is first cultured in a growth medium of reduced water activity. One possible way of introducing the effects of physiological states would be to introduce new sub-populations of cells differing in their physiological state, such as unadapted and pre-adapted cell populations with a transition probability for passing from the unadapted to preadapted state. This approach also requires further experimental fitting to determine parameters, which weakens its fundamental deductive nature.

1.3.1 The effects of microgradients in food systems

One major advantage of deductive models is the ease with which they can be generalized to incorporate new phenomena. For example, it is known that microorganisms consume nutrients from the surrounding food matrix and also liberate metabolic end-products into the surrounding matrix. The implication is that the actual values of environmental variables, such as pH, and the concentration of carbon and nitrogen sources as well as the oxygen concentration experienced by the growing microorganisms will vary and this will feed back onto the growth behaviour. Empirical models cannot describe this phenomenon, but it is easily incorporated into the previous mechanistic model by converting a set of ordinary differential equations into partials, including diffusive transport terms. By way of illustration the set of coupled partial differential equations used in Hills *et al.*[2] and Hills and Wright[3] are:

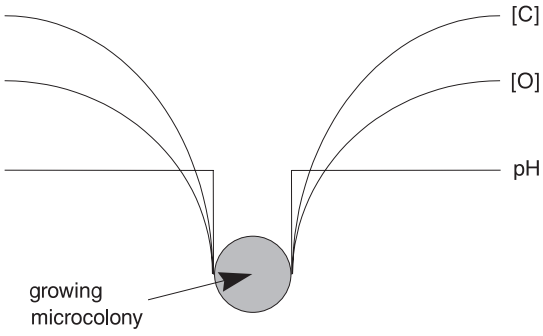


Fig. 1.4 A schematic showing microgradients surrounding a spherical colony.

$$\begin{aligned}
\partial m(r,t)/\partial t &= Am \\
\partial n(r,t)/\partial t &= k_n(m-n) \\
\partial o(r,t)/\partial t &= -y_0 Am + \mathrm{Div}[D_0\,\mathrm{Grad}o] \\
\partial c(r,t)/\partial t &= -y_c Am + \mathrm{Div}[D_c\,\mathrm{Grad}c] \\
\partial h(r,t)/\partial t &= -k_h hm + \mathrm{Div}[D_h\,\mathrm{Grad}h]
\end{aligned} \tag{1.25}$$

where the variables are dimensionless quantities such that $m(r,t)$ is a biomass, $n(r,t)$ a viable cell number, $o(r,t)$ the oxygen concentration, $c(r,t)$ the concentration of carbons sources, $h(r,t)$ is the concentration of hydrogen ions at a certain location r and at time t, so is, of course, related to pH as $10^{(7-\mathrm{pH})}$ (see also Chapter 3). Details can be found in Hills and Wright.[3]

Various methods are available for solving these equations numerically. Apart from the obvious finite element method, the equivalent cellular automaton can be designed[4, 12] and this has the advantage of flexibility in defining irregular spatial boundary conditions. The results show that microgradients are rapidly established around the growing colonies, as illustrated schematically in Fig. 1.4. The existence of microgradients means that the actual values of environmental variables, such as pH, experienced by the exponential phase cells may not be the pH of the bulk medium. The model represented by equations (1.25) also predicts that the microcolony itself has an internal structure, such that cells in the centre are more stressed than those in the surface layer. This, in turn, implies that even when the total plate count of the viable cell population in a food matrix, $N(t)$ is in the exponential growth phase, there may actually be sub-populations of viable and non-viable as well as stationary phase and exponential phase cells present within the colonies. Very similar ideas can be used to model the structure and development of biofilms which can be a major source of contamination in production plants.

1.3.2 The effect of food microstructure and the microscopic water distribution

So far we have considered statistical variations arising from sub-populations of cells. However, the spatial heterogeneity of the food matrix can also give rise to

statistical variations in the local environment and therefore in the survival and growth behaviour of the cells distributed within it. Such microstructural effects are of interest because recent data shows that varying the food microstructure can have dramatic effects on the survival of microorganisms. For example, Fig. 1.5 compares the survival of an innoculum of 10^7 cells/ml of *E.coli* when added to a synthetic growth medium (MOPS) poised at a water activity of 0.94, which is sufficiently low to stress the organism osmotically and prevent growth. In the upper curve in Fig. 1.5 the water activity was poised at 0.94 by addition of sucrose, which, of course dissolved and formed a spatially homogeneous solution. In the lower curves the water activity was lowered to 0.94 by addition of Sephadex microspheres and inert silica powder respectively. These formed a structured inert matrix having heterogeneity on the distance scale of tens to hundreds of microns.[5] The 'microstructural stress' effect for Sephadex results in complete death after about 100 hours; in contrast to the unstructured medium where substantial numbers of cells survive. The results suggest that a microstructured food matrix, such as a partially saturated packed bed of Sephadex microspheres, has a distribution of microscopic environments that differ in their local water content (and hence water activity),[5] nutrient availability and oxygen concentration so that if, by chance, a bacterial cell happens to be located in an unfavourable environmental 'niche', it will suffer stress and injury

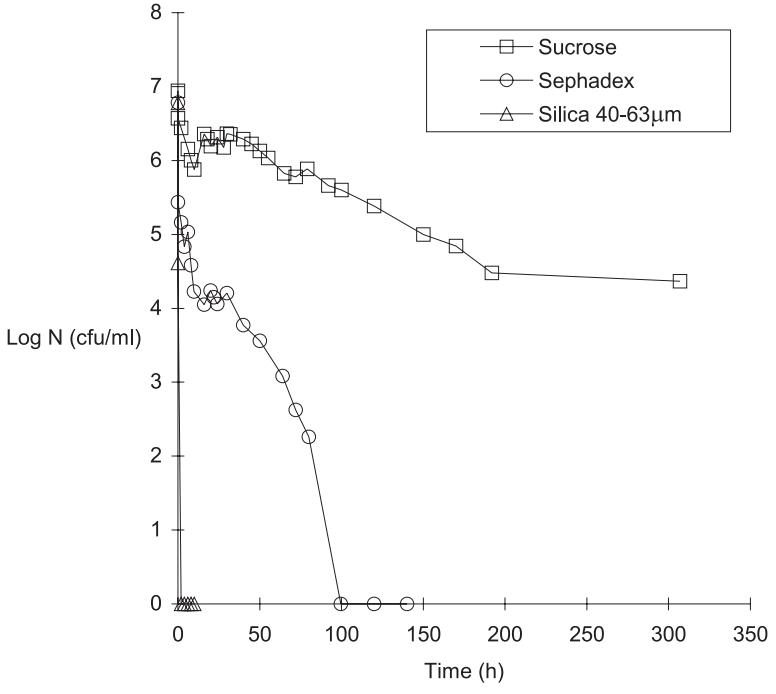


Fig. 1.5 The microstructural stress effect. Challenge test of E.coli K-12 (frag 1) in sucrose solution, Sephadex G25-50 and silica all poised at a water activity of 0.94, 20ºC.

and possible loss of viability. The implication for food safety and shelf-life prediction is obvious, yet, at the present time, no theoretical models are capable of predicting this 'microstructural stress effect'. The cellular automaton approach[4] offers considerable promise here, since it permits microstructure to be incorporated directly into the automaton. In effect, this places the spatial dependence on the environmental variables in equations (1.21) to (1.24).

It is surprising, at least to this writer, that so many fundamental questions concerning the microbial safety of foods still remain to be answered. These issues will clearly need to be addressed before deductive mechanistic modelling can be used in a genuinely predictive way to predict microbial shelf-lives.

We now turn to another, equally complicated, issue affecting food quality, namely food flavour. This is also of vital importance because there is little point in optimizing the processing stage and maximizing microbial stability if the end-product tastes and smells awful!

1.4 Deductive modelling of flavour release from foods

The past few years have seen rapid progress in the development of deductive mechanistic models of flavour release from foods in the mouth based on the fundamental physics of interfacial mass transfer.[6–9] The ultimate goal of these efforts is to be able to predict mathematically the effect of varying food composition, food structure and mastication behaviour on the perceived time-intensity flavour release profile. If successful, it would then be possible to use computer simulations to formulate foods for a desired flavour profile, taking account of individual or group differences in mastication behaviour.

There are several steps in the development of this simulation package:

1. Understanding the physical mechanism of aroma release from a food into the saliva and mouth headspace
2. Modelling the effects of in-the-mouth mastication. This, in turn, involves modelling several processes, namely,
   - Saliva flow
   - Chewing, which involves the selection of food particles and their fragmentation behaviour
   - Swallowing, which removes saliva and some food particles
   - Partitioning of aroma between saliva, mucus and air
   - Transport of aroma to the olefactory epithelium
3. Modelling the relationship between the actual aroma concentration-time profile and the perceived, physiological response of the consumer.

In principle each step needs to be applied to every conceivable food, which is, of course, an enormous and daunting task. However, the problem can be made more tractable by developing fundamental models for various general release mechanisms and then adapting and calibrating them to particular foods. These general models include flavour release from solid foods by

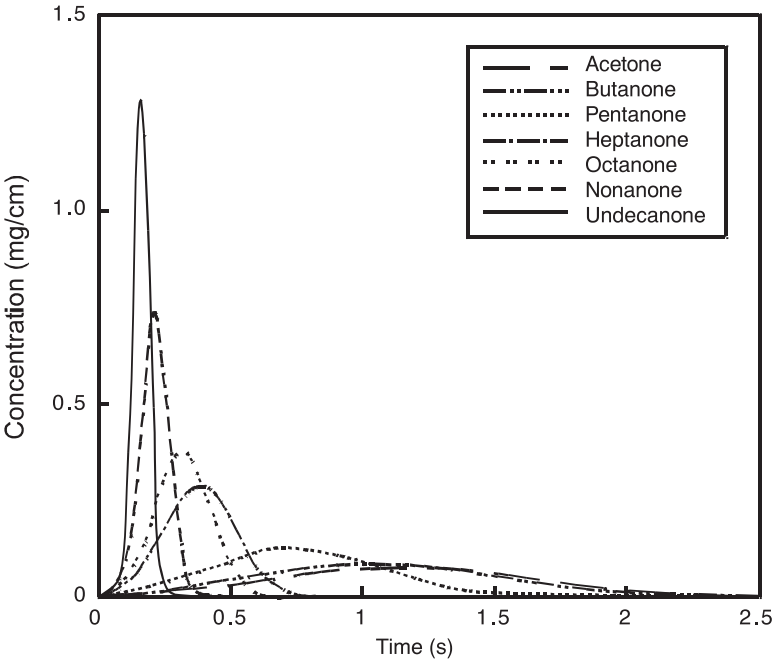


Fig. 1.6 The theoretical flavour release profiles for several flavour compounds released from aqueous solution in the mouth.

- Dissolution (e.g. of a boiled sweet)
- Diffusion-melting (e.g. of a jellied sweet comprising a gelatin-pectin matrix containing sucrose and flavour)
- Fragmentation (e.g. of a crumbly biscuit)
- Bolus formation (e.g. from chewing gum or bread) and
- flavour release from liquids (soft drinks and liquid emulsions).

The effects of mastication are being simulated using a computer model[9] which incorporates effects listed above. This model can be adapted to each of the food release mechanisms listed above. Figure 1.6 shows a typical output from the simulation, and is reproduced from Harrison *et al.*[9]

The third step, the relationship between the perceived response and the actual physical concentration-time profile, is less well developed and much research remains to be done in understanding and modelling this relationship. This takes us from modelling food behaviour in the mouth (physics) to the physiological response on the tongue and nose (physiology) to modelling the consumer behaviour and appreciation while eating the food (psychology and sensory science). Recent advances in functional magnetic resonance imaging (MRI) of the physiological responses of the brain to flavour and aroma perception may help in this endeavour.

1.5 Future trends

It is clear from the previous examples that although considerable progress is being made in the rules that govern the building of fundamental models and in using the deductive mechanistic approach to model various aspects of processing, spoilage and flavour release, these models should not be treated in isolation. Extending the microbial shelf-life of a food by, for example, increasing its porosity and increasing the concentration of preservatives is useless if the food then has an unacceptable flavour profile. The development of integrated computer simulations capable of optimizing several different aspects of a food's production and quality must therefore be considered an outstanding future challenge. Invariably, the process of developing these models will highlight gaps in our knowledge about the underlying physics and chemistry controlling food production and quality. As an example, consider the introduction of new crops (e.g. by genetic modification). Quite apart from the safety issues, it is not yet possible to predict the effect of modifying a raw material on its subsequent processing response because we lack a quantitative understanding of the fundamental relationships between the raw material's biomolecular structure, its microstructure and its macroscopic behaviour which characterizes the food quality factors such as texture and rheology. Understanding how macroscopic processing and storage responses are controlled by microscopic and molecular factors therefore remains another outstanding problem in food science. Modifying raw ingredients such as starch or flour by GM, chemical means or otherwise could also affect both the microbial stability by changing 'water availability' as well as the flavour release profile by altering starch-protein-flavour binding coefficients and flavour transport coefficients which once again highlights the huge amount of basic research required when developing realistic deductive models.

Although we have focused the discussion on the quality of processed foods, very similar considerations apply when developing models for optimizing the production of raw foods in the field. Here environmental factors such as irrigation patterns, harvesting times, soil types and differences in climate affect the quality of the raw food materials. An example would be cassava starch where the processing response of the cassava starch and the rheological properties of the cassava starch pastes and gels depends critically on the granule structure which, in turn, is affected by rainfall patterns and harvesting times.[10] Understanding and optimizing these relationships 'from field to fork' will present many exciting future research possibilities.

1.6 References

1. HILLS B P, *Magnetic Resonance Imaging in Food Science*, John Wiley & Sons, New York, 1998.

2. HILLS B P, WRIGHT K M, BROCKLEHURST T F, WILSON P D G, MACKIE A R, ROBINS M M, 'A new model for bacterial growth in heterogeneous systems', Volume 2, 727–741 in *Proceedings of the 1993 Food Preservation 2000 Conference, 19–21 October, Natick, Mass. USA*, Sci. and Tech. Corp., Virginia, USA, 1996.
3. HILLS B P, WRIGHT K M, A new model for bacterial growth in heterogeneous systems. *J. Theoretical Biology*, 168 (1994) 31–41.
4. BARKER G C, GRIMSON M J, A cellular automaton model of microbial growth. *Binary: Computing in Biology*, 51 (1993) 132–7.
5. HILLS B P, MANNING C E, NMR oxygen-17 studies of water dynamics in heterogeneous gel and particulate systems. *J Molecular Liquids*, 751 (1998) 61–76.
6. HILLS B P, HARRISON M, Two-film theory of flavour release from solids. *Int J Food Sci and Technol*, 30 (1995) 425–36.
7. HARRISON M, HILLS B P, A mathematical model to describe flavor release from gelatine gels. *Int J Food Sci Technol*, 31 (1996) 167–76.
8. HARRISON M, HILLS B P, BAKKER J, CLOTHIER T, A mathematical model to describe flavor release from liquid emulsions. *J Food Science*, 61 (1997) 653–64.
9. HARRISON M, CAMPBELL S, HILLS B P, Computer simulation of flavor release from solid foods in the mouth. *J Agric and Food Chem*, 46 (1998) 2736–43.
10. SRIROTH K, SANTISOPASRI V, PETCHALANUWAT C, KUROTJANAWONG K, PIYACHOMKWAN K, OATES C G, Cassava starch granule structure-function properties: influence of time and conditions at harvest on four cultivars of cassava starch. *Carbohydrate Polymers*, 38 (1999) 161–70.
11. FOOD MICROMODEL, 1992: *A computer-based information resource for predicting microbial growth*, Campden Food and Drink Research Association, Chipping Campden, Gloucestershire, UK.
12. VAN DER SMAN R G M, 1999. Lattice Boltzmann schemes for convection-diffusion phenomena; application to packages of agricultural products. Thesis Wageningen. ISBN 90-5808-048-X.

2

Problem decomposition

M. Sloof, Everest B. V., 's-Hertogenbosch

2.1 Introduction

Almost all food processes are too complex to be modelled in one step. Therefore, a modeller will analyse the phenomenon under study to divide the complex problem into smaller problems that are easier to solve or for which solutions are already available. Just dividing the phenomenon into smaller problems does not take advantage of the full potential that can be reached with a conscious and complete decomposition. A conscious and thorough decomposition of the complex phenomenon may lead to a better understanding of the problem at hand, enabling the modeller to make sound decisions about where to make the simplifications needed to arrive at a practical model for the phenomenon under study.

In information technology, this technique has already proved to be invaluable for modelling physical systems. Because such systems are composed of well-defined components, decomposition of these systems is generally easy and straightforward. However, when applying the same technique to modelling complex phenomena in food products, two main difficulties arise. First, the fundamental subprocesses are not completely understood or described. Second, the interactions between these processes are complex and even less understood.

In this chapter two methods will be discussed that are widely used in information technology for decomposing a complex system into manageable units. The first method is task decomposition and models the system as a collection of interacting tasks. The second method is called object-oriented analysis and design. This method models the system as a collection of interacting objects, where each object represents a concept or entity in the real-world system.

It is shown how and where these decomposition methods can be used in modelling food processes. A number of modelling cases will be presented to illustrate the rules and the possible pitfalls of the decomposition methods. Knowledge of these rules and pitfalls will enable the modeller to develop sound and complete decompositions of the problems under study. It is only with such a complete decomposition of the problem under study that simplifications, necessary to formulate a model, are made correctly and understood thoroughly.

2.2 Decomposition in information technology

In information technology two techniques are widely used to decompose a complex system into manageable subproblems. The first technique is called task decomposition and focuses on the tasks performed by the system. The second technique is called object-oriented analysis and design. With this technique the system is decomposed into a number of objects, where each object is responsible for one or more tasks performed by the system.

The aim of both techniques is to divide a complex problem into a number of so-called building blocks. Task decomposition uses tasks and subtasks as the building blocks to construct a model for the system. In object-oriented analysis and design the building blocks are the objects. Each building block must describe a well-defined part of the complex system that can be studied and understood independent of other building blocks. In this way parts of the complete behaviour can be studied in isolation and may be redesigned without affecting other parts of the model.

When several models have been developed for different phenomena in a particular problem domain, it may appear that some building blocks have been reused in some of these models. These building blocks have been abstracted from the particular problem for which they were originally developed. These building blocks can be used as a starting point for a more extensive analysis of the problem domain to identify a collection of generic building blocks that can be used to build models for phenomena in that domain. A generic building block is a model of a well-defined piece of behaviour, independent of the problem or system for which this building block was originally developed. When such a collection of generic building blocks is available, decomposition of any problem in that domain should aim at selecting the generic building blocks that are relevant for the problem at hand rather than inventing problem-specific solutions.

2.2.1 Task decomposition

The task decomposition method describes the system under study as a collection of interacting tasks or processes. Each task represents a piece of behaviour performed by the system. A task has an effect on one or more attributes of the system, and may itself be influenced by other attributes in the system or by external attributes.

If a task is too complex to be described immediately, it can be decomposed into a sequence of two or more subtasks that are easier to describe. The first subtask in the sequence will have the same input variables as the complex task and the last task will have the same output variables as the complex task. The subtasks in the sequence are connected through intermediate variables. The sequence of subtasks must exert at least the same behaviour as the complex task.

The intermediate variables introduced in the decomposition of the complex task were hidden in the original, more abstract, task. Hence, by decomposing a task into subtasks some variables that were implicit in the abstract task are made explicit. Hence, task decomposition will enhance the understanding of the task and, more importantly, will provide handles to include additional interactions between tasks.

Task decomposition produces a task decomposition tree of the system under study. At the top of this tree is the complex task to be modelled, below this are the subtasks into which the complex task is decomposed, and at the bottom of the tree are the subtasks that cannot, or need not, be further decomposed into subtasks. When all tasks in a phenomenon under study are decomposed into subtasks until the tasks cannot be decomposed further, the resulting task decomposition tree can be seen as a comprehensive model of that phenomenon. When developing a model for the phenomenon the modeller will select subtasks from the decomposition tree at the appropriate level of detail. Subtasks may be selected on different levels of detail. In this way only that part of the comprehensive model is reused that is relevant for the phenomenon under study. Processes that are irrelevant and models for processes that are described in too much detail are left out.

2.2.2 Object-oriented analysis and design

A second important method to decompose a complex system is object-oriented analysis and design. An object-oriented analysis and design results in a description of the system as a collection of interacting objects, where each object has attributes that together define the state of the object, and methods that implement the behaviour of the objects.

Object-oriented analysis starts with an investigation into the main processes in the system under study. For each process in the system under study it has to be decided which object will be responsible for this process. In the design phase a method will be developed in the responsible object to implement this process. If the object needs information from other objects to perform the process, methods must be designed in the other objects to provide this information. In this way a pattern of collaboration between the objects is designed and the complex behaviour of the system under study is decomposed into a number of processes distributed over the objects. The collection of objects with attributes and methods and the collaboration patterns between the objects form the conceptual model for the system under study. The conceptual model is the input for the construction of an object-oriented simulation model.

An object-oriented simulation model consists of classes and instances of classes. A class is a template for one or more objects in the conceptual model with the same behaviour. Each instance of a class represents one object in the conceptual model. The classes are usually organized in a class hierarchy. Classes higher in the hierarchy are more generic than classes lower in the hierarchy. A class inherits all members (attributes and methods) from its ancestors in the hierarchy and adds new members to represent the specialized behaviour. New members hide members with the same name from the ancestor classes.

For example, a class hierarchy for modelling the behaviour of apples could consist of a generic class containing generic models for the processes that occur in all apple cultivars. To model the behaviour of a specific apple cultivar, the generic class can be specialized and one or more methods can be adapted to describe the behaviour of this cultivar that differs from the generic behaviour as described in the generic class.

A class hierarchy for a particular domain can be seen as a template model for the construction of specific models for systems or phenomena in that domain. The generic classes in the class hierarchy contain methods for the generic processes that may be relevant in the phenomenon under study. In the analysis and design of a conceptual model for the phenomenon under study the modeller should use as many generic classes as possible. In cases where the observed behaviour cannot be described with one of the generic classes, the modeller must design a new class that is a specialization of one of the generic classes and add new members to represent the observed behaviour. In this way the class hierarchy can be used as a collection of reusable models for the processes in the domain of interest, thereby reducing development time and effort and reducing the risk of errors in the models.

More extensive discussions of object-oriented analysis and design can be found in the literature, together with descriptions of the Unified Modelling Language (UML) that has emerged as the standard notation for (object-oriented) modelling, and descriptions of various design patterns that are ‘best practice’ solutions for common problems in object-oriented design.[1,2,3]

2.3 Modelling food processes

Modelling starts with a specification of the phenomenon under study. The specification generally implies a limitation of the area of interest. When modelling food processes, this specification may include the product in which the phenomenon is studied, the conditions the product is likely to be subjected to, the aspects of the product behaviour that have to be included and, most difficult, those that may be excluded.

The following sections describe a comprehensive view on the modelling of behaviour of food products during post-harvest distribution. During distribution the product may be traded several times. Each time the product is traded the quality of the product will be evaluated. Furthermore, a product may be

distributed together with other products that may influence the behaviour of the product under study. Therefore, modelling the behaviour of a food product may involve modelling quality change, an object-oriented decomposition to identify products with similar behaviour, and modelling the physiological behaviour of these products.

2.3.1 Modelling quality change

To describe the changes in the quality of the distributed product, models are developed for the processes that may affect the product attributes that are used to assess the quality. The behaviour of a quality attribute not only depends on the physiological processes occurring in the food product, but also on the user assessing the quality attribute, and may even depend on the market situation (see Chapter 17).[4,5]

For example, to assess the quality of a tomato a consumer may evaluate the firmness, the colour, the price and the origin of the tomato. Furthermore, a consumer may compare the product with other tomatoes.

Firmness and colour are quality attributes that depend on the concentrations of various biochemical substances. When assessing the firmness the consumer will also take into account the intended use of the tomato. A firm tomato may have a high quality for use in salads, whereas the same tomato may have a low quality for use in tomato soup. Hence, whereas the firmness depends solely on the concentrations of various biochemical substances, the value assigned to the quality attribute firmness depends on the intended use. The biochemical substances that are responsible for the observed firmness are called intrinsic product properties. When these properties change, the state of the product will change also, leading to a different quality. The quality of a product that is determined by taking into account only quality attributes that depend solely on intrinsic product properties, is called the assigned quality.

Price and origin of the tomato are called extrinsic product properties. Extrinsic properties can change without affecting the state of the product and are therefore not important for the assigned quality of the product.

Furthermore, a user may compare the quality of the product with the quality of competing products. The combination of assigned quality, the extrinsic product properties and the market situation yields the acceptability of a product: an assessment of the product in relation to its price and to other products. Independent of the assigned quality of the product, the acceptability will decrease or increase if other products are assigned a better or worse quality, respectively. The process of determining the assigned quality and the acceptability of a product is depicted in Fig. 2.1. The grey bars in this figure indicate the active role the user plays in this process. By defining criteria for the intrinsic product properties the user determines what quality is assigned to the product. The second grey bar represents socio-psychological factors, such as quality-awareness, status-awareness and previous experiences, that may influence the user when deciding whether or not to buy the product.

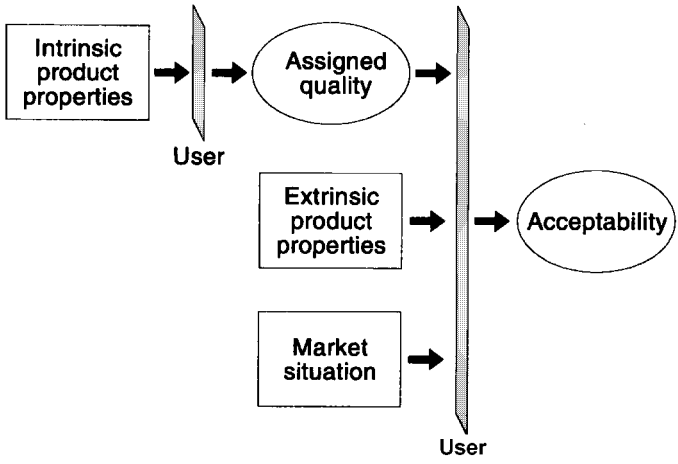


Fig. 2.1 Assigned quality of a food product depends solely on the evaluation of intrinsic product properties by the user. The acceptability of a food product depends on the assigned quality and on extrinsic product properties and the market situation.

From this discussion it becomes evident that it is important to separate clearly these effects in a model for the food process. A model describing the quality change of agricultural products should consist of three submodels:

1. an environment model describing the changes in the immediate environment of the product,
2. a dynamic product model describing the reaction of intrinsic properties of the product on changes in the environment, and
3. a quality assignment model describing the relation between the quality of the product and the values of the product properties.[4,5]

An example of an environment model is a model for a Modified Air package (see Chapter 14), describing the relation between the conditions outside the package, the package material, and the resulting conditions inside the package.

The advantages of separating the environment and quality assignment from the product behaviour can be seen when considering a distribution chain. A distribution chain contains a number of 'quality assessment points', at which different definitions of 'good quality' are applied. For example, early in the distribution chain tomatoes must be reddish, hence unripe, whereas in a supermarket tomatoes must be red and firm. The quality definitions differ, but the physiological processes causing the quality change are the same. Following the generic structure for a quality change model, the model for the distribution chain would consist of two quality assignment models connected to one dynamic product model. However, if the quality attributes colour and firmness were modelled as influenced directly by the processes in the tomato, the models for the physiological processes would differ because of the differing quality definitions, even though the underlying physiological processes are the same.

Models in which quality assignment and product behaviour are not clearly separated can have good prediction performance, as is shown by the FLORES model that is widely used to predict the keeping quality of cut flowers in distribution chains.[6] The FLORES model is a generic model describing the loss in the keeping quality of FLOwers in RESponse to various and varying external conditions. The model is applicable to more than twenty different flower cultivars. The keeping quality is the number of days during which the flowers remain acceptable when kept at an optimal condition. In FLORES the keeping quality is called the vase-life of the cut flowers. Loss of vase-life occurs with the passing of time and if the environmental conditions are in some way suboptimal. The FLORES model consists of seven submodels. Each submodel describes the loss in vase-life due to one effect. The total loss in vase-life is the sum of the losses calculated by these submodels. The first submodel describes the basic decrease of vase-life over a period of time and a certain temperature path. This effect occurs continuously, also at optimal conditions. The other submodels describe external effects that further decrease vase-life: dry storage, infection by fungi, bacteria growth, suboptimal temperatures, exposure to ethylene, and absence of flower preservative in the vase water. These effects occur if the flowers are in some way subject to suboptimal conditions.

The FLORES model is a generic model based on a decomposition of the complex phenomenon of quality change into processes. However, the fact that the functions in FLORES describe direct relationships between the environmental conditions and the loss of the quality attribute vase-life, and the fact that these functions do not explain the effects in terms of underlying physiological processes, makes it impossible to reuse the functions in FLORES for other products. For example, not only flowers but also a lot of fruits are sensitive to ethylene exposure. However, as the ethylene function in FLORES is a direct relation between ethylene concentration and loss of vase-life, the function cannot be used in other models to calculate the effect of ethylene exposure in other products. As a consequence, the models describe the effects of environmental conditions on the product using an implicit definition of ‘good quality’. This combination of physiology and quality assignment has two consequences. Firstly, a model for one product cannot be reused for another product with a similar physiology, because the other product has a different definition of quality. Secondly, a model can be used only to simulate quality change according to one definition of quality. If different definitions of quality can be applied to the same product, then different models have to be developed for each definition of quality.

2.3.2 Object-oriented analysis of the phenomenon

When the quality assignment is clearly separated from the product behaviour, the modeller can focus on modelling the product behaviour. Before focusing on the processes that lead to the observed product behaviour, it has to be decided whether an object-oriented analysis may be necessary. Applying an object-

oriented analysis will describe the phenomenon as a collection of interacting objects. The behaviour of each individual object may then be described as a collection of interacting processes by decomposing the behaviour of each object into processes.

An object-oriented analysis may be applied to the following cases. These cases are described in more detail in the sections below.

1. When studying the effects of a distribution chain on the quality of the distributed product. Each activity in the distribution chain can be modelled as an object that changes the environment of the product, thereby affecting the physiological processes in the product.
2. When the phenomenon under study involves different products and the interactions between the products are important. Each product can then be modelled as an object with its own physiological behaviour. The physiological processes in one product may cause changes in the immediate environment of that product. These changes may affect the physiological processes in the adjacent products. With an object-oriented design of the model these changes can be described explicitly. The object-oriented model may contain one object representing the environment and objects representing the individual products. Through interactions between the latter objects and the environment object, the environmental conditions are changed thereby affecting the physiological processes occurring in the products.
3. When the phenomenon involves a batch of one type of product, with different external conditions at different locations in the batch, or with differing initial quality. The complete batch is modelled as a collection of objects. Each object represents an area in the batch where the products are subjected to similar external conditions.

Modelling a distribution chain

A distribution chain is a sequence of activities performed to deliver a product at a destination with the highest possible quality. Usually, the 'behaviour' of an activity leads to a change in the environment of the distributed product. In some cases the behaviour of an activity may have an immediate impact on the product. An example of the latter is recutting of cut flower stems.

An object-oriented model of a distribution chain may contain objects representing the individual handling activities and an object representing the distributed product. Activities that change the environment of the distributed product are modelled by interactions between the activity objects and an object representing the environment. For example, a temperature-controlled transport can be modelled as follows: the object representing the transport sends a message to the environment object to set a new temperature. To determine the physiological behaviour during the transport the product object interacts with the environment object to retrieve the temperature during the transport.

In cases where the activity leads to a direct change in the product, there will be a direct interaction between the activity object and the product object.

Recutting a flower stem would be modelled as follows: the activity object sends a message to the product object to reduce the bacteria concentration. Upon receipt of this message the product object will use the new bacteria concentration in the simulation of the physiological behaviour.

The advantage of using an object-oriented model for a distribution chain is the clarity of the model. Normally the effects of a distribution chain are modelled as a sequence of environmental conditions at different time points during the distribution. By using an object-oriented model the activities causing these changes in the environment are made explicit. With an object-oriented model it becomes easier to add or remove activities from the distribution chain. Finally, the objects representing the activities may be reused to model distribution chains for other products.

Modelling interactions between products

When a product is stored in a container with other products, not only the interaction between the product and its environment is important, but also the interactions between the products become important. In such a mixed load situation, the different products have to be modelled separately and the environmental conditions between the products have to be modelled explicitly.

An object-oriented model describing the behaviour of products in a mixed load may contain objects representing the individual products in the container and an object representing the environmental conditions inside the container. Each object representing a product implements a dynamic product model describing the physiological processes occurring in the product. These processes are affected by the conditions in the immediate environment of the product. The processes may themselves affect the immediate environment. The single object representing the container environment monitors all changes in the environmental conditions.

The advantage of using an object-oriented model for a mixed load is that products can be added or removed easily without the need to adapt the model for the container environment. The products interact with the environment rather than directly with the adjacent products. Not only is this a more accurate description of the real-world situation, it also allows reuse of product models that originally were not designed to be used in mixed load situations.

Modelling a batch of products

When the phenomenon involves a batch of one type of product, the modeller has to decide whether all products in the batch have the same initial quality and whether the external conditions are the same at all locations in the batch. If this assumption can be made, then the batch can be modelled as a single product with a single environment and a further object-oriented analysis is not necessary. If this assumption cannot be made, then the batch has to be divided into smaller units such that the products in each unit have the same initial quality and are subject to the same external conditions.

An object-oriented model describing the behaviour of products in the batch may contain an object for each area in the batch in which the products have a

similar initial quality and are subject to the same environmental conditions. Each object describes the behaviour of part of the products in the batch.

The advantage of using an object-oriented model for describing the behaviour of a batch of products is that the physiological processes occurring in the products can be modelled once for all areas in which the batch is divided. The division in areas of the complete batch can be manipulated easily by creating or deleting instances of the class that implement the model of the products in the batch.

2.3.3 Process decomposition

In the previous steps the phenomenon under study has been analysed to separate quality assignment from the product behaviour and to identify relevant objects in the phenomenon. The last step in the analysis is to describe the product behaviour as a collection of interacting processes, such that their combined action describes the observed phenomenon and that each of the subprocesses can be fully understood in their own description. The type of subprocesses is largely defined by the fundamental laws and the generally accepted rules in a particular discipline.

Each process influences one or more quantities of the product, and may itself be influenced by other quantities. Hence, the interactions between the processes occur at the quantities. To be able to combine effects of several processes on one quantity, to add processes to the model, or to remove processes from the model, the processes must be described in terms of changes imposed on the affected quantities. Therefore, the processes have to be formulated in terms of differential equations rather than in terms of algebraic equations. An algebraic equation calculates a new value of the output quantity, which makes it impossible to include other effects on that quantity other than by reformulating the model equations.

Process decomposition may start with identifying high-level processes that may occur in the phenomenon. One way to find these processes is by starting from the intrinsic product properties that correspond with the quality attributes in the phenomenon under study. By using knowledge about the food product the modeller derives high-level processes that influence these product properties. Only those high-level processes are included that are affected by one or more of the external influences on the food product identified in the phenomenon under study. The result is a first decomposition of the phenomenon into abstract or high-level processes, describing how the attributes of interest are affected by the external influences on the food product. This process decomposition will contain several high-level variables through which the processes interact.

The next step is to develop a model for each abstract process, either by a further refinement into subprocesses, or by selecting an existing model for the process. For example, a model for a chemical process can be composed into at least three submodels: one model describing the changes in the concentration of the consumed reactant, one model describing the changes in the concentration of

the produced reactant, and one model describing the influence of temperature on the chemical process. An example of the latter model is Arrhenius' law.

As seen above in the example of the FLORES model, process decomposition must be aimed at finding the generic physiological processes in the product. This is illustrated with a model describing the occurrence of chilling injury in cucumber fruits and bell peppers.[7]

Chilling injury is a general term for visible forms of damage that may occur when products are stored at too low temperatures. The injury normally appears after a chilling period, when the product may already be stored at optimal conditions. This deferred appearance makes chilling injury difficult to comprehend and to model. However, by making a number of assumptions and by using generic processes, it proved possible to develop a quantitative simulation model for the complex phenomenon of chilling injury. This model also correctly explains chilling injury phenomena that were not accounted for in the development of the simulation model, which proves the validity of the approach. The complex phenomenon was decomposed into the following processes:

1. Chilling injury is the visible effect of too many free radicals that are generated by reactions in the living cells of the product. The occurrence of chilling injury was modelled as a chemical reaction that consumes the generated radicals. Hence, a higher concentration of free radical increases the visible chilling injury.
2. The generation of radicals was modelled as an autocatalytic process with respect to the amount of free radicals.
3. At normal conditions no chilling injury is observed, so that the free radicals must be removed or inactivated in some way. This radical scavenging process was assumed to be an enzymatic process, affected by the amount of free radicals and by the enzyme activity.
4. The enzyme in the radical scavenging process was assumed to denaturate irreversibly at low temperatures. This accounts for the fact that chilling injury only occurs after a period of too low temperatures.

Modelling interactions between processes

Decomposing a process into more detailed subprocesses will introduce additional variables in the model. For each variable it has to be decided which processes have an effect on the variable, and whether these processes should be included in the model. For example, including an enzymatic process into the model will also introduce a quantity representing the enzyme. Normally, the enzymatic process itself does not affect the enzyme concentration. However, other processes may influence the enzyme concentration. It has to be decided whether such processes are relevant and have to be included in the model. Including processes that influence the enzyme concentration increases the applicability of the model, but also increases the complexity of the model.

When reusing a model for a process, it must be investigated whether the variables in the model are combinations of other variables. A combined variable

hides an interaction between variables at a more detailed level that was assumed to be constant in the situation for which the model was originally developed. Such an interaction may, however, not be constant in the current situation for which the model is reused. Examples of such variables are $V_{\max}$ and K_m in the Michaelis Menten equation (2.1):

$$\frac{dS}{dt} = -\frac{V_{\max} S}{K_m + S}, \frac{dAC}{dt} = 0, P = S_0 + P_0 - S \tag{2.1}$$

The above equation is a well-known model for an enzymatic process that can be applied when the enzymatic process can be modelled as an equilibrium reaction forming an active complex (AC) that in turn decays into product P and free enzyme E, according to equation (2.2) (see also Chapter 3).

$$S + E \underset{k_{s2}}{\overset{k_{s1}}{\rightleftarrows}} AC \xrightarrow{k_p} P + E \tag{2.2}$$

The variables $V_{\max}$ and K_m in the Michaelis Menten equation are combined variables: K_m is a function of k_{s1}, k_{s2} and k_p. $V_{\max}$ is a combination of the initial amount of enzyme E_0 and the specific rate of the enzymic reaction k_{p} (see equation (2.3)).[8,9]

$$V_{\max} = k_p E_0 \quad K_m = \frac{k_{s2} + k_p}{k_{s1}} \tag{2.3}$$

Using the laws of fundamental kinetics, the specific reaction rate k_p is likely to depend on temperature. The amount of enzyme is determined by the amount initially present and by the changes conceivably occurring during the process. Combining the variables E_0 and k_p into $V_{\max}$ obscures this interaction. As long as the exact meaning and relation between the original variables and the combined variable is recognized and acknowledged, no problem arises. However, if temperatures are allowed to vary, the situation changes drastically. The apparent rate of reaction ($V_{\max}$) will increase with increasing temperature, but at still higher temperatures the rate will decrease with increasing temperature. This complex process can be decomposed into two separate processes:

1. a normal temperature-dependent increase of the rate constant k_p, as for all (bio)chemical reactions, and
2. an increasingly important decrease in amount of active enzyme by inactivation at higher temperatures.

A similar line of reasoning holds for the variable K_m. This variable is a combination of reaction rates k_p, k_{s1} and k_{s2} that each will, most probably, change with temperature according to Arrhenius' law (equation 2.4):

$$k = k_{\mathrm{ref}} e^{\frac{E_a}{R}\left(\frac{1}{T_{\mathrm{ref}}} - \frac{1}{T_{\mathrm{abs}}}\right)} \tag{2.4}$$

Finally, the standard Michaelis Menten equation (equation 2.1) is formulated for a non-denaturating enzyme. The amount of enzyme is hidden in V_{max} (equation 2.3). When enzyme denaturation is taken into account the constant amount of enzyme E_0 has to be replaced with the variable amount of still active enzyme.

From this example it becomes evident that it may be hazardous to combine different variables (k_p and E_0), constant for a certain situation, into one variable (V_{max}). Unless its meaning remains clear and evident, further development of theory and application can be seriously hampered. Therefore, it is recommended to keep all variables separated in the mathematical formulation, to encourage the proper use of their separate and individual meanings and to allow for extending the model to new applications and situations.

2.4 Benefits for modelling food processes

In this chapter two methods for problem decomposition were described. Task decomposition divides a complex task into easier to solve subtasks. Object-oriented analysis and design identifies concepts and entities in the system under study and models them as separate objects. Each object is made responsible for one or more tasks in the system under study.

When modelling food processes the emphasis lies on the task decomposition to find the generic processes underlying the observed behaviour. Object-oriented analysis and design is seen as a step before this task decomposition. A decomposition into objects can reveal aspects of the phenomenon under study that may not be recognized easily when only task decomposition is applied. Once the objects are known, the modeller can focus on the behaviour of each object. An object-oriented design of the model has the advantage that the model can be easily extended, and that parts of the model can be reused more easily. Object-orientation facilitates reusability and extendibility, but it is still the modeller who has to keep in mind that an object that may be reused in the future must already be designed for reuse. An object can only be reused if the behaviour of the object is modelled independent of the actual phenomenon under study and if the assumptions underlying the model are explicitly documented.

The major benefit of a complete decomposition of the phenomenon under study into processes is that the modeller gains insight into the complex behaviour of the system. Because all interactions between the processes are made explicit, the modeller can make a sound decision about which processes have to be included in the model and which processes can be ignored. These decisions should be documented with the model to enable a future reuse of the model. For each variable that connects two processes the modeller should investigate whether other processes are relevant that also affect that variable. In this way additional behaviour and processes may be found that the modeller did not anticipate in a first analysis of the problem. Hence, by working towards a complete decomposition of the phenomenon under study a comprehensive

model for the complex process can be derived. Whether all processes are indeed included in the actual model depends on the required level of detail and on the availability of sufficient experimental data to calibrate the model.

A model formulation that is based on a thorough decomposition of the complex phenomenon into fundamental process also has advantages in the statistical analysis of experimental results. As explained earlier, the model should be formulated as a set of differential equations to be able to combine effects from different processes on one variable. Whenever an analytical solution of this set of differential equations can be obtained for constant external conditions, that analytical solution can be applied to the experimental data in nonlinear regression analysis. The estimated parameters constitute a reliable calibration of the formulated model. Furthermore, application of complex, multidimensional nonlinear regression with, for example, time and temperature as simultaneous explaining variables provides the statistical system with more information than is otherwise possible. With this technique the statistical system is able to place the measuring uncertainties where they belong according to the model.[10, 11] If the developed model accurately describes the observed phenomenon an explained part (R^2_{adj}) of up to 95% can frequently be obtained.[7]

Decomposing a complex process into subprocesses might result in the introduction of intermediate state variables that cannot yet be measured in the real world. In these cases the modeller must apply his knowledge about the problem domain and the fundamental laws applied to validate that the behaviour predicted by the model is correct. It constitutes on the one hand a disadvantage of problem decomposition, certainly at the level of model implementation. On the other hand, it may guide and force experimental research to new and unexplored areas.

2.5 Future trends

Application of the problem decomposition methods for modelling food processes described in this chapter may help to improve knowledge about the mechanisms underlying the observed behaviour of food products. Not only knowledge of the physiological processes and their interactions may improve, but also knowledge about which mathematical models can be used for these processes in various situations. Problem decomposition as described in this chapter will lead to detailed models with a high explanatory value. This level of detail may not always be needed. In that case, a detailed, fundamental and comprehensive model that was derived from a complete decomposition of a complex process, may be simplified in different ways for different dedicated purposes, yielding several models for the complex process at different levels of detail.

Knowledge about physiological, physical-chemical and biochemical processes, their interactions and the applicable mathematical models can be stored in some sort of library to facilitate reuse. With such a library, developing a

model for a phenomenon under study may evolve into describing the phenomenon as a collection of interactions between (generic) processes, followed by selecting the appropriate mathematical models for these processes. This two-stage approach to modelling complex processes has been applied in a system for automated modelling of post-harvest physiological behaviour.[4, 12]

The discussion about object-oriented analysis and design shows that in some cases the observed behaviour of a phenomenon under study can be described as interactions between objects in the real world. For example, the behaviour of a packaged product may be modelled as the result of interactions between an object representing the product and an object representing the package. The latter object will contain a model describing how the environment inside the package is related to conditions outside the package and the package material (see also Chapter 14). As proper model development based on proper problem decomposition is very much an expert task that requires a lot of knowledge of the problem domain, it is likely that the product model and the package model are developed by different persons or research groups. Hence, a proper decomposition of a phenomenon may also enable the distribution of the model development to different modellers.

In order to use the product model and the package model in one system, both models must have a specification of the interfaces and of the assumptions underlying the model formulation. The specification of the model interface defines which variables in the model can be varied from outside and which variables can be output from the model. The specification of the underlying assumptions can be used to evaluate whether the model is suitable for the phenomenon under study.

Such models together with their specifications of the interfaces and the underlying assumptions can be seen as reusable components. In information technology, building reusable software components is common practice. A recent method for software development that aims to facilitate the reuse of such off-the-shelf software components is component-based development.[13] A component is defined as a piece of software together with clear specifications of its functions, its interfaces and the requirements it needs to operate. The idea of component-based development is that this specification of the interfaces and requirements forms a kind of contract between the user and the developer of the component. The developer may adapt the internal implementation of the component as long as the component still adheres to this contract. In this way a component can be changed without affecting the functionality of the complete software system. Component-based development leads to software systems that are more flexible and can be maintained better.

Although component-based development focuses on building large software systems, the concepts may also be applied to modelling food processes. If models describing physiological behaviour of different products would adhere to a common interface, those models could be exchanged and easily used in different systems. It would then be possible to develop generic systems to simulate the effects of distribution chains on the distributed product or to

develop generic systems to simulate the effects of a package on the quality of the packaged product. A further advantage of using a component-oriented approach to developing simulation models concerns the update of simulation models with results from research into the mechanisms underlying the observed behaviour of food products. If the submodels of a simulation model are developed as reusable components, submodels can be replaced easily with new versions containing improved descriptions for the represented processes.

2.6 References

1. LARMAN C, *Applying UML and Patterns: an introduction to object-oriented analysis and design*, Prentice Hall, 1998.
2. FOWLER M, SCOTT K and BOOCH G, *UML Distilled, Second Edition: a brief guide to the standard object modeling language*, Addison-Wesley, 1999.
3. GAMMA E, HELM R, JOHNSON R, VLISSIDES J and BOOCH G, *Design Patterns*, Addison-Wesley, 1995.
4. SLOOF M, *Physiology of Quality Change Modelling: automated modelling of quality change of agricultural products*, Ph.D. Thesis, Vrije Universiteit Amsterdam, 1999.
5. SLOOF M, TIJSKENS L M M and WILKINSON E C, Concepts for modelling the quality of perishable products. *Trends in Food Science and Technology*, 1996, **7** (5), 165–71.
6. VAN DOORN W G and TIJSKENS L M M, FLORES: A Model on the Keeping Quality of Cut Flowers. *Agricultural Systems*, 1991, **35**, 111–27.
7. TIJSKENS L M M, OTMA E C and VAN KOOTEN O, Photosystem II quantum yield as a measure of radical scavengers in chilling injury in cucumber fruits and bell peppers: A static, dynamic and statistical model. *Planta*, 1994, **194**, 478–86.
8. CHANG R, *Physical Chemistry with Applications to Biological Systems*. Macmillan, New York, 1981.
9. KEEN R E and SPAIN J D, *Computer Simulation in Biology*. Wiley-Liss, New York, 1992.
10. ROSS G J S, *Nonlinear Estimation*, Springer Verlag, New York, 1990.
11. TIJSKENS L M M and EVELO R G, Modelling colour of tomatoes during post-harvest storage. *Postharvest Biology and Technology*, 1994, **4**, 85–98.
12. SLOOF M, Automated modelling of physiological processes during postharvest distribution of agricultural products. *Artificial Intelligence Review*, 1998, **12**, 39–70.
13. HERZUM P and SIMS O, *Business Component Factory: a comprehensive overview of component-based development for the enterprise*, John Wiley & Sons, 1999.

3

Kinetic modelling

M. A. J. S. van Boekel, Wageningen University, and
L. M. M. Tijskens, ATO-BV, Wageningen

3.1 Introduction

Kinetic modelling is a technique that is very useful in relation to food processing and food quality. The reason is twofold. First, changes in foods as a result of processing and storage lead to a change in quality (usually a quality loss). The processes involved are mainly (bio)chemical and physical reactions. Such changes proceed at a certain rate and with certain kinetics. Kinetic modelling enables us to describe these changes and their rates quantitatively. Second, with kinetic modelling we have a powerful tool that can help to unravel basic reaction mechanisms. The understanding of the basic mechanisms is vital for quality modelling and quality control.

To understand the progress of reactions, knowledge of thermodynamics and kinetics is required. Thermodynamics is helpful in describing and understanding in which direction a reaction will proceed and the energy and entropy changes that are involved. Thermodynamics thus explains the driving force for a reaction. However, thermodynamics cannot tell anything about the speed at which a reaction proceeds. This is the domain of kinetics. The rate with which a reaction proceeds is the resultant of the driving force and the resistance against change. There is thus an intimate link between thermodynamics and kinetics.

The field of chemical kinetics originated in the second half of the nineteenth century, for instance by scientists such as Arrhenius and van 't Hoff. In the early twentieth century Ball,[1] Stumbo[2] and Bigelow pioneered kinetic principles in food processing (mainly the canning industry). They introduced parameters such as the *D*-value (decimal reduction time, a measure for reaction rates) and the *z*-value (a parameter characterizing temperature sensitivity). These *D*- and *z*-value parameters are still used by industry and legislation, despite the known deviance

from reality. The use of kinetics is not limited to (bio)chemical reactions; it is also applicable to microbiological as well as physical changes such as crystallization, aggregation and coalescence.[3]

Nowadays numerical procedures are so reliable and computers so fast that it has also become possible to model very complicated reactions in foods. A basic rule in modelling should, however, not be forgotten, namely that a model should be as simple as possible ('Ockham's Razor'),[4] but nevertheless comply with the occurring processes. So, even in the era of computer-aided modelling, models should not be overparameterized. Kinetics has developed into a powerful tool for modelling of quality attributes in foods. Nevertheless, (bio)chemical and physical insight is a prerequisite for correct application of food quality modelling and kinetic parameters have to be extracted from experiments to calibrate the models. It is thus of utmost importance to establish a fundamental reaction mechanism and to derive kinetic parameters in the most accurate way. Correct application of kinetic principles is essential. In this chapter we will go deeper into these principles in relation to food and food processing.

3.2 Key principles and methods

3.2.1 Simple kinetics

Chemical reactions are basically monomolecular or bimolecular, very rarely termolecular. A monomolecular reaction results from an internal change in a molecule, and a bimolecular reaction is the resultant from two interacting molecules. The fundamental principles of kinetic modelling are based on the conversion of chemical reaction mechanisms into the constituting differential equations, applying the law of mass action. The rules of this conversion can be found in any good textbook on chemical or enzyme kinetics.[5–8]

The most simple example is an irreversible first order (monomolecular) decay or conversion reaction:

$$\mathrm{A} \xrightarrow{k} \mathrm{B} \tag{3.1}$$

This mechanism results in the following set of differential equations.

$$\frac{\mathrm{d[A]}}{\mathrm{d}t} = -k[\mathrm{A}]$$

$$\frac{\mathrm{d[B]}}{\mathrm{d}t} = k[\mathrm{A}] \tag{3.2}$$

Zero-order formation, where a product B is formed out of reactant A present in excess, can be represented by the same reaction mechanism:

$$\mathrm{A} \xrightarrow{k} \mathrm{B} \tag{3.3}$$

The constituting differential equation, assuming an excess and therefore constant concentration of component A, is:

$$\frac{\mathrm{d[B]}}{\mathrm{d}t} = k[\mathrm{A}] = k' \tag{3.4}$$

with k' the (pseudo) zero order rate constant; the rate is seen effectively to be independent on concentration of B. The rate will, however, depend on the constant concentration of A. In this model simplification, already a steady-state approximation is applied to the concentration of component A.

For an irreversible second order (bimolecular) reaction the following representation applies:

$$2\,\mathrm{A} \xrightarrow{k} \mathrm{B} \tag{3.5}$$

with the constituting differential equations:

$$\frac{\mathrm{d[A]}}{\mathrm{d}t} = -2k[\mathrm{A}]^2$$

$$\frac{\mathrm{d[B]}}{\mathrm{d}t} = k[\mathrm{A}]^2 \tag{3.6}$$

All these sets of differential equations can be solved easily at constant external conditions (mainly temperature and pH).

3.2.2 Complex kinetics

Reversible reactions are already more complex. Suppose we have the following reversible reaction (not yet at equilibrium)

$$\mathrm{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{B} \tag{3.7}$$

The differential rate equations are:

$$\frac{\mathrm{d[A]}}{\mathrm{d}t} = -k_1[\mathrm{A}] + k_2[\mathrm{B}]$$

$$\frac{\mathrm{d[B]}}{\mathrm{d}t} = k_1[\mathrm{A}] - k_2[\mathrm{B}] \tag{3.8}$$

The integrated rate equations, assuming an initial concentration of $\mathrm{A} = \mathrm{A}_0$ and of $\mathrm{B} = 0$, are:

$$[\mathrm{A}] = \frac{[\mathrm{A}]_0}{k_1 + k_2}\left(k_2 + k_1 \mathrm{e}^{-(k_1+k_2)t}\right)$$

$$[\mathrm{B}] = \frac{[\mathrm{A}]_0 k_1}{k_1 + k_2}\left(1 - e^{-(k_1+k_2)t}\right) \tag{3.9}$$

At a certain stage, the rates for the forward and the reverse reaction become equal, equilibrium is reached and the equilibrium constant K_{eq} is given by equation (3.10):

$$K_{eq} = \frac{k_1}{k_2} = \frac{[\mathrm{B}]}{[\mathrm{A}]} \tag{3.10}$$

An example of a reversible reaction relevant for foods is the mutarotation of reducing sugars.

Even more complex reactions are consecutive and parallel reactions. Consecutive reactions are reactions in which products are formed as intermediates, which then react further. The simplest example is:

$$\mathrm{A} \xrightarrow{k_1} \mathrm{B} \xrightarrow{k_2} \mathrm{C} \tag{3.11}$$

The differential rate equations for this case are:

$$\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = -k_1[\mathrm{A}]$$

$$\frac{\mathrm{d}[\mathrm{B}]}{\mathrm{d}t} = k_1[\mathrm{A}] - k_2[\mathrm{B}]$$

$$\frac{\mathrm{d}[\mathrm{C}]}{\mathrm{d}t} = k_2[\mathrm{B}] \tag{3.12}$$

The analytical solution at constant external conditions or integrated rate equations are:

$$[\mathrm{A}] = [\mathrm{A}]_0 e^{-k_1 t}$$

$$[\mathrm{B}] = [\mathrm{B}]_0 e^{-k_2 t} + k_1[\mathrm{A}]_0 \frac{e^{-k_1 t} - e^{-k_2 t}}{k_2 - k_1}$$

$$[\mathrm{C}] = [\mathrm{C}]_0 + [\mathrm{B}]_0(1 - e^{-k_2 t}) + [\mathrm{A}]_0 \left(1 + \frac{k_1 e^{-k_2 t} - k_2 e^{-k_1 t}}{k_2 - k_1}\right) \tag{3.13}$$

Figure 3.1 gives an example of a (hypothetical) consecutive reaction $A \rightarrow B \rightarrow C$ at various temperatures, assuming an activation energy of 5000 for the step $A \rightarrow B$ and 15000 for $B \rightarrow C$ (see section 3.2.5 for a further discussion on activation energy).

Parallel reactions imply that a reactant is subject to two or more different elementary reactions at the same time:

$$\begin{aligned} \mathrm{A} + \mathrm{B} &\underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathrm{P} \\ \mathrm{A} &\longrightarrow \mathrm{Q} \end{aligned} \tag{3.14}$$

The differential equation for component A is

$$-\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} = k_1[\mathrm{A}][\mathrm{B}] + k_2[\mathrm{A}] \tag{3.15}$$

but an analytical solution for the integrated rate equation is not found easily; one has to resort to a numerical solution in this case. An example of such a parallel

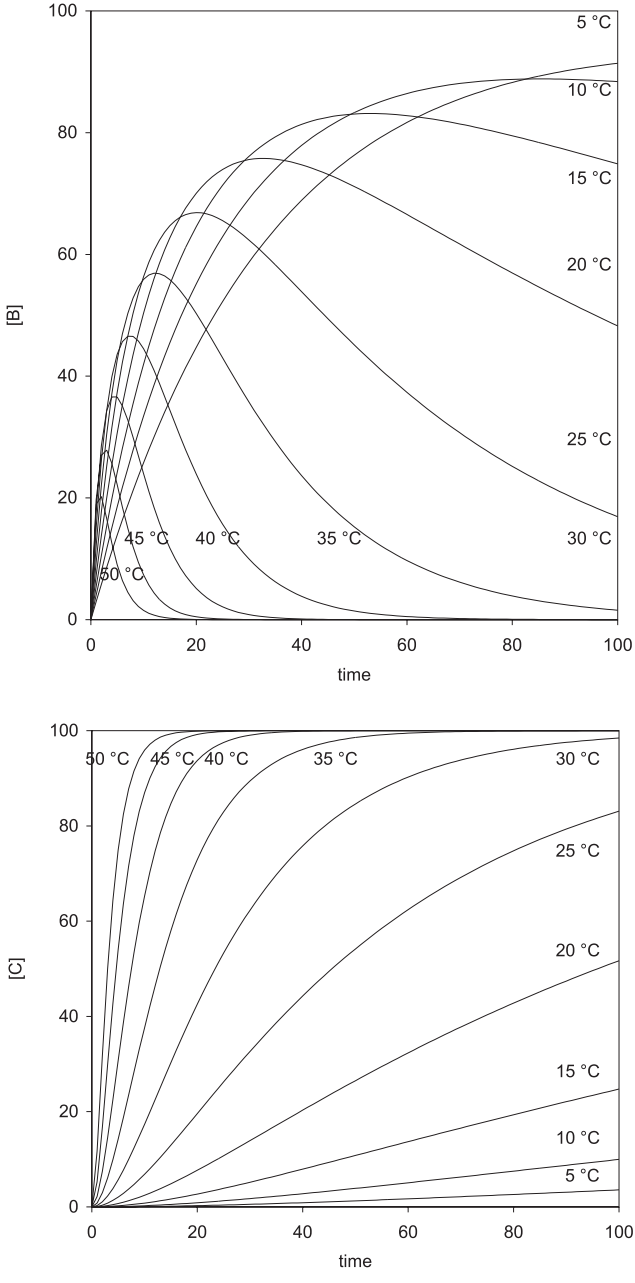


Fig. 3.1 Time course of a hypothetical consecutive reaction A → B → C for components B and C as a function of temperature with $E_a = 5000$ for the first reaction and $E_a = 15000$ for the second reaction (arbitrary units)

reaction in foods is the simultaneous isomerization of glucose and its participation in the Maillard reaction.[9] Maillard reactions may occur during sterilization of foods and result in colour and flavour changes.

There are many examples of parallel and consecutive reactions in foods, for instance, again, the Maillard reaction,[9] or degradation of chlorophyll during heating.[10] Another practical example is the change in activity of the enzyme polygalacturonase (PG) during storage.[11] Parallel and consecutive reactions are particularly amenable to multiresponse analysis, which has distinct advantages, as discussed in section 3.2.7.

Even for relatively simple cases, the analytical solutions of the constituting differential equations at constant conditions can be quite complicated. Many more possibilities of complex reactions exist. However, for more complex reactions than the ones given, it will be very tedious, if not impossible (for instance in cases such as equation (3.15), to derive analytical solutions. The only option that is left is numerical integration of the differential equations. Fortunately, this is no problem any more with modern computers and software.[12–16]

Bimolecular reactions require that molecules will have to come together before they can interact. Encounters may occur due to temperature-induced movement, diffusion and flow. If particles would react immediately upon an encounter, the rate of reaction is controlled by diffusion; such reactions are called diffusion controlled. On theoretical considerations, one can deduce that for the fastest bimolecular reaction rate possible in water at 20ºC the reaction rate constant is $6.6 \times 10^9\,\mathrm{dm^3\,mol^{-1}\,s^{-1}}$ and at 100ºC $3 \times 10^{10}\,\mathrm{dm^3\,mol^{-1}\,s^{-1}}$. These should be roughly the upper limits for bimolecular reaction rates at the temperature indicated.[3] For monomolecular dissociation in solution, the rate is determined by the rate at which the products can diffuse away. The upper limit for monomolecular reaction rate constants (uncharged reactions)[3] would then be roughly $10^{12}\,\mathrm{s^{-1}}$.

3.2.3 Steady-state approach

In the literature, one frequently approximates kinetic equations for consecutive reactions by assuming the so-called steady-state, or quasi-steady-state approximation (QSSA). In the above example of the consecutive reaction mechanism (equation (3.13)), intermediate B could be very reactive and have a fast turn-over rate. This effectively comes down to the situation that after some initial induction period, $\mathrm{d[B]}/\mathrm{d}t \approx 0$. Such an assumption greatly simplifies the resulting rate equations, and that is the very reason for introducing steady-state assumptions. Another advantage of a steady-state approximation is that it gives a 'feel' for the most important steps, and probably makes it more comprehensible. QSSA could also be helpful in 'mechanism reduction', i.e. reduction of the number of species and therefore reduction in the number of differential equations. To achieve this, it is necessary to identify reactants and products as important (those for which accurate calculation of concentrations is the aim), as necessary (those which are necessary to calculate the concentrations of the

important ones), or as redundant (those that can be omitted without appreciable effect on the reaction network). However, a steady-state approach should not really be necessary any more, because, as mentioned above, differential equations can be solved easily by numerical integration, should analytical integration appear to be impossible. Steady-state assumptions need not be made any more in this computer age, at least not from the standpoint that rate equations can otherwise not be derived any more. Some examples of software packages that work with differential rate equations rather than integrated rate equations are given in Stewart *et al.*[15] and Kuzmic.[16]

3.2.4 Enzyme kinetics

Enzyme kinetics are an example where the steady-state approach is used, namely in the famous Michaelis-Menten equation. Michaelis-Menten kinetics accounts for the kinetic properties of many enzymes but certainly not all. It provides, however, more a line of reasoning to deduce possible mechanisms and to develop useful models than a prescription of the mechanism. It is the most simple approach to enzyme kinetics. A relation is sought between the rate of product formation (rate of catalysis) and the concentration of enzyme and substrate. The development of Michaelis-Menten kinetics as it is used today was actually due to more researchers than Michaelis and Menten.[8] The first proposal came from Henry, later refined by Michaelis and Menten who assumed the establishment of a rapid equilibrium between enzyme E and substrate S leading to the formation of an enzyme-substrate complex ES. The active complex ES is then converted into product P liberating the enzyme E again:

$$\mathrm{E} + \mathrm{S} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_3} \mathrm{E} + \mathrm{P} \tag{3.16}$$

(Equation (3.16) describes, in fact, a consecutive reaction, discussed above.) There is also the Van Slyke equation, which differs from equation (3.16) in that the formation of the ES complex is assumed to be irreversible ($k_2 = 0$). It leads eventually to the same equation but with a slightly different meaning of the constants involved. Another assumption was that products do not revert to substrate, so the reaction E+P does not result in ES. That means that the rate of product formation v is:

$$v = k_3[\mathrm{ES}] \tag{3.17}$$

which is a 'normal' first-order rate process. Later, Briggs and Haldane[8] introduced the assumption that the rate of formation of the ES complex equals that of its breakdown (steady-state assumption):

$$k_1[\mathrm{E}][\mathrm{S}] = (k_2 + k_3)[\mathrm{ES}] \tag{3.18}$$

and hence an expression for [ES] is:

$$[\mathrm{ES}] = \frac{[\mathrm{E}][\mathrm{S}]}{(k_2 + k_3)/k_1} \tag{3.19}$$

The Michaelis constant K_M is now introduced:

$$K_M = \frac{k_2 + k_3}{k_1} \tag{3.20}$$

and equation (3.19) can be written as:

$$[\mathrm{ES}] = \frac{[\mathrm{E}][\mathrm{S}]}{K_M} \tag{3.21}$$

Meanwhile, the total concentration of the enzyme ($[\mathrm{E_T}]$) can be written as:

$$[\mathrm{E}_T] = [\mathrm{E}] + [\mathrm{ES}] \tag{3.22}$$

Substituting for [E] in equation (3.21) and solving for [ES] gives:

$$[\mathrm{ES}] = [\mathrm{E}_T]\frac{[\mathrm{S}]/K_M}{1 + [\mathrm{S}]/K_M} = [\mathrm{E}_T]\frac{[\mathrm{S}]}{[\mathrm{S}] + K_M} \tag{3.23}$$

Combining equation (3.23) with equation (3.17) gives:

$$v = k_3[\mathrm{E}_T]\frac{[\mathrm{S}]}{[\mathrm{S}] + K_M} \tag{3.24}$$

The expression $k_3[\mathrm{E_T}]$ represents the maximal rate v_max, namely when [S] is much greater than K_M and consequently $[\mathrm{S}]/([\mathrm{S}]+K_\mathrm{M})$ in equation (3.23) becomes unity so that:

$$v_\mathrm{max} = k_3[\mathrm{E}_T] \tag{3.25}$$

Hence, equation (3.25) can be written as:

$$v = v_\mathrm{max}\frac{[\mathrm{S}]}{[\mathrm{S}] + K_M} \tag{3.26}$$

and this is the famous Michaelis-Menten equation. Equation (3.26) describes the hyperbolic curve for the relation between rate v and [S] that is found with many (but certainly not all) enzymes (Fig. 3.2). The physical significance of K_M is that it represents the substrate concentration at which $v = 0.5v_\mathrm{max}$.

Some interesting features follow from equation (3.26). When $[\mathrm{S}] >> K_M$:

$$v = v_\mathrm{max} \tag{3.27}$$

and the result is a zero-order reaction for this condition. On the other hand, when $[\mathrm{S}] << K_M$:

$$v = [S]\frac{v_\mathrm{max}}{K_M} \tag{3.28}$$

Since v_max and K_M are constants, the rate is directly proportional to [S], i.e. a first-order reaction appears under this condition.

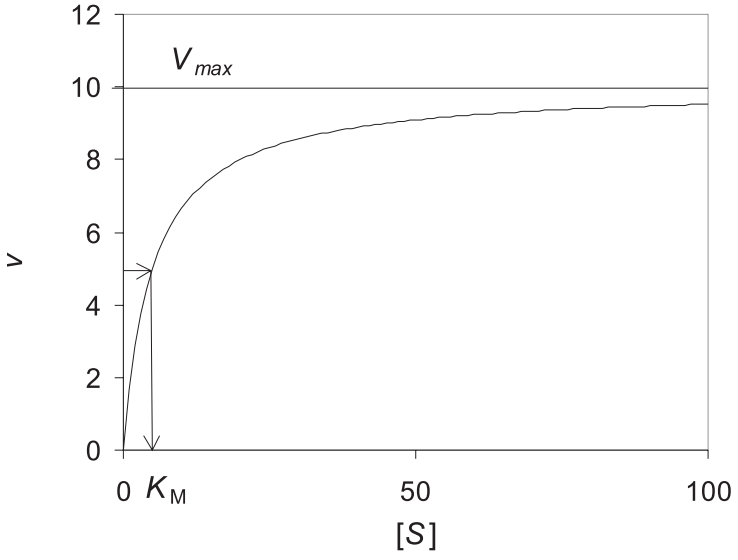


Fig. 3.2 Graphical depiction of Michaelis-Menten kinetics (arbitrary units).

3.2.5 Effects of temperature

Many foods are, for various reasons, heat treated or stored in chilled rooms and the effect of heat treatment on food quality is a very important issue. Temperature is also of importance with regard to keeping quality[23] or shelf-life. In general, temperature is the most important of all external factors. Therefore, knowledge of how the kinetics of reactions are affected by temperature is essential. Most food scientists would tend to use Arrhenius' law and derive activation energy from it. Arrhenius derived his equation empirically. Later on it was put into a theoretical perspective, especially for gas reactions, based on the collision theory, which incorporates time via molecular velocities and the number of favourably-oriented high-energy collisions (kinetic theory of gases). Arrhenius' equation appears to fit many reactions and is therefore used frequently. Although it may be a perfectly good choice in many cases (but not all), it seems appropriate to start at a somewhat more fundamental level by explaining relevant aspects of transition state theory, also referred to as the activated complex theory or absolute rate theory. This theory bridges the gap between thermodynamics and kinetics by postulating an equivalence between energy E and frequency of atomic motions ν, making it possible to deduce rate data from energy data (using the Planck expression $E = h\nu$, with h Planck's constant).

The transition state theory forms a theoretical basis to deduce more practical equations (such as Arrhenius' law). The theory is well suited for reactions in solutions and is not concerned with rates of encounters (like in gas reactions) but considers thermodynamic and statistical mechanics principles to predict how

many combinations of reactants will be present in the so-called transition state. This is a type of high-energy state in which molecules can be present in an unstable but activated condition; they will either turn back into reactants or undergo some molecular change to return as products. Consider the reactants A, B that are transformed into products P, Q via a transition state $AB^{\ddagger}$, as follows,

$$A + B \rightleftharpoons AB^{\ddagger} \rightarrow P + Q \tag{3.29}$$

Figure 3.3 shows schematically how the potential energy of the system changes with the reaction coordinate, i.e., the path along the potential energy curve. The reaction coordinate indicates the state of the molecules in the transition from reactants to products. When they start to interact the potential energy increases, and a maximum is reached in the activated, or transition state. It decreases again when products are formed.

The first step is considered to be a (quasi) equilibrium between the transition state and the reactant molecules, characterized by a practical equilibrium constant $K^{\ddagger}$ with dimension concentration^{-1} (hence not a thermodynamic one)

$$K^{\ddagger} = \frac{[AB^{\ddagger}]}{[A][B]} \tag{3.30}$$

The rate at which this equilibrium is established is fast compared to the rate of conversion of $AB^{\ddagger}$ to P, Q, so the position of the equilibrium is not perturbed

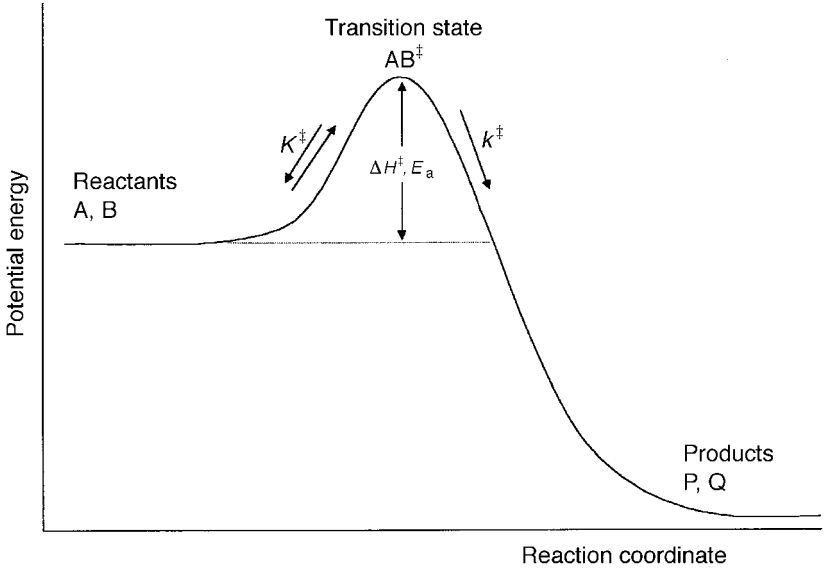


Fig. 3.3 Schematic presentation of the potential energy of reactants A, B, transition state $AB^{\ddagger}$ and products P, Q along the reaction coordinate. The energy barrier is the activation enthalpy $\Delta H^{\ddagger}$ (in the case of transition state theory) or the activation energy E_a (in the case of Arrhenius' equation).

significantly. To be explicit: $[AB^{\ddagger}]$ is not an intermediate that can be isolated, because of its instability. The second step, the formation of products, is considered to be unimolecular with a rate constant $k^{\ddagger}$ without any barrier, and considerations based on statistical mechanics result in

$$k^{\ddagger} = \frac{k_B T}{h} \tag{3.31}$$

in which k_B is Boltzmann's constant ($1.3807 \times 10^{-23}\,\mathrm{J\,K^{-1}}$), h is Planck's constant (6.626×10^{-34}J s) and T the absolute temperature (K). $k^{\ddagger}$ has dimension of frequency (s^{-1}). The rate of formation of products is thus

$$\frac{d[P]}{dt} = k^{\ddagger}[AB^{\ddagger}] = \frac{k_B T}{h}[AB^{\ddagger}] = \frac{k_B T}{h} K^{\ddagger}[A][B] \tag{3.32}$$

Comparing equation (3.32) with a 'normal' rate equation for a bimolecular one shows that the rate constant k can be expressed as

$$k = \frac{k_B T}{h} K^{\ddagger} \tag{3.33}$$

The equilibrium constant relates to the Gibbs energy, and consequently the enthalpy and entropy, of activation, as follows,

$$\frac{K^{\ddagger}}{(c^{\theta})^{1-m}} = e^{\frac{\Delta G^{\ddagger}}{RT}} = e^{\frac{\Delta S^{\ddagger}}{R}} e^{\frac{-\Delta H^{\ddagger}}{RT}} \tag{3.34}$$

The factor $(c^{\theta})^{1-m}$ is necessary to turn the practical equilibrium constant $K^{\ddagger}$ into a thermodynamic one, i.e., dimensionless. c^{θ} is the concentration in the standard state, usually chosen as mol dm^{-3}, and m is the molecularity (in this case 2, see equation (3.29)). Combination of equations (3.33) and (3.34) then gives

$$k = \frac{k_B T}{h} e^{\frac{-\Delta S^{\ddagger}}{R}} e^{\frac{\Delta H^{\ddagger}}{RT}} (c^{\theta})^{1-m} \tag{3.35}$$

Equation (3.35) has the correct units for a rate constant of any order because of the factor $(c^{\theta})^{1-m}$, the concentration in the standard state to which the thermodynamic parameters are referred. This equation is referred to as the Eyring equation, after one of the developers of the transition state theory. The importance of this equation is that it relates the effect of temperature on the reaction rate constant to fundamental terms of enthalpy and entropy changes. If, for instance, a high enthalpy of activation exists, this would make the reaction quite slow at moderate temperatures, but this may be compensated by an increase in activation entropy such that the reaction can still proceed at a measurable rate. A striking example of such a phenomenon is the unfolding of proteins. This indeed requires a high activation enthalpy because of the high number of bonds being broken simultaneously upon unfolding but, at the same time, the entropy of the unfolded chain increases enormously. In other words, high activation

enthalpies *and* entropies are characteristic for protein unfolding. On the other hand, bimolecular reactions usually have a negative activation entropy (because of bond rearrangements and bond formation, entropy of the two reactants is lost), and a moderate activation enthalpy (breaking old bonds will release energy, but forming new ones will cost energy). Unimolecular reactions are usually characterized by a moderate activation entropy (either slightly negative or positive, depending on intramolecular changes, the exception being protein unfolding) and an activation enthalpy depending on the type of mechanism.

The activation enthalpy and entropy are usually assumed to be independent of temperature, which in general is probably not true, but for the heat treatment of foods the temperature range is mostly not so large, so the approximation may hold. A notable exception is, again, protein unfolding in an aqueous environment, because interaction with water comes into play. Upon unfolding, hydrophobic groups are exposed and cause increased structuring of water. There is thus also a contribution of enthalpy and entropy changes of the solvent water which may oppose the positive enthalpy and entropy for protein unfolding. The ordered solvent structure around hydrophobic groups is broken down as temperature increases. Hence the difference in heat capacity between unfolded and folded (native) proteins is quite large, resulting in temperature dependency of (activation) enthalpy and entropy.

Arrhenius' law was derived empirically, but it found its roots in the kinetic theory of gases. It has proven to be very worthwhile in chemical kinetics. Arrhenius' law states that

$$k = Ae^{-\frac{E_a}{RT}} \tag{3.36}$$

in which A is the so-called pre-exponential factor (sometimes also called the frequency factor), and E_a the activation energy. It is very instructive to compare Arrhenius' law equation (3.36), with the expression derived from transition state theory equation (3.35). The dimension of A should be the same as that of the rate constant k; it therefore does have units of frequency only in the case of a first-order reaction. In fact A represents the rate of reaction at infinite temperature. To get rid of this rather undetermined factor, in practical applications the Arrhenius equation is often reparameterized in the reference form:

$$k = k_{ref} e^{\frac{-E_q}{R}\left(\frac{1}{T_{ref}} - \frac{1}{T}\right)} \tag{3.37}$$

where the index ref refers to an arbitrarily chosen value of T, preferably in the middle of the studied temperature region.

Obviously, E_a relates to the activation enthalpy $\Delta H^{\ddagger}$ and the exact relationship is found as follows. From equation (3.35) it follows that,

$$\ln(k) = \ln\left(\frac{k_B}{h}\right) - \ln\left(\frac{1}{T}\right) + \ln\left(\frac{K^{\ddagger}}{(c^{\theta})^{1-m}}\right) \tag{3.38}$$

hence,

$$\frac{\mathrm{dln}(k)}{\mathrm{d}(1/T)} = -T + \frac{\mathrm{dln}\left(\frac{K^{\ddagger}}{(c^{\theta})^{1-m}}\right)}{\mathrm{d}(1/T)} \tag{3.39}$$

and combining the temperature effect on $K^{\ddagger}$ (the van 't Hoff equation):

$$\frac{\mathrm{dln}(k)}{\mathrm{d}(1/T)} = -T - \frac{\Delta H^{\ddagger}}{R} \tag{3.40}$$

From the Arrhenius equation (3.36), it follows that

$$\frac{\mathrm{dln}(k)}{\mathrm{d}(1/T)} = -\frac{E_a}{R} \tag{3.41}$$

and consequently combination of equations (3.40) and (3.41) results in

$$E_a = \Delta H^{\ddagger} + RT \tag{3.42}$$

Likewise it can be deduced that the pre-exponential factor A is related to the activation entropy $\Delta S^{\ddagger}$. This makes the factor A much more comprehensible. (The physical meaning of A as such seems to be experienced as somewhat vague, which probably accounts for the fact that the factor A is very often not reported as a result in food science literature. It gives however as much useful information as does E_a.)

Another difference between Arrhenius' and Eyring's expressions is that the temperature T appears in the pre-exponential factor in Eyring's equation (3.35). This has a consequence in the way results are presented and analysed. Very often, Arrhenius' law is presented as a plot of ln(k) versus $1/T$, which should result in a straight line (if the relationship holds). With Eyring's relationship, $\ln(k/T)$ versus $1/T$ should be plotted. We would like to remark here that it is not a good idea to derive the activation energy parameters from linear regression of $\ln(k)$ or $\ln(k/T)$ versus $1/T$ because of the weighting of data points through logarithmic transformation; rather, non-linear regression should be used.[18] Another remark in this respect is that the two-step procedure of first deriving rate constants at several constant temperatures and then regressing them versus temperature usually results in very wide confidence intervals when only a limited number of temperatures have been studied, as is frequently the case. A better approach is to substitute expressions (3.35), (3.36) and (3.37) directly into the appropriate rate equations and perform a non-linear regression. In this way, all data are used to estimate the activation parameters and an estimate of these parameters of much higher precision is obtained. For representation purposes, it probably remains a good idea to provide Arrhenius' or Eyring's expression in the form of a plot of $\ln(k)$ or $\ln(k/T)$ versus $1/T$ because any deviation of the data from these expressions becomes immediately apparent. In doing so, however, the values of the parameters estimated by non-linear regression should be used to construct the plot. Deviations of Arrhenius' and Eyring's relationship are

indeed possible, but very unlikely. It probably indicates that another reaction influences the one under study, and that problem decomposition is conducted improperly. It is the responsibility of the researcher to check this. In the case that Arrhenius' or Eyring's equations are not applicable (for instance, because an undetected change in mechanism occurs at the higher temperatures), the resulting parameter estimates are worthless. The undetected changes in mechanism have to be incorporated first (see Chapter 2 on problem decomposition) to obtain reliable estimates. So, the first step should always be to check the validity of the laws of Arrhenius/Eyring, and only if they appear to be correct would the next step be the estimation of the activation parameters. Obvious as this may seem, this rule is not always obeyed.

3.2.6 pH effects in kinetic modelling

In many reactions systems pH is of utmost importance. Avoiding interference of changing pH is one of the reasons many experiments are conducted in buffered systems. By the buffering capacity of the applied solutions, the pH remains almost constant at the desired level.

Although the fundamentals of the effects of hydrogen ions on buffering solutions are well understood (as can be taken from almost any textbook on physical chemistry, e.g. Chang[7]), the effects of pH on reactions are still mostly modelled with empirical models. Especially in modelling microbial growth, enzyme activity and quality behaviour in agricultural products, polynomials are frequently used to incorporate the effects of constant but different levels in pH values.[24, 25] The major disadvantage of these models and their estimated parameters is, of course, the fact that it is almost impossible to transfer the values of parameters determined in one experiment to another situation. This limits drastically the application of such models.

The essential action for incorporating the effects of pH on various reaction systems is, again, to find the appropriate reaction mechanism for that system. As with the effect of temperature on the behaviour of reaction rate constants, when the effects of pH are not described by a basic mechanism, one has to attempt to postulate a mechanism, and not just a mathematical fit function.

To incorporate pH into kinetic modelling, one has to realize that pH is defined as $-\log([H^+])$. The concentration of H^+ ions is thus what is important, and all anticipated effects of pH have to be incorporated as H^+ ions. The simplest example is a H^+ catalysed conversion:

$$A + H^+ \xrightarrow{k} B + H^+ \tag{3.43}$$

The constituting differential equations that can be derived from this mechanism are:

$$\frac{\mathrm{d[A]}}{\mathrm{d}t} = -k\left[\mathrm{H+}\right][\mathrm{A}] \tag{3.44}$$

As the hydrogen ions are not consumed, the concentration of H^+ can be considered constant, and the analytical solution becomes:

$$[A] = [A]_0 e^{-k\left[H_0^+\right]t} \tag{3.45}$$

Exchanging $[H^+]$ for the appropriate pH expression gives:

$$[A] = [A]_0 e^{\left(-k10^{-pH}t\right)} \tag{3.46}$$

So, with this fundamental approach, it becomes not only clear what the effects of pH are, but also by what type of functions: in this mechanism it is effective via a double exponential function.

Another example is when the hydrogen ions are used up in a reaction:

$$A + H^+ \xrightarrow{k} B \tag{3.47}$$

The differential equation for [A] is the same as in the previous example. Now, when the pH is kept constant by buffering action, exactly the same solution as for the previous example emerges: the actual pH does not change. However, in unbuffered systems, the hydrogen ions are used up. By using the mass conservation law, that is $[H^+]-[A]$ is constant at any time, the analytical solution now becomes:

$$[A] = [A]_0 \frac{[H]_0 - [A]_0}{[H]_0 e^{k([H]_0 - [A]_0)t} - [A]_0} \tag{3.48}$$

Here one can see that a relatively small change in the reaction mechanism studied has a tremendous impact on the behaviour of the reaction components and on deduced equations.

These principles of applying the concentration of hydrogen ions instead of pH directly, has recently been used for describing the combined effects of pH and activating and denaturing temperatures on the activity of the enzyme phytase from various sources.[26] The principles and basic equations had previously been deduced,[5, 6, 7] separately for temperature and for pH, but it was not applied to experimental data hitherto. The combined effects of temperature and pH were validated on seven sets of independently measured data. The fit of the model on the obtained data was remarkable (see Fig. 3.4).

3.2.7 Multiresponse modelling

As indicated above, reactions in foods are usually quite complicated and the use of simple uniresponse kinetics, in which only one response is analysed, can only be an approximation of the underlying mechanism. A very useful technique for studying more complex changes in foods is the multiresponse modelling technique. If it is possible to measure more reactants and/or products simultaneously rather than taking one reactant or one product, one can apply multiresponse modelling. The major advantages of this are that proposed reaction

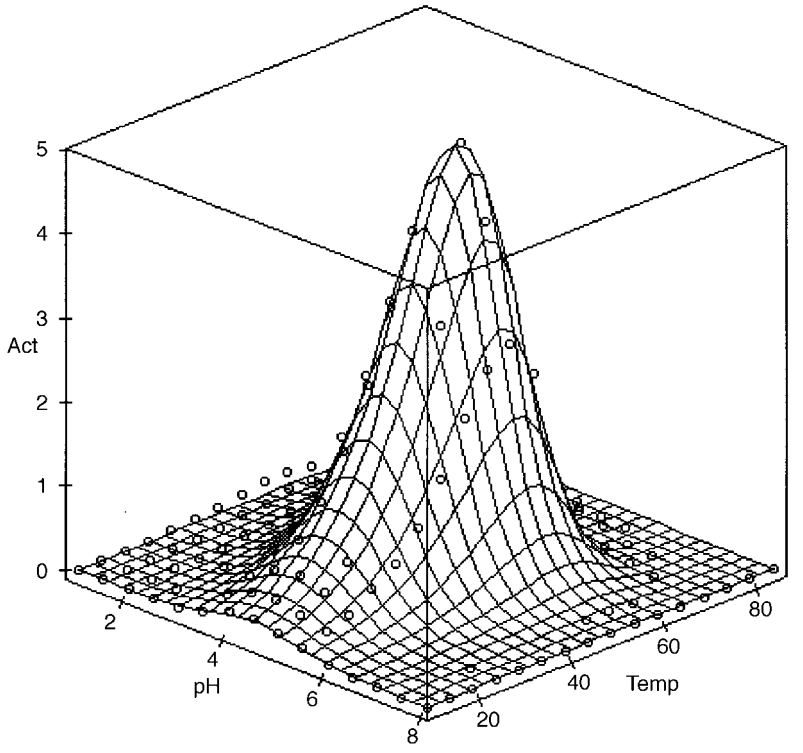


Fig. 3.4 Measured (symbols) and simulated (lines) activity for phytase produced by *E. coli*.

models can be tested much more rigorously, and that resulting parameter estimates are much more precise. The reason for these advantages is that the information that can be extracted from data is increased considerably. A disadvantage is that another regression criterion than the familiar least squares must be used, namely the determinant criterion. This is in itself not more difficult, but there are only a few software packages that handle this criterion. It is however very rewarding to apply multiresponse modelling in appropriate cases.[10, 20, 21]

An important aspect for multiresponse modelling is to take variances and covariances of the various responses into account. To clarify this point, a hypothetical reaction scheme is discussed first. Suppose three reactions take place at the same time, and we are able to measure all six components during the course of the reaction:

$$\mathrm{A} + \mathrm{B} \xrightarrow{k_1} \mathrm{C} + \mathrm{D}$$

$$\mathrm{C} \xrightarrow{k_2} \mathrm{E}$$

$$D + B \xrightarrow{k_3} F \tag{3.49}$$

with k_i as reaction rate constants. Then the following differential equations can be set up:

$$\frac{d[A]}{dt} = -k_1[A][B]$$

$$\frac{d[B]}{dt} = -k_1[A][B] - k_3[D][B]$$

$$\frac{d[C]}{dt} = k_1[A][B] - k_2[C]$$

$$\frac{d[D]}{dt} = k_1[A][B] - k_3[D][B]$$

$$\frac{d[E]}{dt} = k_2[C]$$

$$\frac{d[F]}{dt} = k_3[D][B] \tag{3.50}$$

These coupled ordinary differential equations (ODEs) can be solved by numerical integration. A well-suited algorithm is, for instance, the Gear routine, especially designed for so-called stiff differential equations (in which the parameters may have largely different values, which is frequently the case for kinetic rate constants).[12] Following this approach we can find a numerical solution describing the evolution of our six compounds over time.

Next, the model (i.e. the numerically integrated rate equations) should be fitted to the experimental data points. The question is how to do that properly. The 'natural' procedure for this would seem to be the method of least squares, for instance to minimize for component A:

$$\sum_{u=1}^{n} (y_A - \hat{y}_A)^2 \tag{3.51}$$

in which u $(1 \ldots n)$ is the number of experimental runs, y_A the experimental data points for component A, and $\hat{y}_A$ the predictions of component A by the model (as predicted by numerical integration). In the above example, there are several responses at the same time (the concentrations of components A, B, C, D, E, F at each time interval studied). The question is now whether the best fit criterion in the above example is simply to minimize the combined sum of squares for all responses (like for component A in (equation 3.51). There are several, rather strict, requirements for application of least squares, and these turn out to be very strict, especially in the case of multiresponse modelling. It is not always appreciated by researchers that the fit criterion to be used depends on the experimental error structure: the covariance matrix of the experimental errors, $\mathbf{Cov_{ee}}$, is of importance. For our hypothetical example, this matrix is

$$\mathbf{Cov_{ee}} = \begin{bmatrix} \sigma_{AA} & \sigma_{AB} & \sigma_{AC} & \sigma_{AD} & \sigma_{AE} & \sigma_{AF} \\ \sigma_{BA} & \sigma_{BB} & \sigma_{BC} & \sigma_{BD} & \sigma_{BE} & \sigma_{BF} \\ \sigma_{CA} & \sigma_{CB} & \sigma_{CC} & \sigma_{CD} & \sigma_{CE} & \sigma_{CF} \\ \sigma_{DA} & \sigma_{DB} & \sigma_{DC} & \sigma_{DD} & \sigma_{DE} & \sigma_{DF} \\ \sigma_{EA} & \sigma_{EB} & \sigma_{EC} & \sigma_{ED} & \sigma_{EE} & \sigma_{EF} \\ \sigma_{FA} & \sigma_{FB} & \sigma_{FC} & \sigma_{FD} & \sigma_{FE} & \sigma_{FF} \end{bmatrix} \tag{3.52}$$

The diagonal elements in the matrix $\mathbf{Cov_{ee}}$ represent the variances of each response (i.e., $\sigma_{AA} = \sigma_A{}^2$ and the off-diagonal elements the covariances (i.e., $\sigma_{AB} = \rho\sigma_A\sigma_B$ with ρ the correlation coefficient). The point is that in most cases the covariance matrix of the experimental errors will be unknown. It would be reasonable to assume for multiresponse measurements that measurements in different runs are not correlated, but components measured within one run are expected to be correlated (for instance because several components are determined in one sample). Hence, the covariances $\neq 0$ *within* a run; then the best-fit criterion is not least squares minimization, but minimization of the determinant of the so-called dispersion matrix **C** with elements

$$c_{ij} = \sum_{u=1}^{n} [y_u^i - \hat{y}_u^i] \cdot [y_u^j - \hat{y}_u^j] \tag{3.53}$$

in which i,j is the index of responses ($i,j = 1..r$) and u the index of experimental runs ($u = 1..n$). It is perhaps worth noting that the diagonal elements of matrix **C** correspond to the sum of squares for each of the responses. The point is that not only the sum of squares for each of the responses is taken into account but also the crossproducts of the responses. This analysis is due to Box and Draper,[17] and since then further developed by several authors. See, for instance, the review article by Stewart *et al.*[15]

If the covariance matrix, equation (3.52), happens to be known, the best-fit criterion is minimization of

$$\sum_{i=1}^{r} \sum_{j=1}^{r} \sigma^{ij} c_{ij} \tag{3.54}$$

in which σ^{ij} are the elements of the inverse of the matrix $\mathbf{Cov_{ee}}$. If, in addition, no correlation exists between responses ($\sigma^{ij} = 0$ for $i \neq j$) and the variances of the responses are known, minimization of the following is appropriate

$$\sum_{i=1}^{r} \sigma^{ii} \sum_{u=1}^{n} [y_u^i - \hat{y}_u^i]^2 \tag{3.55}$$

Equation (3.55) represents the case of weighted least squares. Finally, if σ^{ii} is equal for all responses the minimization criterion is:

$$\sum_{i=1}^{r} \sum_{u=1}^{n} [y_u^i - \hat{y}_u^i]^2 \tag{3.56}$$

and this is actually the least squares criterion for all responses, analogous to equation (3.51) for one component. Coming back now to the question whether or not least squares is the best fit criterion, the answer appears to be: only under the (rather strict) conditions that all variances are the same and that covariances within a run are zero. This is a situation normally not encountered in practice if several responses are measured at the same time. In conclusion, it turns out that in multiresponse modelling the determinant criterion, i.e., minimization of the determinant of matrix (3.53), is the alternative for least squares. In other words, the method of least squares is not well suited for dealing with multiresponse data.[15]

There are two major advantages of multiresponse modelling. The first is that kinetic models can be tested much more rigorously because all the information contained in the data is linked and used. The second advantage is that the resulting estimates of the parameters (once a model is acceptable in terms of goodness of fit and scientific understanding) are much more precise than with uniresponse modelling. Several examples have been given for multiresponse analysis of reactions occurring in foods.[9, 18–21]

3.2.8 The engineering approach

Above, we have described how kinetic equations can be derived from postulated chemical mechanisms. Mathematically, reactions can also be described in terms of the change in concentration c of a component over time t without considering an actual reaction mechanism:

$$-\frac{\mathrm{d}c}{\mathrm{d}t} = kc^n \tag{3.57}$$

with n the order of the reaction and k the reaction rate constant. This applies to an irreversible reaction of one component (in this case a decomposition, but formation is of course also possible). Integration of equation (3.57) at constant external conditions, gives:

$$\begin{aligned} c_t^{1-n} &= c_0^{1-n} + (n-1)kt \quad (n \neq 1) \\ c &= c_0 \exp(-kt) \qquad (n = 1) \end{aligned} \tag{3.58}$$

with c_0 the initial concentration. There are in fact two types of orders, the first one is found if one determines *initial* rates $-\mathrm{d}c/\mathrm{d}t$ for various concentrations and then one can determine the order n_c from equation (3.57), which is called the order with respect to concentration. The second possibility is to use equation (3.58) and follow the change in concentration over time, and find the best fit for the order n_t, called the order with respect to time.[3, 18, 22]

When the reaction mechanism reflects the processes occurring in reality, the two orders are necessarily the same. However, when the reaction mechanism applied is a simplification of the reaction process, both orders need not be the same. If $n_t < n_c$ then the reaction rate increases as the reaction moves on (auto-

catalysis); if $n_t > n_c$ the reaction rate decreases as the reaction progresses (auto-inhibition). This indicates that it is useful to determine both types of orders because the comparison of their values should make clear whether or not the mechanism used in the modelling effort changes during the course of the reaction when, for example, autocatalysis or inhibition occurs. If both orders appear to be the same, one can conclude that the reaction under investigation seems to be a simple one and does reflect the occurring processes. However, if both orders show discrepancies, this could be a starting point for further mechanistic investigation.

The value of the order n is usually reported to be between 0 and 3; it need not be an integer value. However, if the order is not integer, it is a clear indication that one should dig deeper into the mechanism at work. It is important to note that equations (3.57) and (3.58) only give a mathematical description, not a mechanistic description of what is going on. In mechanistic terms, a reaction is either monomolecular or bimolecular (very rarely termolecular). An order of 1 or 2 may thus be an indication of a mono- or a bimolecular reaction, but not necessarily. An order that is not 1 or 2 indicates a more complex behaviour, usually because one observes a combination of several reactions. Strictly speaking, the above analysis is only valid for simple irreversible reactions with only one reactant. This is not very realistic for foods. We call this approach the 'engineering approach' because it comes down to mathematical modelling rather than kinetic modelling and is useful only for calculation of conversion rates and the like, which is relevant for engineering purposes, but not for obtaining insight into reaction mechanisms. One should thus realize that it is a crude approximation of the underlying physical or chemical mechanism. It is certainly not valid to deduce mechanistic conclusions and it would be most dangerous to extrapolate beyond the experimental region for which the relation was established. In any case, it may be a starting point for further kinetic analysis. If one wants to dive into more basic reaction mechanisms, one has to take more complex kinetic equations into account, as discussed above.

3.3 Areas of application

Kinetics are very useful in describing changes occurring during food processing and storage. Processing almost always requires a compromise because besides the desired changes, undesired changes will also occur. For instance, during heat sterilization, enzymes and microorganisms are inactivated (desired reactions) but at the same time nutrients such as vitamins and amino acids are degraded (undesired reactions) and undesirable flavour and colour compounds may be formed. Especially during the heating up phase, activation and denaturation of enzymes occur in the intermediate temperature regions.[27–29]

When the effect of temperature on reactions in foods has been established, preferably in the form of the parameters discussed, i.e., activation energy/enthalpy and activation entropy/pre-exponential factor, the value of the

parameters needs some discussion. Occasionally, there seems to be some misunderstanding regarding interpretation. For instance, if a high activation energy is found, the conclusion is sometimes drawn that the reaction will proceed slowly or with difficulty. This is a wrong conclusion, because the reaction may actually proceed quite fast, namely at high temperature. The point is that a high activation energy indicates a strong temperature dependence, that is to say it will run very slowly at low temperature, but very fast at high temperature. Such differences in activation energy are exploited in processes such as HTST (high temperature short time pasteurization) and UHT (ultra-high temperature processing). These processes are designed based on kinetic knowledge, and result in products with the same shelf-life as their traditional counterparts but with a better quality.

The background of these processes is as follows. Chemical processes leading to quality loss usually have an activation energy in the range of 100 kJ/mol, whereas the inactivation of enzymes and microorganisms has a much higher activation energy of, say, between 200 and 500 kJ/mol. This means that chemical reactions are less temperature sensitive than inactivation of enzymes/microorganisms, in other words by employing a high temperature, microorganisms are inactivated rapidly in a short time, whereas this time is too short to result in appreciable chemical changes.[3] Figure 3.5 illustrates this. Temperatures employed in UHT processing are in the range of 140–150ºC and the times needed are then in the order of a few seconds. Another consequence of importance for foods is that reactions with a relatively low activation energy will proceed at a measurable rate at low temperatures, for instance during storage at low and moderate temperature. In other words, chemical reactions do not stop at low temperature. To prevent spoilage due to chemical reactions in such cases, other measures need to be taken such as drying and induction of glassy states (preventing diffusion of reactants) or by taking away a reactant (e.g. oxygen to prevent chemical oxidation).

When kinetics are applied in food processing, it is necessary to consider additional aspects such as the residence time distribution in equipment, the type

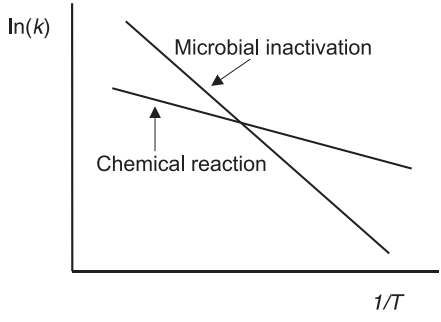


Fig. 3.5 Schematic presentation of the temperature dependence of a chemical reaction and microbial inactivation.

of flow (laminar v. turbulent) and heating-up and cooling-down effects in heat exchangers. These are typical scaling-up engineering problems. These problems are dealt with in Chapters 15 and 16.

When kinetics are applied to foods and living materials, a typical problem is that of compartmentalization, meaning that reactants are physically separated in different compartments or cells, so that they cannot interact. When foods are subsequently processed or damage of cells occurs, the reactants may come together and the reaction proceeds. This can be both desirable and undesirable. In any case it makes the application of kinetics more complicated.

3.4 Pros and cons of kinetic modelling

The advantages of kinetic modelling are manifold.

- The rules for building kinetic models are well rooted in the theories on chemical kinetics and thermodynamics. Consistent application of these rules leads to fundamental and generic models.
- The vast knowledge and information available in the literature and the expertise of experimental researchers can be applied in a way that is consistent with the prevailing and accepted theories.
- With generic models, extrapolations in areas outside the testing area are allowed, provided the processes are governed by the same mechanisms. This also means that model parameters can be validated on separate data sets, obtained for example in favourable laboratory circumstances, and applied in practical situations. Transfer of parameter values is then possible without any difficulty.[30]
- Consistent application of kinetic modelling will avoid, or at least diminish, the burden of scaling-up problems.
- A distinction can be made between kinetic parameters (all rate constants and energies of activation) for fundamental processes and batch parameters (depending on the material in which the reactions take place, i.e. matrix effects). Kinetic parameters are specific for a certain process and therefore have the same value for each repetition or duplication. Batch parameters will depend strongly on the composition of actual batches of agricultural products.[31]

Some disadvantages of kinetic modelling are as follows.

- It is often difficult, if not impossible, to detect and deduce the mechanism at work. Problem decomposition is a major assisting technique to overcome this disadvantage.
- Simplifying the mechanism, without including unnecessary processes and without excluding necessary processes is often very difficult, which is probably the reason that the ‘engineering approach’ is still widely popular.
- Correct application of kinetic modelling in foods requires insight in chemical kinetics, biochemistry, physics, mathematics and statistics and engineering,

as well as knowledge of the food matrix. It may be difficult to unite all this knowledge in one researcher. It may therefore be better to work in a team with specialists in each of these fields.

3.5 Future trends

One future trend in kinetic modelling is definitely that more complex models will be developed and applied; both developments in software (numerical methods) and in hardware (processor speed) make it possible to construct, analyse and apply very complex models. The bottleneck will, in fact, not be the modelling part, but the experimental validation of models. This remains, of course, a very important issue. The situation calls for far more intense cooperation between specialists in experimental design, experiment conduction, kinetic modelling and statistical analysis. A future trend should be that more attention is paid to the accuracy and precision of experimental determinations, to the precision of the parameters obtained, and to the precision of predictions. In other words, the statistical aspects of modelling and modelling applications should receive more attention. It makes no sense to have very fancy models incapable of making reliable predictions.

Another upcoming trend is molecular modelling with which it becomes possible to predict (bio)chemical changes based on the simulation of molecular behaviour. Although it is not yet possible, one could imagine the possibility of predicting chemical reactivity from the outcome of molecular modelling.

3.6 References

1. BALL C O and OLSON F C W, *Sterilization in Food Technology: theory, practice and calculations*, New York, McGraw-Hill, 1957.
2. STUMBO C R, *Thermobacteriology in food processing* (2nd edn), New York, Academic Press, 1972.
3. VAN BOEKEL M A J S and WALSTRA P, Use of kinetics in studying heat-induced changes in foods. IDF Monograph *Heat-induced changes in milk*. (2nd edn) P. F. Fox (Ed.), Brussels, International Dairy Federation, 1995.
4. RATKOWSKI D A, *Handbook of Nonlinear Regression Models*, New York, Marcel Dekker, 1990.
5. WHITAKER J R, *Principles of Enzymology for the Food Sciences* (2nd edn), New York, Marcel Dekker, 1994.
6. FERSHT A W H, *Enzyme Structure and Mechanism*, New York, Freeman & Co, 1984.
7. CHANG R, *Physical Chemistry with Applications to Biological Systems*, New York, Macmillan, 1981.
8. SEGEL I H, In *Enzyme Kinetics. Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*, New York, John Wiley & Sons, 1993.

9. VAN BOEKEL M A J S and BRANDS C M J, Heating of sugar-casein solutions: isomerization and Maillard reactions. In: *The Maillard Reaction in Foods and Medicine* (Eds. J. O'Brien, H. E. Nursten, J. Crabbe and J. F. Ames), Cambridge, Royal Society of Chemistry, 1998.
10. VAN BOEKEL M A J S, Modelling of chemical reactions in foods: a multiresponse approach. *Acta Horticulturae* (1998) **476** 149–55.
11. TIJSKENS L M M, RODIS P S, HERTOG M L A T M, KALANTZI U and VAN DIJK C, Kinetics of polygalacturonase activity and firmness of peaches during storage. *J Food Eng* (1998) **35** 111–26.
12. CHESICK J P, Interactive program systems for integration of reaction rate equations. *J Chem Ed* (1988) **65** 599–602.
13. TRNKA O, HARTMAN M and SVOBODA K, An alternative semi-implicit Euler method for the integration of highly stiff nonlinear differential equations. *Computers Chem Engng* (1997) **21** 277–82.
14. ESPOSITO W R and FLOUDAS C A, Global optimization for the parameter estimation of differential-algebraic systems. *Ind Eng Chem Res* (2000) **39** 1291–1310.
15. STEWART W E, CARACOTSIOS M and SØRENSEN J P, Parameter estimation from multiresponse data. *AIChE J* (1992) **38** 641–50.
16. KUZMIC P, Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal Biochem*, 1996 **237** 260–73.
17. BOX G E P and DRAPER N R, The Bayesian estimation of common parameters of several responses. *Biometrika* (1965) **52** 355–65.
18. VAN BOEKEL M A J S, Statistical aspects of kinetic modeling for food science problems. *J Food Sci* (1996) **61** 477–85, 489.
19. VAN BOEKEL M A J S, Kinetic modeling of sugar reactions in heated milk-like systems. *Neth Milk Dairy J* (1996) **50** 245–66.
20. VAN BOEKEL M A J S, Kinetic modelling in food science: a case study on chlorophyll degradation in olives. *J Sci Fd Agric* (2000) **80** 3–9.
21. VAN BOEKEL M A J S, Testing of kinetic models: usefulness of the multiresponse approach as applied to chlorophyll degradation in foods. *Food Res Int* (1999) **32** 261–9.
22. VAN BOEKEL M A J S, Kinetics of heat treatment. *Heat treatment and alternative methods*, Brussels, International Dairy Federation, 1996.
23. TIJSKENS L M M and POLDERDIJK J J, A generic model for keeping quality of vegetable produce during storage and distribution. *Agric Systems* (1996) **51** 431–52.
24. ZWIETERINK M, *Modeling of Microbial Quality*. PhD thesis, Wageningen Agricultural University, The Netherlands, 1993.
25. WIJTZES T, *Modelling the Microbial Quality and Safety of Food*. PhD thesis, Wageningen Agricultural University, The Netherlands, 1996.
26. TIJSKENS L M M, GREINER R, BIEKMAN E S A and KONIETZNY U, Modelling the effect of temperature and pH on the activity of enzymes: the case of phytases. *Biotechnology & Bioengineering* (2000) **72** (3) 323–30.
27. TIJSKENS L M M, RODIS P S, HERTOG M L A T M, WALDRON K W, INGHAM L,

PROXENIA N and VAN DIJK C, Activity of peroxidase during blanching of peaches, carrots and potatoes. *J Food Eng* (1997) **34** 355–70.

28. TIJSKENS L M M, WALDRON K W, NG A, INGHAM L and VAN DIJK C, The kinetics of pectin methyl esterase in potatoes and carrots during blanching. *J Food Eng* (1997) **34** 371–85.
29. SCHOKKER E P and VAN BOEKEL M A J S, Heat inactivation of the extracellular proteinase from Pseudomonas fluorescens 22F: inactivation during heating up and cooling periods. *J Dairy Res* (1999) **66** 467–72.
30. TIJSKENS L M M, HERTOG M L A T M and VAN DIJK C, *Generic Modelling and Practical Applications*. Proceedings COST 915 & COPERNICUS/CIPA Workshop on Food Quality Modelling, B.M. Nicolaï, J. de Baerdemaeker (eds.), Brussels, European Commission, 1998.
31. TIJSKENS L M M and WILKINSON E C, Behaviour of biological variability in batches during postharvest storage. *Aspects Applied Biology* (1996) **46** 267–8.

4

The modelling of heat and mass transfer

B. M. Nicolaï, P. Verboven, N. Scheerlinck, Katholieke Universiteit Leuven

4.1 Introduction

Heat and mass transfer processes are among the most important physical phenomena that occur during production and processing of foods. As a consequence of these processes, several important variables such as the temperature and the moisture concentration inside the food depend on time as well as on the position inside the food system. Since many product properties and quality attributes of foods (see Chapter 17 for definitions), such as microbial load, nutritional value, texture and organoleptic quality, are affected by these variables, they also depend on both time and space. The product properties and quality attributes of the food can, hence, no longer be considered as *lumped* variables which are homogeneous inside the food. For example, during thermal sterilisation of canned foods in steam retorts, the temperature close to the boundary of the can is higher than that in the centre of the can; as a consequence, the inactivation of pathogenic microorganisms such as *Clostridium botulinum* will be less advanced in the centre of the can. The design of the thermal process is, therefore, always based on the temperature course in that position in the can which receives the least intense heat treatment.

The objective of this chapter is to give an overview of mathematical models and principles for describing the transport of heat and mass in foods and their environment. Although the underlying physics of these processes have been well understood for many years, the governing partial differential equations cannot be solved analytically except for simple but unrealistic cases. However, several software packages are now available to solve realistic heat and mass transport problems by numerical means. These methods will be discussed as well.

The outline of the chapter is as follows. In Section 4.2. the Fourier equation for conduction heat transfer will be introduced, along with its corresponding boundary and initial conditions. Some analytical solutions will be given. It will be shown that mass diffusion is governed by the same equation, which is then called Fick's equation. The general transport equations – the continuity equation, the momentum equation and the enthalpy equation will be described in Section 4.3. It will be shown how the transport equations can be modified to take into account turbulence effects. Several types of boundary conditions which are relevant to food processes will be described as well.

In Section 4.4 Luikov's equations for coupled heat and mass transfer will be introduced. These equations are basically coupled diffusion equations which have been proposed to model heat and moisture transfer in porous media. In Section 4.5 several numerical methods to solve heat and mass transfer problems will be introduced. Special attention will be paid to the finite difference, finite element and finite volume method. An overview will be given of commercially available software packages. Applications in the area of thermal processing (heating, cooling and freezing) will be discussed in Chapters 15 and 16.

4.2 The diffusion equation

4.2.1 Derivation

In 1811 the French Académie de Sciences initiated a scientific contest with the following question: 'Donner la théorie mathématique des lois de la propagation de la chaleur et comparer le résultat de cette théorie à des expériences exactes' (To establish the mathematical theory of the laws that describe the propagation of heat, and to compare the results of this theory with exact experiments).

Joseph Fourier submitted on 28 September a study which consisted of a memoir, which was previously submitted (in 1807) to the Académie and which was partly published in 1808, and some additional notes. The jury, which included some famous mathematicians such as Lagrange, Laplace and Legendre, awarded him the prize at a public event on 6 January 1812, although they expressed their reservation because of the lack of mathematical rigor of Fourier's theory. Nevertheless, the study, which was in 1822 reprinted under the title *Théorie analytique de la Chaleur* (*The Analytical Theory of Heat*), is now considered as one of the most important scientific works of mathematical physics.[1]

Fourier considered only heat conduction processes in which heat is transported by molecular diffusion processes. Under this assumption, transient heat conduction in an isotropic object Ω with boundary Γ is governed by the equation which now bears his name and which is given by the following partial differential equation

$$\rho c \frac{\partial T}{\partial t} = \nabla k \nabla T + Q \quad \text{on } \Omega \tag{4.1}$$

where

ρ = density ($\mathrm{kg\,m^{-3}}$)
c = heat capacity ($\mathrm{J\,kg^{-1}{}^{o}C^{-1}}$)
k = thermal conductivity ($\mathrm{W\,m^{-1}{}^{o}C^{-1}}$)
Q = volumetric heat generation ($\mathrm{W\,m^{-3}}$)
T = temperature (°C)
t = time (s)

The *thermophysical parameters* k, ρ and c may be temperature dependent so that the problem becomes nonlinear. Thermophysical properties of various agricultural and food products are compiled in various reference books (e.g., the compilation made by the ASHRAE).[2] Further, equations have been published which relate the thermophysical properties of agricultural products and food materials to their chemical composition. In general, both the heat capacity and the density can be calculated with sufficient accuracy, but the models for the thermal conductivity require some assumptions about the orientation of the different main chemical constituents with respect to the direction of heat flow which is not always obvious.

In conventional thermal food processes the heat generation Q is zero. However, in the case of volumetric heating techniques such as microwave and ohmic heating, Q is the driving force of the heat transfer. The modelling of these techniques is a very active research area.[3–6]

The initial condition for the Fourier equation can be described as a spatial dependent function at time $t = 0$:

$$T(x, y, z, t) = T_0(x, y, z) \text{ at } t = 0 \tag{4.2}$$

At the boundary Γ of the heated or cooled object, fixed temperature (Dirichlet), convection or radiation conditions may apply:

$$\begin{aligned} &T(x, y, z, t) = f(x, y, z, t) \\ &\qquad\qquad\qquad\qquad \text{on } \Gamma \\ &k\frac{\partial}{\partial n_{\perp}}T = h(T_{\infty} - T) + \varepsilon\sigma(T_{\infty}^4 - T^4) \end{aligned} \tag{4.3}$$

with $f(x, y, z, t)$ a known function (e.g., it was measured, or it is known from control procedures), $n_{\perp}$ the outward normal to the surface, h the convection coefficient ($\mathrm{W/m^2\,{}^{o}C}$), T_{∞} the (known) ambient temperature, ε the emission coefficient, and σ the Stefan-Boltzmann constant. The surface heat transfer coefficient h must be considered as an empirical parameter.

Diffusion also occurs during the transport of species (water, gases) in tissue or food materials. It is governed by Fick's second law for diffusion mass transfer:[7]

$$\frac{\partial C_a}{\partial t} = \nabla D_a \nabla C_a + r_a \text{ on } \Omega \tag{4.4}$$

where

$C_a =$ molar concentration of component a (mol m^{-3})
$D_a =$ mass diffusion coefficient of component a (m^2 s^{-1})
$r_a =$ rate of production of component a (mol m^{-3} s^{-1})
$t =$ time (s).

Note that the above equation is only valid for diffusion in solids or stationary liquids with the assumptions of constant density ρ and zero mass velocity. Therefore, the above equation is only valid for slow pure diffusion of a single species with negligible changes in the total density. For other conditions, more complex transport equations should be applied.[7,8] In general, the mass diffusion coefficient D_a is not a constant but depends on the temperature and the concentration of the components in the mixture, as well as on pressure in gas systems. The production rate r_a depends on the metabolic activity of the product, which is a function of temperature and composition.

The initial condition for Fick's equation can be described as a spatial dependent function at time $t = 0$:

$$C_a(x, y, z, t) = C_{a0}(x, y, z) \text{ at } t = 0 \tag{4.5}$$

At the boundary Γ of the object, a fixed concentration (Dirichlet) or convection conditions may apply:

$$\begin{aligned} &C_a(x, y, z, t) = f(x, y, z, t) \\ &\qquad\qquad\qquad\qquad \text{on } \Gamma \\ &D_a \frac{\partial}{\partial n_\perp} C_a = h_{ma}(C_{a\infty} - C_a) \end{aligned} \tag{4.6}$$

with h_{ma} the empirical surface mass transfer coefficient (m s^{-1}), which, for water vapour in air, can be related to the surface heat transfer coefficient under the condition of low mass transfer rates or in a turbulent flow.[2] The Lewis relation then applies:

$$\frac{h}{h_{ma}\rho c} = \left(\frac{k}{\rho c D_a}\right)^{\frac{2}{3}} = Le^{\frac{2}{3}} \tag{4.7}$$

The ambient concentration $C_{a\infty}$ can be obtained from the perfect gas law.

4.2.2 Analytical solutions

Equations (4.1) and (4.4) can be solved analytically under a limited set of initial and boundary conditions for simple geometries only. Several solution techniques such as separation of variables, Green functions and variational methods are discussed in the many books on partial differential equations.[9,10] A large number of analytical solutions of the Fourier equation were compiled by Carslaw and Jaeger.[11]

Usually the Fourier equation is rewritten in dimensionless coordinates by introducing a dimensionless temperature ϑ and a dimensionless time Fo which is called the Fourier number

$$\vartheta = \frac{T - T_\infty}{T_0 - T_\infty} \tag{4.8}$$

$$Fo = kt/\rho c L^2 \tag{4.9}$$

with L a characteristic length, e.g., the half-thickness of a slab. For different geometries such as slab, cylinder and sphere, it can be shown that there exists a linear relationship between the logarithm of ϑ and Fo. For example, for a slab of half-thickness L subjected to convection boundary conditions, ϑ is given by

$$\vartheta = \sum_{n=1}^{\infty} \frac{4\ \sin(\zeta_n)}{2\zeta_n + \ \sin(2\zeta_n)} \exp(-\varsigma_n^2 Fo) \cos(\varsigma_n \frac{x}{L}) \tag{4.10}$$

and the discrete values of ζ_n are positive roots of the transcendental equation

$$\zeta_n \tan(\zeta_n) = Bi \tag{4.11}$$

where the Biot number Bi is defined as

$$Bi = \frac{hL}{k} \tag{4.12}$$

For $Fo \geq 0.2$, it can be shown that the infinite series in equation (4.10) can be approximated by the first term of the series. The graphical representation of the resulting relationship is commonly known as a Heissler chart and can be found in any standard textbook on heat transfer.[12]

4.3 The Navier-Stokes equations

4.3.1 Conservation equations

In fluids, transport of heat and mass is more complicated than in solid foods, as besides diffusion also convective transport of liquid particles may take place. The driving force behind convective transport is a pressure gradient in the case of forced convection, e.g. due to a fan in an oven, or density differences because of, e.g. temperature gradients. Navier and Stokes independently derived the equations for convective transport which now bear their names. For simplicity we will restrict the discussion to a single Newtonian fluid system. This means that we will only consider fluids for which there is a linear relationship between shear stress and velocity gradient, such as water or air. More complicated fluids such as ketchup, starch solutions, etc., are so-called non-Newtonian fluids, and the reader is referred to standard books on rheology for more details.[13]

When we apply the conservation principle to a fixed infinitesimal control volume $dx_1 dx_2 dx_3$ we obtain the continuity, momentum and energy equations, written in index notation for Cartesian coordinates x_i ($i = 1, 2, 3$ for the x-, y- and

z-direction, respectively), and whenever an index appears twice in any term, summation over the range of that index is implied (for example, $(\partial \rho u_j / \partial x_j)$ becomes $(\partial \rho u_1 / \partial x_1) + (\partial \rho u_2 / \partial x_2) + (\partial \rho u_3 / \partial x_3))$:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u_j}{\partial x_j} = 0 \tag{4.13}$$

$$\frac{\partial \rho u_i}{\partial t} + \frac{\partial \rho u_j u_i}{\partial x_j} = \frac{\partial}{\partial x_j} \eta \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) - \frac{\partial}{\partial x_i} \left(p + \frac{2}{3} \eta \frac{\partial u_j}{\partial x_j} \right) + f_i \tag{4.14}$$

$$\frac{\partial \rho H}{\partial t} + \frac{\partial \rho u_j H}{\partial x_j} = \frac{\partial}{\partial x_j} \left(k \frac{\partial T}{\partial x_j} \right) + \frac{\partial p}{\partial t} + Q \tag{4.15}$$

where

$u_i (i = 1, 2, 3)$	=	Cartesian components of the velocity vector **U** ($\mathrm{m\,s^{-1}}$)
T	=	temperature ($^{o}\mathrm{C}$)
H	=	static enthalpy ($\mathrm{J\,kg^{-1}}$)
p	=	pressure (Pa)
ρ	=	density ($\mathrm{kg\,m^{-3}}$)
K	=	thermal conductivity ($\mathrm{W\,m^{-1}\,{}^{o}C^{-1}}$)
η	=	dynamic viscosity ($\mathrm{kg\,m^{-1}\,s^{-1}}$)
f_i	=	external body forces ($\mathrm{N\,m^{-3}}$)
Q	=	heat source or sink ($\mathrm{W\,m^{-3}}$)

For a full derivation of these equations we refer to any textbook on fluid mechanics.[14, 15]

The system of five equations ([4.13]–[4.15], three equations for the velocity components plus the continuity and the energy equation) contains seven variables (u_1, u_2, u_3, p, h, T, ρ). We therefore need additional equations to close the system. The thermodynamic equation of state gives the relation between the density ρ and the pressure p and temperature T. The constitutive equation relates the enthalpy h to the pressure and the temperature. For an ideal gas we can use the following equations:

$$\rho = \frac{pM}{RT} \tag{4.16}$$

$$c = \left(\frac{\partial H}{\partial T} \right)_p \tag{4.17}$$

with M the molecular weight of the fluid ($\mathrm{kg\,mol^{-1}}$) and R the universal gas constant ($\mathrm{J\,mol^{-1}\,K^{-1}}$). When the heat capacity is assumed constant, the constitutive equation reduces to a linear relation between H and the difference between the actual temperature T and a reference temperature. Since only relatively low velocities are encountered in the food processes under consideration, the flow is often assumed incompressible and these equations can be applied.

For isothermal fluids we can assume that the density ρ is constant so that the continuity equation vanishes. In the case of non-isothermal flows the Boussinesq approximation is often applied, in which it is assumed that density is the only parameter which depends on the temperature.[15]

4.3.2 Turbulence

Many heat transfer processes in food operations often involve turbulent flow of air or water. Turbulence can be induced by the presence of flow obstructions such as baffles, shelves and the foods themselves. Turbulence is a state of the flow which is characterised by fluctuations of the flow variables (eddies) over a large range of scales, both in time and space. This complex pattern of motion enhances heat transfer rates considerably but also causes additional pressure drops which must be taken into account in the design of the equipment. Turbulence must therefore be incorporated in the governing models unless a laminar flow regime can be guaranteed.

Although the Navier-Stokes equations are general conservation equations which are equally well applicable to turbulent flow, the large variation of spatial scales introduces severe numerical problems, and only for simplified cases and low Reynolds numbers is it currently possible to perform such direct numerical simulations on supercomputers.[16] Simulation shortcuts are possible at different levels of complexity and approximations. The least approximations are needed in large eddy simulations, in which case the largest eddies are resolved but the effects of smaller eddies are estimated by additional models.[15] This approach is now being used more widely, since it is almost within reach of current computer power.

The most popular approach is based on the Reynolds Averaged Navier-Stokes (RANS) equations, which are obtained from averaging out the governing equations (Eqns 4.13–4.15) and including the effect of the turbulent fluctuations by additional models for the new terms appearing in the RANS equations. In the Boussinesq approach, the turbulence is accounted for by a turbulent 'viscosity' which is incorporated in the viscous and thermal diffusion transport terms. In $K-\varepsilon$ models, originally proposed by Jones and Launder, the turbulent viscosity η_t is obtained as a function of the turbulent variables K, which represents the turbulent kinetic energy associated with the fluctuating components of the flow velocities, and ε, the turbulent energy dissipation rate:[17]

$$\eta_t = \rho C_\eta \frac{K^2}{\varepsilon} \tag{4.18}$$

The constant C_η may be assumed constant for equilibrium conditions, where the turbulence production nearly equals the turbulence dissipation.

Additional transport equations have been derived for these turbulent flow variables. Several undefined constants appear in the model equations, which together with several assumptions and the specific near-wall treatment render this model empirical. There are three popular $K-\varepsilon$ models, namely the standard

$K-\varepsilon$ model, a RNG (Renormalisation Group) $K-\varepsilon$ model and a LRN (Low Reynolds Number) $K-\varepsilon$ model.[17–19] Verboven *et al.* compared these three turbulence models for a typical forced convection heating process of complexly shaped foods, and concluded that the boundary layers are badly represented by the wall function approach and the departure from local equilibrium is not accounted for.[20] A correction function can be added to correct for the latter behaviour in conjunction with the Low Reynolds Number model (see the work of Yap).[21] Nevertheless, it was found that experimental input for these corrections is needed in order to determine important constants.

More complex closures for the RANS models are based on dynamic equations for the Reynolds stresses and fluxes themselves in the RANS equations. In addition to the equations for the mean flow, this approach results in seven more partial differential equations. These models are believed to be more accurate but require a better insight into the process of turbulence and care must be taken with their numerical solution. Finally, it must be noted that new turbulence models are constantly proposed and tested.

4.3.3 Initial and boundary conditions

Unlike the diffusion equations, there are no conclusive general rules for the implementation of boundary conditions for the Navier-Stokes equations in order to have a well-posed problem because of their complex mathematical nature. For a full account, we refer to Hirsch.[22] For incompressible and weakly compressible flows, it is possible to define Dirichlet boundary conditions (fixed values of the variables, mostly upstream), Neumann boundary conditions (fixed gradients, mostly downstream) and wall boundary conditions (a wall function reflecting the behaviour of the flow near the wall). Initial values must be provided for all variables.

Difficulties arise when the exact conditions are unknown. This is especially true in turbulent flows, where the exact values of the turbulence energy and energy dissipation rate are often unknown at the inlet, and need to be guessed using information about the velocity and the flow geometry. The direction of the flow at boundaries may be difficult to specify, but may have considerable influence when the flow contains swirls. The effect of the pressure resistance (e.g. in a cool room) on the fan flow rate may be considerable and cannot always be taken into account appropriately. In any case a sensitivity analysis can be useful to obtain an error estimate associated with approximate or guessed boundary conditions.

4.3.4 Additional equations

In the case of an air flow through a bulk of products (e.g. cooling of horticultural products), or a flow of multiple fluids (e.g. the dispersion of disinfectants in a cool room or the injection of water for air humidification), the separate phases need to be considered. Depending on the flow conditions, a *multi-phase modelling* or a *mixed-fluid modelling* approach can be applied.

The mixed-fluid modelling approach is the least complicated and assumes that the mixture shares the same velocity, total pressure and temperature field. This is valid for moist air flow, where the water vapour transport is considered by means of an additional convection-diffusion conservation equation for the mass fraction of water vapour:

$$\frac{\partial \rho X_a}{\partial t} + \frac{\partial}{\partial x_j}\rho u_j X_a = \frac{\partial}{\partial x_i}\rho D_a \frac{\partial}{\partial x_i} X_a + r'_a \tag{4.19}$$

with X_a the mass fraction of water vapour, and r'_a [kg m^{-3} s^{-1}] a source/sink of water vapour (e.g. evaporation or condensation). Note that this equation is the full conservation form of equation (4.4), considering a variable density and allowing a mass flow of the mixture.

In the case of a dispersion of particles (e.g. water droplets from a humidifier, a spray of moist particles that need to be dried or the dispersion of disinfectants), the mixed model approach should include the particle-fluid interaction. The simplest approach is to include a *Langrangian model* for the individual particle motion, directly from Newton's second law:

$$m_p \frac{du_{pi}}{dt} = F_i \tag{4.20}$$

with m_p (kg) the particle mass, u_{pi} (m/s) a component of the velocity of the particle and F_i (N) the total force on the particle. The major component of this force is due to drag on the particle exerted by the moving fluid. Other contributions arise from a pressure gradient force, a buoyancy force and an added mass force. In addition, mass and heat may be exchanged between the particle and the fluid (consider exchange of a single mass component and a lumped system):

$$\begin{aligned} \frac{dm_p}{dt} &= -A_p h_{ma}(X_{ap} - X_a) \\ m_p c_p \frac{dT_p}{dt} &= -A_p h(T_p - T) + \frac{dm_p}{dt} h_{fg} \end{aligned} \tag{4.21}$$

where A_p (m^2) is the surface area of the particle, X_{ap} is the mass fraction of transferred component on the surface of the particle in equilibrium with the air, c_p (J/kg°C) is the specific heat of the particle, T_p (C) its temperature and h_{fg} (J/kg) the latent of evaporation of the transferred component.

The above exchanges of momentum, mass and heat are of course sources to the continuum fluid phase, which have to be included in the system of conservation equations (4.13)–(4.15). In a turbulent flow, the effect of eddies on the dispersion of the particles has to be taken into account. One straightforward solution is to randomly sample the turbulent components of the fluid velocity and calculate the effect on the particle motion. The disadvantage of this method is the large number of particles that has to be tracked to obtain a meaningful result. Variability in particle size can be taken into account by means of appropriate size distributions from which particles are sampled.

When the different phases have distinct velocity and temperature fields (e.g. two viscous liquids that are mixed or the airflow through a bulk of apples), the above approach is not valid and a *multi-phase approach* should be applied. In this case, the Navier-Stokes equations need to be solved for the separate phases. The coupling of momentum, mass and heat transfer remains and appropriate empirical formulations should be introduced for these inter-phase transfers. Furthermore, a volume fraction is assigned to each phase, which can vary in time (in the case of a mixed flow of fluids) or is a known fixed parameter (in the case of a bulk of products). In both cases, the volume fraction can vary with the spatial coordinates.

Finally, the problem may contain *chemical kinetics* (e.g. microbial activity and active components of disinfectants in cool rooms). In this case, the active species have to be tracked by means of a transport equation similar to equation (4.19). In addition, the chemical reaction must be solved. Therefore the reaction rates, property changes and heat releases must be calculated as part of the solution. Consider the following reaction:

$$A + B \longrightarrow C \tag{4.22}$$

The reaction rate R_c (mol s^{-1}) is defined as

$$R_c = -\frac{d}{dt}[A] = -\frac{d}{dt}[B] = \frac{d}{dt}[C] = k_f [A]^n [B]^m [C]^o - k_b [A]^p [B]^q [C]^r \tag{4.23}$$

with k_f the forward rate constant and k_b the backward rate constant. The rate constants can be modelled by the following Arrhenius-like expression:

$$k_{f,\,b} = aT^b e^{(E/RT)} \tag{4.24}$$

with a and b empirical constants and E the empirical activation energy.

The heat of reaction can be calculated from the heats of formation of the species and depends on temperature. The reaction leads to sources/sinks in the species conservation and energy equations.

4.4 Heat and mass transfer in porous media: Luikov's equations

Because most solid foods contain moisture, the heat that is applied to it forces the moisture to evaporate inside the porous product and causes moisture gradients. Hence, heat is also transferred due to moisture transfer and vice versa. The Fourier equation does not apply in this case. More complicated models should be introduced to take into account the simultaneous transfer of heat and mass inside foods and their surroundings.

Luikov's approach to heat and mass transfer in capillary-porous bodies is based on irreversible thermodynamics. Luikov considered a system consisting of a capillary-porous body and a bound substance, as shown in Fig. 4.1. In the

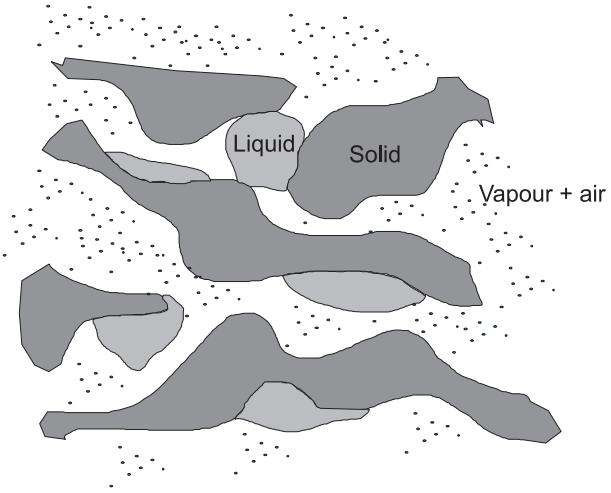


Fig. 4.1 Scheme of a capillary-porous body.

range of positive temperatures, the substance bound with the capillary-porous body can be in the form of liquid, vapour or gas. In a capillary-porous body transfers of the bound substance take place simultaneously in different states.[23]

Setting up for each bound substance the differential equations of heat and mass transfer, applying the principles of irreversible thermodynamics and taking the overall sum, results in Luikov's coupled system of partial differential equations.

$$c_{11}\frac{\partial T}{\partial t} = \frac{\partial}{\partial x_i}\left(k_{11}\frac{\partial}{\partial x_i}T + k_{12}\frac{\partial}{\partial x_i}\Psi\right) \tag{4.25}$$

$$c_{22}\frac{\partial \Psi}{\partial t} = \frac{\partial}{\partial x_i}\left(k_{21}\frac{\partial}{\partial x_i}T + k_{22}\frac{\partial}{\partial x_i}\Psi\right) \tag{4.26}$$

where

$$\begin{aligned} c_{11} &= \rho c & c_{22} &= \rho c_m \\ k_{11} &= k + \frac{\zeta h_{fg} k_m \delta}{c_m} & k_{12} &= \zeta h_{fg} k_m \\ k_{21} &= \frac{k_m \delta}{c_m} & k_{22} &= k_m \end{aligned} \tag{4.27}$$

In Table 4.1 the thermophysical parameters and variables appearing in equations (4.25)–(4.27) are compiled.

Table 4.1 Thermophysical properties and variables appearing in Luikov's model

c_m	moisture capacity	$\mathrm{kg\,kg^{-1}\,{}^oM^{-1}}$
c	heat capacity	$\mathrm{J\,kg^{-1}\,{}^oC^{-1}}$
δ	thermo gradient	$\mathrm{{}^oC^{-1}}$
ζ	ratio of vapour diffusion to total moisture diffusion	
k_m	moisture conductivity	$\mathrm{kg\,m^{-1}\,s^{-1}\,{}^oM^{-1}}$
k	thermal conductivity	$\mathrm{W\,m^{-1}\,{}^oC^{-1}}$
h_{fg}	latent heat	$\mathrm{J\,kg^{-1}}$
ρ	density	$\mathrm{kg\,m^{-3}}$
t	time	s
T	temperature	^{o}C
Ψ	moisture potential	^{o}M

Table 4.2 Thermophysical properties and variables related to the boundary conditions applied to Luikov's model

h_m	convective mass transfer coefficient	$\mathrm{kg\,m^{-2}\,s^{-1}\,{}^oM^{-1}}$
h	convective heat transfer coefficient	$\mathrm{W\,m^{-2}\,{}^oC^{-1}}$
$n_\perp$	outward normal to the surface	
T_∞	ambient temperature	^{o}C
Ψ_∞	ambient moisture potential	^{o}M
Γ_C	convective heat transfer boundary	
Γ_T	specified temperature boundary	
Γ_Ψ	specified moisture potential boundary	

Luikov conducted a large number of investigations to validate the theory and to determine experimentally the values of the parameters for a number of materials. The approach appeared to model the physical process well.[24]

At the boundary of the capillary-porous object, two types of boundary conditions can be applied, namely specified potentials of heat and mass transfer,

$$T = T_\infty \text{ on } \Gamma_\mathrm{T},\ \Psi = \Psi_\infty \text{ on } \Gamma_\Psi \tag{4.28}$$

or convection heat and mass transfer,

$$k_q \frac{\partial T}{\partial n_\perp} + h(T - T_\infty) + (1-\varepsilon)\lambda h_m(\Psi - \Psi_\infty) = 0$$

$$k_m \frac{\partial \Psi}{\partial n_\perp} + \frac{k_m \delta}{c_m}\frac{\partial T}{\partial n}\frac{\partial T}{\partial n} + h_m(\Psi - \Psi_\infty) = 0 \qquad \text{on } \Gamma_\mathrm{C} \tag{4.29}$$

The thermophysical parameters and variables related to the boundary conditions are compiled in Table 4.2.

Although Luikov's model has been applied successfully to thermal food processes, few parameter sets are available in the literature.[25–28] Moreover, some parameters have no clear physical meaning and it is, hence, difficult to assign reasonable values to them without prior knowledge.

4.5 Numerical methods

4.5.1 Numerical discretisation

For realistic – and thus more complicated – heat and mass transfer problems usually no analytic solution is available, and a numerical solution becomes mandatory. For this purpose the problem is reduced significantly by requiring a solution for a discrete number of points (the so-called *grid*) rather than for each point of the space-time continuum in which the heat and mass transfer proceed. The original governing partial differential equations are accordingly transformed into a system of *difference equations* and solved by simple mathematical manipulations such as addition, subtraction, multiplication and division, which can easily be automated using a computer. However, as a consequence of the discretisation the obtained solution is no longer exact, but only an *approximation* of the exact solution. Fortunately, the approximation error can be decreased substantially by increasing the number of discretisation points at the expense of additional computer time.Various discretisation methods have been used in the past for the numerical solution of heat conduction problems arising in food technology. Among the most commonly used are the finite difference method, the finite element method, and the finite volume method. It must be emphasised that – particularly in the case of nonlinear heat transfer problems – the numerical solution must always be validated. It is very well possible that a plausible, convergent but incorrect solution is obtained. At least a grid dependency study must be carried out to verify whether the solution basically remains the same when the computational grid is refined.

4.5.2 The finite difference method

Principle

The finite difference method is the oldest discretisation method for the numerical solution of differential equations and had been described as long ago as 1768 by Euler. The method is based on the approximation of the derivatives in the governing equations by the ratio of two differences. For example, the first time derivative of some function $T(t)$ at time t_i can be approximated by

$$\left.\frac{dT}{dt}\right|_{t_i} \cong \frac{T(t_{i+1}) - T(t_i)}{\Delta t} \tag{4.30}$$

with $\Delta t = t_{i+1} - t_i$. This expression converges to the exact value of the derivative when Δt decreases. The power of Δt with which the so-called truncation error decreases is called the order of the finite difference approximation, and can be obtained from a Taylor series approximation of T at time t_i.

Equation (4.30) is called a forward difference as it uses the future value of the function and it is of order 1. A backward difference of order 1 is given by

$$\left.\frac{dT}{dt}\right|_{t_i} \cong \frac{T(t_i) - T(t_{i-1})}{\Delta t} \tag{4.31}$$

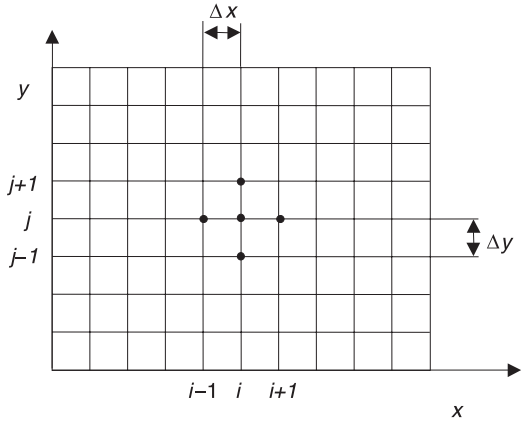


Fig. 4.2 Finite difference grid of a two-dimensional rectangular region. The nodes which are involved in the computation of the temperature at position (i, j) are indicated by dots.

Equations (4.30) and (4.31) are also called forward and backward Euler schemes. Likewise, finite difference formulas can be established for second order derivatives. The so-called central difference formula is of order 2 and is defined by

$$\left.\frac{d^2T}{dt^2}\right|_{t_i} \cong \frac{T(t_{i+1}) - 2T(t_i) + T(t_{i-1})}{\Delta t^2} \tag{4.32}$$

The finite difference method will be illustrated for a 2D heat conduction problem. For this purpose the computational domain is subdivided in a regularly spaced grid of lines which intersect at common nodal points (Fig. 4.2). Subsequently, the space and time derivatives are replaced by finite differences. For example, if central differences are used it is easy to see that the following expression is obtained for the Fourier equation:

$$\frac{\partial T_{i,j}}{\partial t} = \frac{k\Delta t}{\rho c}\left(\frac{T_{i+1,j} - 2T_{i,j} + T_{i-1,j}}{(\Delta x)^2} + \frac{T_{i,j+1} - 2T_{i,j} + T_{i,j-1}}{(\Delta y)^2}\right) \tag{4.33}$$

Similar equations can be established for all interior nodes of the grid, and special procedures are available to discretise the boundary conditions in the nodes which are on the boundary of the grid. The large number of equations (equal to the number of nodal points) can conveniently be ordered into a differential system of the general form

$$\mathbf{C}\frac{d}{dt}\mathbf{u} + \mathbf{K}\mathbf{u} = \mathbf{f} \tag{4.34}$$

with $\mathbf{u} = [u_1\, u_2 \cdots u_N]^T$ the nodal temperature vector. This vector differential equation can be discretised in time, and typically leads to a system of algebraic

equations which must be solved by appropriate means. The system matrices contain many zeros, and this feature can be exploited advantageously to reduce the required number of computations.

Applications
The finite difference method has been used for the prediction of the temperature course in the centre of canned foods in cylindrical containers, and rectangular bricks under various processing conditions.[29–31]

4.5.3 The finite element method

Principle
The mathematical foundations of the finite element method were established at the beginning of the twentieth century by Ritz and Galerkin. Based on variational calculus, Ritz developed in 1909 a method for the solution of partial differential equations.[32] He assumed that an approximate solution of the governing differential equation could be represented by a series of analytical functions (*trial functions*), with unknown coefficients. He then determined these coefficients by minimising a functional with respect to these coefficients. The method is restricted to problems for which such an equivalent minimisation problem (the so-called *variational principle*) can be established, which is not always the case.

In 1915 the Russian engineer Galerkin presented a related method for the computation of the elastic equilibrium of rods and thin plates. As in Ritz's method, he expressed the approximate solution of the partial differential equation as a function series with unknown coefficients. Substitution of the approximate solution in the differential equation produces in general a non-zero residual. Galerkin determined the unknown coefficients of the series by orthogonalisation of this residual with respect to another set of analytical functions (the *test functions*). The same functions were used for both the test as well as the trial functions. The Galerkin method is applicable to problems (e.g., nonlinear conduction heat transfer) for which no variational principle can be found. However, it can be proven that for each Ritz solution always a corresponding Galerkin solution can be found, as shown by White.[33]

The first application in the Western literature of the Galerkin method to solve transient heat conduction problems can be attributed to Bickley.[34] The Galerkin method was later generalised to the *method of weighted residuals* by using any set of linearly independent and complete functions as test functions, see Crandall.[35] Other popular choices are Dirac functions (the *collocation method*) and power series of the spatial coordinates the (*method of moments*).

In the traditional weighted residual and Ritz method the trial functions are often trigonometric functions or polynomials that span the whole computational domain. This introduces severe difficulties that limit the applicability of the Galerkin method (for a discussion, see Chapter II in Fletcher[36]). Most of these difficulties were eliminated by the introduction of the concept of *finite elements* by Clough.[37] He suggested representing a given domain as a collection of a

number of finite elements, subdomains of variable size and shape, which are interconnected in a discrete number of *nodes*. The solution of the partial differential equation is approximated in each element by a low-order polynomial in such a way that it is defined uniquely in terms of the (approximate) solution at the nodes. The global solution can then be written as a series of low-order piecewise polynomials with the coefficients of the series equal to the approximate solution at the nodes. The weighted residual or Ritz method is then applied using these low-order polynomials as trial and test functions, resulting in a system of algebraic or ordinary differential equations which can be solved using the well-known techniques. These methods are now bundled under the common denominator of *finite element method*.

The Galerkin finite element method

A first step in the construction of a finite element solution of a partial differential equation is the subdivision of the computational domain in a grid of finite elements, which are interconnected at a discrete number of common nodal points. The elements may be of arbitrary size and shape. A large number of element shapes have been suggested in the literature and are provided in most commercial finite element codes. Typical 2D and 3D element shapes are shown in Fig. 4.3, and a finite element grid of a food container is shown in Fig. 4.4.

The unknown solution is expressed in each element as a piecewise continuous polynomial in the space coordinates with the restrictions that (i) continuity between elements must be preserved and (ii) any arbitrary linear function could be represented.[38] In general, the unknown temperature field $T(x, y, z, t)$ can then be approximated by

$$\overline{T}(x, y, z, t) = \mathbf{N}^T(x, y, z)\mathbf{u}(t) \tag{4.35}$$

with $\mathbf{N}$ a vector of so-called shape functions and $\mathbf{u}^j$ a vector containing the temperatures at the nodes of the finite element grid. In general the approximate temperature field $\overline{T}$ is not identical to T, and when $\overline{T}$ is substituted in the diffusion equation, a non-zero residual ε is obtained:

$$r = \rho c \frac{\partial \overline{T}}{\partial t} - \nabla k \nabla \overline{T} - Q \tag{4.36}$$

This residual is subsequently orthogonalised with respect to the shape functions $\mathbf{N}$:

$$\int_\Omega \mathbf{N}\left(\rho c \frac{\partial \overline{T}}{\partial t} - \nabla k \nabla \overline{T} - Q\right) d\Omega \equiv 0 \tag{4.37}$$

It can be shown that after the application of Green's theorem and some matrix algebra a system of the form (4.34) is obtained.[38–40] $\mathbf{C}$ and $\mathbf{K}$ are now called the capacitance matrix and the stiffness matrix, respectively; $\mathbf{f}$ is the thermal load vector. The matrices $\mathbf{C}$, $\mathbf{K}$ and $\mathbf{f}$ are constructed element-wise. As in the case of the finite difference method, the system (4.34) is solved using traditional finite difference methods. Note that $\mathbf{K}$ and $\mathbf{C}$ are positive definite, symmetric and

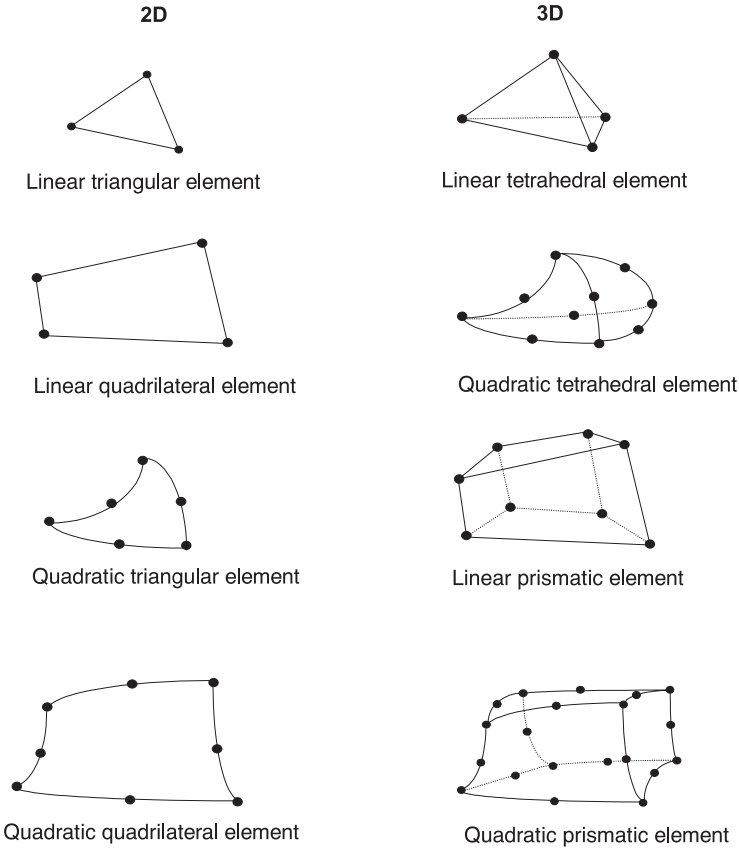


Fig. 4.3 Typical 2D and 3D finite element shapes.

banded. These very important features can be exploited advantageously to significantly reduce the computational effort and memory requirements.

Comparison of finite differences versus finite elements
Puri and Anantheswaran reviewed the use of the finite element method in food processing.[41] They listed the following key advantages of the finite element method compared with the finite difference method:

- spatial variation of material properties can be handled with relative ease
- irregular regions can be modelled with greater accuracy
- it is better suited to nonlinear problems
- element sizes can be easily varied
- spatial interpolation is much more meaningful
- problems with mixed boundary conditions are easier to handle.

Some of the disadvantages are:

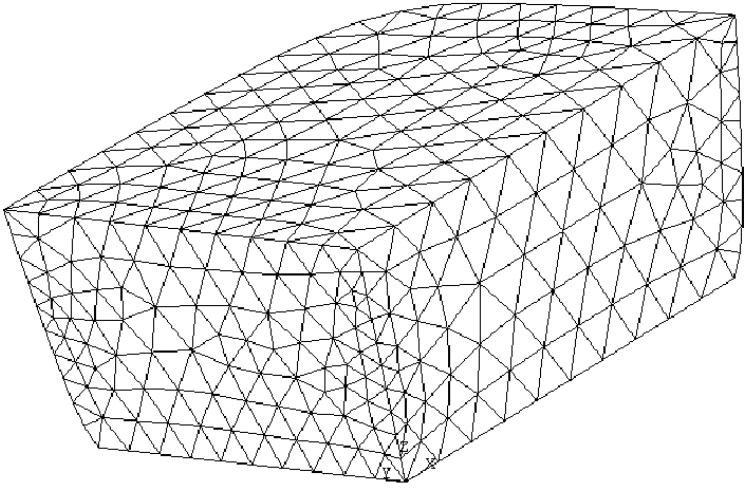


Fig. 4.4 3D finite element grid for a food container. Because of symmetry reasons only a quarter of the food container needs to be modelled.

- the discretised equations are usually mathematically more complex
- the method can require more computer resources for the same problem.

Because of the dramatic increase in computer power during the last decade, most conduction heat transfer problems can now be solved in a reasonable amount of time by means of the finite element method.

Applications

The finite element method was applied first to heat conduction problems by Zienkiewicz and Cheung, Visser, and Wilson and Nickell.[42–44] Early applications to thermal food processing problems were described by Comini *et al.*, Singh and Segerlind, and De Baerdemaeker *et al.*[45–47] Most commercial finite element codes are now based on the Galerkin finite element method.

Applications of the finite element method include the simulation of conduction heat transfer in foods with complicated geometrical shapes such as chicken legs,[47] a baby food jar,[48] broccoli stalks,[49] tomatoes,[50] and lasagna.[51] Special attention has been paid recently to stochastic finite element methods which were developed to take into account random variability of product and process parameters.[51–54]

4.5.4 The finite volume method

Principle

The finite volume method of discretisation is most widely used in commercial CFD (computational fluid dynamics) codes at the moment. It owes its popularity to the fact that it obeys the clear physical principle of conservation on the

discrete scale. The concepts of the method are easy to understand and have physical meaning.

The system of general conservation equations can be written in coordinate-free notation and integrated over a finite control volume V with surface A. Applying Gauss's theorem to obtain the surface integral terms, the equations have the following form, with ϕ the transported quantity:

$$\int_V \frac{\partial \rho\phi}{\partial t} dV + \int_A (\rho \mathbf{U}\phi).\mathbf{n} dA - \int_A (\Gamma \nabla \phi).\mathbf{n} dA = \int_V S_\phi dV \tag{4.38}$$

This equation states the conservation principle on a finite scale for all relevant quantities in the system when the surface integrals are the same for volumes sharing a boundary. Moreover, the finite volume form of the model becomes independent on the coordinate system. When the physical domain is subdivided into control volumes, a grid only defines the boundaries of the volumes. This is advantageous for modelling complex geometries.

The volume integrals are approximated in terms of the volume-centered value of ϕ The values at the volume faces are required for solving the surface integrals in equation (4.38). This requires interpolation in terms of volume-centered values. Some interpolation schemes may be highly accurate, but produce unbounded solutions when grids are too coarse. Others are unconditionally stable, but have a low accuracy and produce erroneous results called numerical or false diffusion. The reader is referred to the literature for a more elaborate discussion about the limits and benefits of different approximating formulas.[14, 15] The time discretisation in the control volume method is carried out using finite differences in the time domain, explained above.

Solution of the discretised equations

Discretisation results in the following set of equations, in matrix-vector notation:

$$\mathbf{A}\boldsymbol{\varphi} = \mathbf{Q} \tag{4.39}$$

where $\mathbf{A}$ is a square sparse matrix containing the coefficients resulting from the discretisation, $\boldsymbol{\varphi}$ is a vector containing the unknowns at the control volume centres and $\mathbf{Q}$ is a vector containing the variable-independent source terms. Equation (4.39) is still non-linear: the flow variables appear in the coefficients. An iterative method is therefore required in which the non-linear terms have to be linearised. The least expensive and most common approach is the Picard iteration. In this method coefficients are updated using the most recent solution of the system. This approach requires more iterations than Newton-like methods, which use a Taylor series expansion, but do not involve the computation of complex matrices and are found to be much more stable.

The solution of the linearised equations can be performed by direct methods, which are computationally very costly and generally do not benefit from the mathematical properties of the linear system. It is therefore advantageous to use an iterative method.

The iterative method should have certain properties in order to guarantee a valid solution. The main requirement for convergence of the solution is that the matrix **A** be diagonally dominant, which has been shown by Scarborough:[55]

$$\frac{\sum |A_{np}|}{|A_P|} \begin{cases} \leq 1 \text{ at all } P \\ < 1 \text{ at one } P \text{ at least} \end{cases} \tag{4.40}$$

where *np* are the neighbouring nodes of the node *P*. Several iterative solvers are available. A detailed discussion is given by Ferziger and Peric.[15]

To verify the validity of the mathematical solution, the solution change during the iterative procedure should be monitored. One can then stop the iteration process, based on a predefined convergence criterion and be assured of a convergent solution of the discretised equations. The convergence error $\boldsymbol{\varepsilon}_c^n$ can be defined as:[15]

$$\boldsymbol{\varepsilon}_c^n = \boldsymbol{\varphi} - \boldsymbol{\varphi}^n \tag{4.41}$$

where $\boldsymbol{\varphi}$ is the converged solution of equation (4.39) and $\boldsymbol{\varphi}^n$ is the approximate solution after *n* iterations. It is not possible to obtain $\boldsymbol{\varepsilon}^n$ directly and it is even hard to calculate a suitable estimation of the value. In practice, the residual r^n can be used to test for convergence:

$$\mathbf{A}\boldsymbol{\varphi}^n = \mathbf{Q} - \mathbf{r}^n \tag{4.42}$$

When the residual goes to zero, the convergence error will be forced to decrease as well, because:

$$\mathbf{A}\,\boldsymbol{\varepsilon}_c^n = \mathbf{r}^n \tag{4.43}$$

The reduction of the norm of the residual is a convergence criterion to stop the iterations. The residual should be reduced by three to five orders of magnitude. It may happen that the residual decreases much faster than the actual convergence error, in which case care should be taken and the iteration procedure continued.

Applications

The finite volume method is most popular for solving transport equations and is the core of most commercial so-called computational fluid dynamics packages. It combines the flexibility of the finite element method with the execution speed of the finite difference method. Early examples are the work of van Gerwen and van Oort, and Wang and Touber, modelling the airflow in cool rooms.[56–58] More elaborate works were published in the mid 1990s. Mirade and co-workers invested considerable validation efforts for the calculated velocity field in a meat chiller and reported an average accuracy of 20–40% for the calculated velocity magnitudes compared to measured ones.[59,60] Also modelling the airflow in a storage room, under working conditions with blockages, Hoang *et al.* found an agreement of 20–30%.[61] Further developments in storage rooms have been made by Chen and Li, Tassou and Xiang and Xu and Burfoot, who

studied the airflow in the room and the heat and mass transfer in porous stacks of agricultural produce.[62–65] Important studies have been performed on modelling the flow of liquid foodstuffs in cans.[66–69] Verboven *et al.* used CFD to predict the air flow and temperature fields inside a force convection oven.[70, 71]

4.5.5 Commercial software

Most commercial CFD codes for fluid flow analysis are available on UNIX as well as NT platforms. Parallel versions are often available as well. Some of the commercial codes dedicated to CFD analysis are described below. Some general-purpose numerical codes, like ANSYS (Swansee, USA) also include CFD features, but are mainly intended for structural and conduction heat transfer analysis.

CFX/TASCflow (http://www.aeat.com/cfx/)
CFX (formerly CFDS-FLOW3D) covers a group of commercial CFD codes. The main solvers are CFX-4, CFX-5, CFX-TASCflow and CFX-Radiation, which are individually supported and each has special features. The software is menu-structured and problems are defined by means of command files containing simple keywords. Additional input can be programmed by means of FORTRAN subroutines. CFX-4 uses block-structured finite volume meshes, while CFX-5 uses a fully unstructured mesh of tetrahedral control volumes. CFX-TASCflow allows unmatched hexahedral meshes, which makes it particularly suited for the analysis of rotating machinery.

Fluent/FIDAP/Polyflow (http://www.fluent.com/)
Fluent Inc. (Lebanon, NH, USA) is the world leader in the rapidly growing field of CFD software. Fluent recently acquired Polyflow, S.A., as well as Fluid Dynamics International Inc. (FDI), the developer of the FIDAP CFD software package. Fluent is a multi-purpose finite volume based menu-structured CFD code and allows either body-fitted structured meshes (FLUENT 4.5) or solution-adaptive unstructured meshes (FLUENT 5). User-defined subroutines are accessible through FORTRAN. FIDAP is a general-purpose finite element CFD code, which allows unstructured meshing in a straightforward way. A special feature of FIDAP is the capability to deal with fluid-structure interactions. POLYFLOW is a finite-element based CFD package for the analysis of polymer processing, including extrusion die design, blow moulding, and fibre spinning, as well as other materials processing applications.

PHOENICS (http://www.cham.co.uk/)
PHOENICS (Parabolic Hyperbolic Or Elliptic Numerical Integration Code Series) appeared in the 1980s as the first general-purpose code after integration of different problem-specific codes. It claims to be the most widely used CFD code in the world, partly because of the availability of a shareware version. PHOENICS is available on MS-DOS, WINDOWS and LINUX in addition to

NT and UNIX platforms. Mesh generation capabilities are limited as only multi-block structured grids are possible. The PHOENICS structure allows users to access more than 50% of the source code (unlike other packages) and supplies entry to a FORTRAN library with new physical or numerical content. This has resulted in, for example, a wide range of turbulence models (mixing-length models, $k-\varepsilon$ models, stress models and more exotic types). The open-source coding requires some programming skills and experience from the user with the PHOENICS Input Language.

STAR-CD (http://www.cd.co.uk/)
STAR-CD is developed by Computational Fluid Dynamics Ltd, which was originally dedicated to the automotive industry but today aims at all sectors of industry. STAR-CD is a multi-purpose control volume based CFD code. STAR-CD provides a LINUX version in addition to the common UNIX and NT platforms. The code STAR-HPC has been developed for parallel machines. It is capable of dealing with unstructured meshes containing cell shapes ranging from tetrahedra and prisms to general polyhedra in any hybrid unstructured mesh with arbitrary interfaces.

4.6 Conclusions

Many software tools are now readily available for solving realistic heat and mass transfer problems. The reliability of the numerical solution, however, largely depends on the availability of suitable thermophysical properties and the complexity of the governing models. The numerical solution of convective transport problems described by Navier-Stokes equations remains a difficult task, particularly when turbulence is involved. The empirical constants involved in the most popular turbulence models necessitate a careful validation of the obtained results.

4.7 Acknowledgements

The authors wish to thank the European Union (projects FAIR-CT96-1192 and INCO IC15 CT98 0912) and the Flemish Minister for Science and Technology (COF 97-08) for financial support.

4.8 References

1. FOURIER J, 1822. *Théorie Analytique de la Chaleur*. Reprinted in *Oevres de Fourier – Tome premier*, Darboux M G, ed., Paris, France, Firmin Didot, père et fils, 1888.
2. ANON., *ASHRAE Handbook – Fundamentals*, Atlanta, American Society of Heating, Refrigeration and Air-Conditioning Engineers, 1993.

3. DATTA A K, PROSETYA H M, HU W, 'Mathematical modeling of batch heating of liquids in a microwave cavity', in Yano, T., Matsuno, R., Nakamura, K. (eds), *Developments in Food Engineering*, London, Glasgow, Weinheim, Blackie Academic and Professional, 1994, 325.
4. OHLSSON T, 'In-flow microwave heating of pumpable foods', in Yano, T., Matsuno, R., Nakamura, K. (eds), *Developments in Food Engineering*, London, Glasgow, Weinheim, Blackie Academic and Professional, 1994.
5. ZHANG L, LIU S, PAIN J-P, FRYER P J, 'Heat transfer and flow in solid–liquid food mixtures', *Food Engineering in a Computer Climate*, New York, Philadelphia, Hemisphere Publishing Company, London, 1992, 79.
6. DE ALWIS, FRYER P J, 'A finite element analysis of heat generation and transfer during ohmic heating of food', *Chemical Engineering Science*, 1990 **45** 1547–60.
7. BIRD R B, STEWART W E, LIGHTFOOT E N, *Transport Phenomena*, New York, John Wiley and Sons, 1960.
8. NI H, DATTA A K, 'Moisture, oil and energy transport during deep-fat frying of food materials', *Trans IchemE*, 1999 **77**(C) 194–204.
9. ZAUDERER E, *Partial Differential Equations of Applied Mathematics*, Singapore, New York, Chichester, John Wiley and Sons, 1989.
10. KEVORKIAN J, *Partial Differential Equations – Analytical Solution Techniques*, Pacific Grove, California, USA, Wadsworth, Brooks and Cole, 1990.
11. CARSLAW H S, JAEGER J C, *Conduction Heat Transfer in Solids*, 2nd ed., London, Oxford University Press, 1959.
12. MILLS A F, *Heat Transfer*, Homewood, Irwin, 1992.
13. CROCHET M J, DAVIES A R, WALTERS K, *Numerical Simulation of non-Newtonian Flow*, Rheology Series 1, Amsterdam, Elsevier, 1984.
14. VERSTEEG H K, MALALASEKERA W, *An Introduction to Computational Fluid Dynamics: The Finite Volume Method*, Harlow, Longman Scientific & Technical, 1995.
15. FERZIGER J H, PERIC M, *Computational Methods for Fluid Dynamics*, Heidelberg, Springer-Verlag, 1996.
16. MOIN P, KIM J, 'Tackling turbulence with supercomputers', *Scientific American*, 1997 **276**(1) 46–52.
17. JONES W P, LAUNDER B E, 'The prediction of laminarisation with two-equation model of turbulence', *Int J of Heat Mass Transf*, 1972 **15** 301–13.
18. YAKHOT V, ORSZAG S A, THANGAM S, GATSKI T B, SPEZIALE C G, 'Development of turbulence models for shear flows by a double expansion technique', *Phys Fluids A*, 1992 **4**(7) 1510–20.
19. LAUNDER B E, SHARMA B I, 'Application of the energy dissipation model of turbulence to the calculation of flow near a spinning disc', *Lett Heat and Mass Transf*, 1974 **1** 131–8.
20. VERBOVEN P, SCHEERLINCK N, DE BAERDEMAEKER J, NICOLAÏ B M, 'Performance of different turbulence models for the calculation of the surface heat transfer coefficient', *Conference/Workshop on Thermal*

Properties and Behaviour of Foods during Production, Storage and Distribution, Prague, Food Research Institute, 1997.

21. YAP C, *Turbulent heat and momentum transfer in recirculating and impinging flows*, Ph.D. Thesis, Manchester, Faculty of Science and Technology, 1987.
22. HIRSCH C, *Numerical Computation of Internal and External Flows. Vol. 1: Fundamentals of Numerical Discretization*, Chichester, John Wiley & Sons, 1991.
23. LUIKOV A V, *Heat and Mass Transfer in Capillary-Porous Bodies*, Oxford, London, Edinburgh, Pergamon Press, 1966.
24. IRUDAYARAJ J, HAGHIGHI K, STROSHINE R L, 'Finite Element Analysis of Drying with Application to Cereal Grains', *J Agri Eng Research*, 1992 **53** 209–29.
25. FRANCA A S, HAGHIGHI K, 'An Adaptive Approach to Finite Element Modeling of Heat Transfer in Food Processing', *1994 International Winter Meeting – Atlanta, Georgia*, Atlanta, American Society of Agricultural Engineers, 1994.
26. ZHOU L, PURI V M, ANATHESWARAN R C, DOORES S, SIGMAN-GRANT M, 'Effect of Temperature Gradient on Moisture Migration During Microwave Heating', *1993 International Summer Meeting*, Spokane, Society of Agricultural Engineers & Canadian Society of Agricultural Engineers, 1993.
27. LITCHFIELD J B, OKOS M R, 'Moisture Diffusivity in Pasta During Drying', *J Food Eng*, 1992 **17** 117–42.
28. HUSAIN A, CHEN C S, CLAYTON J T, 'Simultaneous Heat and Mass Diffusion in Biological Materials', *J Agri Eng Research*, 1973 **18** 343–54.
29. TEIXEIRA A A, DIXON J R, ZAHRADNIK J W, ZINSMEISTER G E, 'Computer optimization of nutrient retention in thermal processing of conduction-heated foods', *Food Technology*, 1969 **23** 137.
30. TUCKER G S, CLARK P, 'Modelling the cooling phase of heat sterilization processes, using heat transfer coefficients', *International Journal of Food Science and Technology*, 1990 **25** 668–81.
31. MANSON J E, ZAHRADNIK J W, STUMBO C R, 'Evaluation of lethality and nutrient retentions of conduction-heating foods in rectangular containers', *Food Technology*, 1970 **24** 109–13.
32. RITZ W, 'Über eine neue Methode zur Lösung gewissen Variationsprobleme der mathematischen Physik', *J. Reine Angew. Math.*, 1909 **135** 1–61.
33. WHITE R E, *An introduction to the finite element method with applications to nonlinear problems*, New York, Chichester, John Wiley & Sons, 1985.
34. BICKLEY W G, *Phil Mag* 1941 **32**, 50–66.
35. CRANDALL S H, *Engineering Analysis*, New York, St. Louis, San Francisco, McGraw-Hill, 1956.
36. FLETCHER C A, *Computational Galerkin Methods*, New York, Berlin, Springer-Verlag, 1984.

37. CLOUGH R W, 'The finite element method in plane stress analysis', *Proceedings of the 2nd ASCE Conference on Electronic Computation*, Pittsburgh, 1960.
38. ZIENKIEWICZ O C, *The Finite Element Method*, Maidenhead, New York, St. Louis, San Francisco, McGraw-Hill Book Company, 1977.
39. CONNOR J J, BREBBIA C A, *Finite Element Techniques for Fluid Flow*, London, Boston, Newnes-Butterworths, 1976.
40. SEGERLIND L, *Applied Finite Element Analysis*, New York, Chichester, John Wiley & Sons, 1984.
41. PURI V M, ANANTHESWARAN R C, 'The finite element method in food processing: a review', *J Food Eng*, 1993 **19** 247–74.
42. ZIENKIEWICS O C, CHEUNG Y K, 'Finite elements in the solution of field problems', *The Engineer*, 1965 507–10.
43. VISSER W, 'A finite element method for determination of non-stationary temperature distribution and thermal deformations', *Proceedings of the Conference on Matrix Methods in Structural Mechanics*, Air Force Institute of Technology, Wright Patterson A.F. Base, Dayton, Ohio, USA, 1965.
44. WILSON E L, NICKELL R E, 'Application of the finite element method to heat conduction analysis', *Nuclear Engineering Design*, 1966 **4** 276–86.
45. COMINI G, DEL GUIDICE S, LEWIS R W, ZIENKIEWICZ O L, 'Finite element solution of non-linear heat conduction problems with special emphasis on phase changes', *International Journal of Numerical Methods in Engineering*, 1974 **8** 613–24.
46. SINGH R P, SEGERLIND L J, 'The finite element method in food engineering', ASAE paper No. 74-6015, St. Joseph, American Society of Agricultural Engineers, 1974.
47. DE BAERDEMAEKER J, SINGH R P, SEGERLIND L J, 'Uses of the finite element method in food processing', *J Food Process Eng*, 1977 **45** 1547–59.
48. NAVEH D, KOPELMANN I J, PFLUG I J, 'The finite element method in thermal processing of foods', *J Food Sci*, 1983 **48** 1086–93.
49. JIANG H, THOMPSON D R, MOREY R V, Finite element model of temperature distribution in broccoli stalks during forced-air precooling. *Transactions of the ASAE*, 1987 30, 1473–7.
50. PAN J C, BHOWMICK S R, 'The finite element analysis of transient heat transfer in fresh tomatoes during cooling', *Trans ASAE*, 1991 **34** 972–6.
51. NICOLAÏ B M, VAN DER BROECK M, SCHELLEKENS M, DE ROECK G, MARTENS T, DE BAERDEMAEKER J, 'Finite element analysis of heat conduction in lasagna during thermal processing', *International Journal of Food Science and Technology*, 1995 **30**(3) 347–64.
52. NICOLAÏ B M, DE BAERDEMAEKER J, 'A variance propagation algorithm for the computation of heat conduction under stochastic conditions'. *International Journal for Heat and Mass Transfer*, 1999 **42** 1513–20.
53. NICOLAÏ B M, VERBOVEN P, SCHEERLINCK N, DE BAERDEMAEKER J, 'Numerical analysis of the propagation of random parameter fluctuations

in time and space during thermal food processes', *Journal of Food Engineering*, 1999 **38** 259–78.

54. NICOLAÏ B M, VERLINDEN B, BEUSELINCK A, JANCSOK P, QUENON V, SCHEERLINCK N, VERBOVEN P, DE BAERDEMAEKER J, 'Propagation of stochastic temperature fluctuations in refrigerated fruits', *International Journal of Refrigeration*, 1999 **22**(2) 81–90.
55. SCARBOROUGH J B, *Numerical Mathematical Analysis*, 4th ed., Baltimore, University Press, 1958.
56. VAN GERWEN R J M, VAN OORT H, 'The use of fluid dynamics simulation models in cold store design', *International Congress of Refrigeration*, Bristol, International Institute of Refrigeration, 1989.
57. VAN GERWEN R J M, VAN OORT H, 'Optimization of cold store design using fluid dynamics', *International Congress of Refrigeration*, Dresden, International Institute of Refrigeration, 1990.
58. WANG H, TOUBER S, 'Distributed dynamic modelling of a refrigerated room', *Int J Refr*, 1990 **13** 214–22.
59. MIRADE P S, DAUDIN J D, ARNAUD G, 'Simulation en deux dimensions de l'aéraulique de deux tunnels de réfrigération des viandes', *Int J Refr*, 1995 **18**(6) 403–12.
60. MIRADE P S, DAUDIN J D, 'Numerical simulation and validation of the air velocity field in a meat chiller', *Int J Appl Sci Comp*, 1998 **5**(1) 11–24.
61. HOANG M L, VERBOVEN P, DE BAERDEMAEKER J, NICOLAÏ B M, 'Numerical analysis of the air flow in a cool store by means of computational fluid dynamics', *Int J Refr*, 2000 **23**(2) 127–40.
62. CHEN L, LI Y, 'Turbulent force flow in an airblast storage room', *ROOMVENT'98: 6th International Conference on Air Distribution in Rooms, Proceedings Volume II*, Stockholm, Pricor, 451–8.
63. TASSOU S A, XIANG W, 'Modelling the environment within a wet air-cooled vegetable store', *J Food Eng*, 1998 **38** 169–87.
64. XU Y, BURFOOT D, 'Simulating the bulk storage of food stuffs', *J Food Eng*, 1999 **39** 23–9.
65. XU Y, BURFOOT D, 'Predicting condensation in bulks of food stuffs' *J Food Eng*, 1999 **40** 121–7.
66. ENGELMANN M S, SANI R L, 'Finite element simulation of an inpackage pasteurisation process', *Num Heat Transfer*, 1983 **6** 41–54.
67. DATTA A K, TEIXEIRA A A, 'Numerically predicted transient temperature and velocity profiles during natural convection of canned liquid foods', *J Food Sci*, 1988 **53**(1) 191–5.
68. KUMAR A, BHATTACHARYA M, BLAYLOCK J, 'Numerical simulation of natural convection heating of canned thick viscous liquid food products', *J Food Sci*, 1990 **55** 392–6.
69. GHANI A G A, FARID M M, CHEN X D, RICHARDS P, 'Numerical simulation of natural convection heating of canned food by computational fluid dynamics', *J Food Eng*, 1999 **41**, 55–64.
70. VERBOVEN P, SCHEERLINCK N, DE BAERDEMAEKER J, NICOLAÏ B M,

'Computational fluid dynamics modelling and validation of the isothermal airflow in a forced convection oven', *J Food Eng*, 2000 **43**(1) 41–53.

71. VERBOVEN P, SCHEERLINCK N, DE BAERDEMAEKER J, NICOLAÏ B M, 'Computational fluid dynamics modelling and validation of the temperature distribution in a forced convection oven', *J Food Eng*, 2000 **43**(2) 61–73.

5

Combined discrete/continuous modelling

R. W. Sierenberg

5.1 Introduction: the big gap

All modellers claim that they model real world systems most accurately in order to predict the status of some specified aspects of that system. This does not imply that modellers create their models according to a common methodology.

If, for instance, the system concerns apples and the aspects are related to quality (colour, juiciness) then, for sure the model is of the type *continuous*. However, if these aspects concern logistics (availability in a supermarket, the use of storage facilities) then the model will be of the type *discrete*. More than 90% of all models are either continuous or discrete. Moreover most modellers can be classified that way and for a long time the two groups were completely separated. That gap has nothing to do with religion but everything with the way they handle parallelism.

5.1.1 How do continuous modellers deal with parallelism?

The answer is most simple: *They do not.*

A continuous model consists of:

- *Endogenous variables*, describing inside processes needed to explain the quality aspects of the apple to be studied; for instance enzymes, concentrations and extractions.
- *Exogenous variables*, describing the surroundings. Such as the change in temperature and humidity during an experiment.
- A set of *functions and/or differential* equations, stipulating the *relations* between these variables, containing
- *Parameters* to be estimated using series of simulations of well registered experiments.

The variables are time dependent, related to each other and supposed to operate simultaneously. Simultaneous calculation of the changes of the endogenous variables is troublesome; therefore, these calculations are performed *one by one* over small time steps.

Evaluation of a differential equation of one of these variables normally involves other variables, which values *are supposed to be known*. That can be true for the exogenous variables, however, the values of the endogenous variables should be estimated. Modern techniques will perform these estimations extremely well, maintaining almost any accuracy stipulated by the modeller, *but only if all variables behave continuously*. From the moment computers could be used to run continuous models, the modellers became more and more happy with this situation because the improving techniques offered more and more splendid and user friendly tools.

The bad thing is that, due to the demand for continuous variables only, these continuous modellers became prisoners in their own continuous field.

5.1.2 How do discrete modellers deal with parallelism?

They cannot use the trick with small time steps because discrete varying values cannot be estimated from previous values. These modellers had to solve the problem of parallelism by themselves, until software producers started offering tools for it. There is only one way: all parallel processes in a system must be replaced by just *one* process in the model being *the* process of successive evaluations of discrete time events. I have to demonstrate this *event description method* first, just to prove that it only works for trivial problems.

Let us consider the system of the counters in a supermarket in order to find a relation between the number of customers waiting to pay for their shopping and the number of active counters. The first step is to define the *state variables*. In this case the state variables are the numbers of waiting customers N_WAIT[I] for each counter I and a logical variable BUSY[I] to express whether the counter is busy or not. The second step is to find the events *depending on time* that cause the change of at least one of these state variables. In our case there will be two types:

1. The arrival of a customer wanting to pay.
2. A leaving customer.

Note that the event ‘a counter starts working’ is not a time event, because it is always a result of one of the two mentioned.

The moment of (future) occurrence of a time event is called its *eventtime*. The eventtimes concerning leaving customers are kept in the variables LEAVE_TIME[I] where I denotes the counter from which a customer will leave. LEAVE_TIME[I] will be 10^{10} if BUSY[I] = false. NEXT_ARRIVALTIME will be used to keep the eventtime of the next arriving customer.

To complete the model we need the distributions IAT (interarrivaltime between successive customers) and PAYTIME. The procedure SAMPLE_FROM will be

used to get samples from these distributions. The following program explains the replacement trick.

```
SEQUENTIAL MECHANISM: (time advance calculation)
Let LEAVE_TIME[IO] BE THE SMALLEST OF ALL LEAVE_TIME[I].
If NEXT_ARRIVAL < LEAVE_TIME[IO]
         NOW ← NEXT_ARRIVAL
         goto ARRIVAL EVENT
end
NOW ← LEAVE_TIME[IO]
goto DEPARTURE EVENT
ARRIVAL EVENT:
Let N_WAIT[IO] BE THE SMALLEST OF ALL N_WAIT[I]
N_WAIT[IO] ← N_WAIT[IO] + 1
NEXT_ARRIVAL ← NOW + SAMPLE_FROM IAT
goto FINISHING

DEPARTURE EVENT:
BUSY[IO] ← FALSE
LEAVE_TIME[IO] ← 10^10
goto FINISHING

FINISHING:
IF BUSY[IO] = FALSE & N_WAIT[IO] > 0
         N_WAIT[IO] ← N_WAIT[IO]–1
         BUSY[IO] ← TRUE
         LEAVETIME[IO] ← NOW + SAMPLE FROM PAYTIME
end
repeat from SEQUENTIAL MECHANISM
```

It looks rather simple. Why doesn't it work? You have to experience it yourself. From here to the end of this paragraph, you have to manage a milk factory. Your factory has two major production departments:

1. The cows' milk receiving department where the milk is temporarily stored in tanks (not longer than a day) to be centrifuged (to get the cream out) and pasteurised to become half-products (product milk and consumers' milk) for the next department. The production in this department is *pushed* by the daily, almost equal, production of the same set of cows.
2. The ready product department where, in several production lines, the half-products are transformed in different kinds of milk, buttermilk, yoghurts and custards. This production is *pulled* by the behaviour of the consumers, filling their refrigerators on Friday to survive the weekend. This is done so successfully that they hardly need anything before Tuesday.

All the milk you deliver to the supermarket chain should be fresh every day, which does not allow buffering in half products. Therefore you buffer in the buttermilk, yoghurt and custard products having longer production times and

later expiry dates. Consequently, your production schedule will be different for each day of the week. Nothing special so far. However, next month one of your four yoghurt production tanks must be replaced by a new one, which will take exactly four days. The standard solution is sending the surplus of product milk to other milk factories. Of course, you want to find out if the problem can be solved without these intra transports. For studying all alternatives, you need a model.

What was the first step? Specifying the state variables! Say you survive that, what is the second step? Finding the time events. Just imagine what that means! Long before you have to help the modeller with the second step, you will conclude that solving the yoghurt problem without a model will be easier.

In conclusion: modelling parallel discrete processes needs special tools too, which are available nowadays as general modelling languages or as packages for special groups of systems. These modelling languages solve the problem of parallel processes by monitoring the model created by the user.

The purpose of the tools for discrete modellers is therefore different in comparison with the tools for continuous modellers. So, modellers are separated from each other by the use of their tools.

The model language PROSIM (PROSIM bv, Zoetermeer, The Netherlands) and a few more are equipped with tools for continuous processes, allowing combined discrete/continuous modelling. Theoretically, the gap can be closed, however most continuous modellers are so tied to their tools that they rather avoid discrete processes than face parallelism. If we want to control systems, getting fresh food on the supermarket shelves in an optimal condition and in acceptable amounts, we need the cooperation of all specialists involved. Parallelism is too important to be denied. The next paragraph will demonstrate why.

5.2 The power of parallel processes

We, human beings, are not able to control more than five interacting processes. We already have trouble in finding the number of people in a room if there are more than five. We need to count. Even counting becomes troublesome if these people are walking around. Complexity is mostly due to parallelism.

Let us enter the world of parallelism carefully considering a simple road map. There are junctions connected by road sections forming a *network*. Each section has a direction and a length; it goes from one junction to another. In case there can be traffic in both directions we use two sections.

At the entrance of each section, there is a dog. If a dog is activated it will carry out the following process:

1. Run to the junction at the other end of the section with the speed of electricity.
2. Activate each dog that can be found there.
3. Return to the entrance of the section with the much slower speed of sound.

Furthermore, each dog will remember how it was activated (which dog or otherwise). (The differences in speed prevent a dog being activated twice in one experiment.)

Now, just choose one of the junctions and activate all the dogs at that junction. Try to figure out what happens before reading the next sentence.

Within a split second, all the dogs will be back in their starting positions. You can choose another junction randomly and ask some dog there: 'Who activated you?' That dog will point to another dog. The questioning can be repeated until some dog answers: 'You did'. At that moment, the chain of dogs questioned stipulates the shortest route from the first junction to the second one.

There are three remarkable issues:

1. The dogs only perform a most simple task. There is no further logic, ruling or whatsoever. The fact that the dogs operate *simultaneously* solves the shortest route problem!
2. There are a lot of shortest-route-solving methods not based on parallelism. The efficiency of all these methods were compared ages ago, by counting the average number of calculations needed for a solution. Neglecting the effort to simulate parallelism, this 'method' is the most efficient one.
3. The method is robust. Even if several dogs are not willing to run, you will still get a fair result.

The next step is to forget about the dogs and to make the sections alive instead. The sections will be called neurones now. Each neurone has a head and a tail and at the junctions, heads and tails are joined together. In this way, our road map is transformed into a very simple model of a neural network. Instead of being activated, a neurone can be *triggered*. If a neurone is triggered it will *fire*. In our simple model, it means that the neurone triggers the neurones which tails are connected to its head. After firing, a neurone will be insensitive for triggering for some time. In fact, the neurones in our new model have almost the same process description as the dogs in the road map, but now it makes sense. If we trigger a neurone from the outside, all other accessible neurones in the network are informed where it happened in a split second. Again: most efficiently, no special control or intelligence, and robust. Our body is full of neural networks. These networks have special tasks, causing for instance, that you withdraw your hand immediately if you touch something very hot. These special tasks are realised by differences in the process descriptions of the neurones and their threshold values for being triggered. Conclusion: *the most fundamental activities in all that lives are based on parallelism.*

In all cases, it turns out that the network is far more 'intelligent' than its components. The next step will be that the components do not need to be physically connected. A weaker form of being connected is being *related*. Without that physical bond, the possibilities for parallelism are manifold, as are the ways for 'triggering'. That can be just smell and/or sound and then we get *swarm intelligence* as known from ants and bees colonies.

The final step makes the network into a community, where the components are human beings related with each other in the most complex ways, triggering their relations using money, power, love or friendship. The community can reach almost everything based on simultaneous performance of activities.

This paragraph started with an example showing how the shortest-route-problem could be solved using parallelism. Nowadays, the power of parallelism is more and more discerned, leading to the construction of computer configurations not being based on the almighty central processor.

More people discover the power of parallelism, creating software tools for solving combinatorial problems based on simultaneous activities.

Modelling languages like PROSIM can be helpful this way, for instance to find an efficient solution for the problem of parameter estimation.

Suppose, we have a model correctly describing the processes causing juiciness (J) of an apple. This model contains the parameters P_1, P_2, ..., P_6 which values are unknown so far. The model is valid if we can find correct values for the parameters. To find these values we need data from one or more experiments that can be compared against simulation by the model. During each experiment the values for the exogenous variables are known and the values for the endogenous variables are measured regularly during the experiments.

A simple way to estimate the parameters P_1, P_2, ..., P_6 could be by trying to find their values one by one. Thus, repeat the simulation of the experiments, varying one of the parameter values as long as the difference between measured and calculated juiciness decreases. Then start with the next parameter. Due to the correlations between the parameters, coming from their functional relationships, normally the problem still cannot be solved by treating all six parameters this way.

Fortunately, Levenberg-Marquard provides a much better method. This method can be used if all partial derivatives $\partial^2 J / \partial P_i \partial P_j$ are known for all combinations of i, j. The problem is how to find these derivatives when the parameters are directly or indirectly *coefficients* of differential equations, which they normally are. That problem can be solved using parallelism by making seven copies C_0, C_1, ..., C_6 of the model and have them all run *simultaneously* to simulate the experiments. The models are all equal except for one parameter value. Model copy C_0 has the best parameter estimates found so far, say E_1, ..., E_6. In model copy C_1, the parameter P_1 has the value $E_1 + \Delta$. In model copy C_2 the value of P_2 equals $E_2 + \Delta$, and so on. Each time the seven models reach the time of a measurement (all simultaneously of course), the differences in simulated juiciness shown with each other and with the value measured give estimates of the partial derivatives to be merged with the results obtained from the previous measurement. At the end of the run the method of Levenberg-Marquard will deliver new estimates for the parameters and the whole method can be repeated until the best fit is found.

5.3 The 'world view' of system theory

> Dear Descartes,
> When you are thinking, you have to formulate your thoughts in words, sentences, abstracts and conceptions as you learnt from the community that raised you. So, your thinking proves the existence and even more the functioning of that community. That community is timeless and if your spirit survived your dead body, you may still be a part of it.

Starting in 1975 Bernard Zeigler tried to convince modellers all over the world that each system can be considered as a set of related components. These components simultaneously perform processes but should be described one by one because humans cannot do better. According to their relationships the components form a network. Because the relations may change and components may die or be created, the structure of the network is time dependent. That concept should be the base of all modelling. Nobody ever denied it, only a few followed him.

Nowadays, we know that nature is composed in this way and we can only conclude that 'can be considered' is an understatement that should be replaced by 'is'. The concept forms the basis of what is called the 'system theory' being lectured in most universities of technology.

The next example demonstrates some fundamental concepts of this theory. It concerns the supermarket again but with a more interesting goal. We need a model to investigate the effect of logistics rules on missing sales and overload. Overload may cause *overcode* being the amounts of an article that must be removed from the shelves because the last day for selling has expired.

The supermarket is a network of related components. The first modelling step concerns the question: '*Which of these components do we need in the model?*' It is obvious that we need articles. Most supermarkets deal with 3000 to 6000 different articles. Not all of them will be interesting to be used in each run. Suppose, we are just interested in articles that can cause overcode. This will reduce the scope of the model to about 600 articles. In modelling terms the chosen articles form a *class of components*. That means: the way to *specify* an article will rule for all of them. Components of the same class differ from each other because their specifications have other *values*.

Do we need customers? No, we only need a *class of sales*. Introducing customers would create the need for a shopping list for each of them. That may be interesting for marketing purposes but will be a burden to our goal. Instead of a class of customers we introduce one single component, the *sales generator*, to create all the sales at the beginning of each day as explained below.

Sales will remove items from the shelves but not overcode. That will be done by the single component *chief* at the end of each day. The sales and the chief will cause empty shelves, so we need some more components taking care of stockpiling new goods. The single component chief (or his computer) will cause distribution centres to deliver supplies by trucks. Because our model is meant to inspect the logistics rules and not to simulate a distribution centre, distribution

centres are assumed to work perfectly. Only the trucks are modelled delivering all that is ordered at schedule.

As shown, single components and classes of components are chosen according to the goal of the investigation. Although this is called 'the first step', it does not mean that this step should be completed before starting the second one. Generally, modelling is a cyclic activity.

The second step concerns the specifications of the components, by linking *attributes* to each of them. As all components of one class have similar specifications it is sufficient to define attributes for each class and for each single component. This example shows how this can be done for the articles. Each article will at least have the following attributes:

code:	containing a unique series of figures identifying the article.
in_units:	having the value TRUE if the article is sold in units instead of weight
n_sales:	the number of sales of the article during the current day
m_quantum:	the mean number of units in a sale if **in_units** and the mean weight in a sale differ
missed:	the amount (units or weight) of missed sales so far because the article was out of stock
overcode:	units or weight
stock:	the available amount (units or weight)
sold:	the amount sold on the current day so far
shelfcode:	the number of days (including the delivery day) the article can be sold
out[30]:	out[1] the amount that must removed at the end of the current day when not sold; out[2] the amount to be removed tomorrow, etc.
overload:	the amount of stock still available at the moment of delivery

These attributes give a rough idea about the meaning of the second step and illustrate some further features outlined below. A complete supermarket model needs at least hundred attributes for each article. Do not even try to get all attributes together at once. Just start with enough components and classes to get something working, to let the sales generator create sales and to let the chief create orders to be delivered by trucks. That will allow the development of a working model to start with. Afterwards, you will find out:

- that more sophisticated logistics rules will ask for more attributes,
- that articles should be grouped according to delivery schemes
- that articles should be grouped together to solve problems about substitution,
- that articles are related because they only differ in packing,
- that articles need more attributes describing their behaviour during sales actions,
- that sales of articles can depend on weather conditions,
- .....

As you see, one hundred attributes is not extreme! However, it also shows another aspect; the amount of data involved with running the model will need special attention. At the start of every simulated day, the number of sales of each article must be obtained from outside. Furthermore, every article will produce time series for output purposes. As a consequence, the model has to interact with data-bases. Modelling tools should be equipped for that purpose. For instance, the amount of software in PROSIM needed for the data handling concepts is much bigger than the amount of software needed for all of the continuous concepts.

The third step in creating a model concerns the process descriptions of the single components and a process description for each class component containing 'living' components. Components not performing activities (including decision-making) on their own do not need a process description because their processes will be caused by the activities of the living components. As the articles do nothing on their own behalf they do not need a process description.

As an example of a process description we use the sales generator. This component has to deal with articles and sales. So we need to specify some attributes of a sale first:

shop_art: referring to the article being sold
quantum: the number of items or the weight
sale_time: the moment of the sale

That will be enough to explain the description of the process of the sales generator. This fictive component represents all shopping customers. At first sight we may try to model this component as follows:

```
Wait until the supermarket opens this day
While the supermarket is open
   wait some time
   create a sale
end
```

The first activity is simple. In the next one we have to specify the value of 'some time' by sampling from some distribution function of 'inter sales time'. The problem is that such a distribution function does not exist because in some way it should express the variation in the arrival density of customers (see Fig. 5.1) belonging to that day of the week. If we try to use a distribution while varying the mean during the day, we are guilty of committing sin number four of the pitfalls (see section 5.5) by introducing a false variance in the number of sales. If we still continue we commit sin number four again in the next statement when selecting the SHOP_ART of the sale by introducing a false variance in the number of sales of that article.

For realistic results, the creation of sales must be modelled more carefully. The sales generator will start its activities each day at time 00.00. First of all, the number of sales (N_SALES) of each article for that day must be obtained from an external data-base. That may be historical data or data based on statistics, the

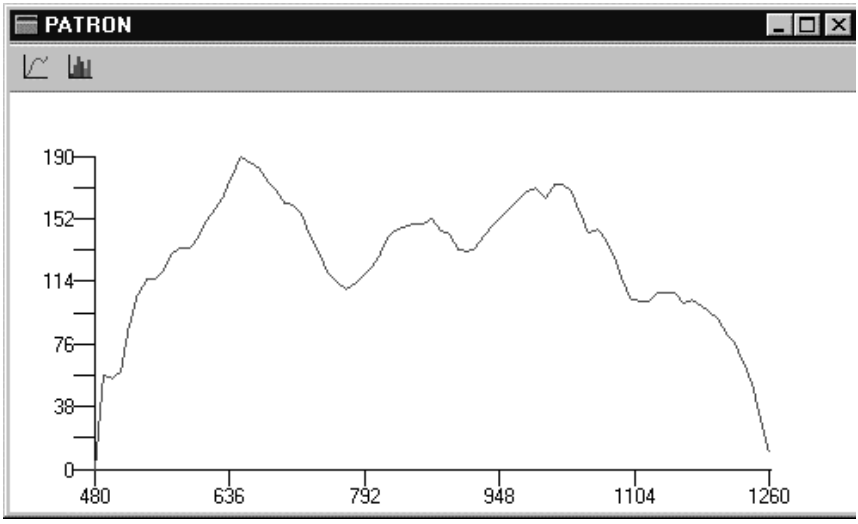


Fig. 5.1 Average number of sales per 10 minutes during the opening hours (08–21) of a series of Fridays.

model has nothing to do with. Besides the N_SALES, each article has a M_QUANTUM, representing the average amount (items or weight) in a sale. It is not realistic to expect that this number is known for each day of the week. That could be possible for some 'fast runners' only. Therefore, this figure is supposed to be independent of the day of the week, because with the data available we cannot do better. The best way to obtain the number of items in a sale is by keeping the variance as low as possible. Let, for instance, the M_QUANTUM of a milk product be 1.7 packs. R is a random number equally distributed between 0 and 1. If R is smaller than 0.7 the QUANTUM of the sale will be 2 packs and 1 otherwise.

The sales generator will use the function of arrival densities as a distribution function. Each sample from this distribution represents the time during the day of a sale. The process description of the sales generator could be as follows:

- obtain the N_SALES for each component of the class article
- for each article
  for count ← 1 to N_SALES
    create a sale
    SHOP_ART ← THIS ARTICLE
    QUANTUM ← M_QUANTUM if IN_UNITS = false
    calculate QUANTUM as shown above otherwise
    SALE_TIME ← sample from arrival_density
    activate this sale at SALE_TIME this day
  end
 end
- wait until next morning
- repeat from top

When a sale comes alive it will update the attributes of the article SHOP_ART points to, according to the sufficiency of the value of STOCK OF SHOP_ART.

On arrival, a truck will update the attributes of the articles it delivers, not forgetting to increase OUT[SHELFCODE] with the delivered amount. At the end of each day the chief will compare the value of OUT[1] with SOLD. If OUT[1] is greater than SOLD, the amount OUT[1]–SOLD must be subtracted from stock and added to the overcode. Otherwise the amount SOLD–OUT[1] must be subtracted from OUT[2]. After that the chief will set the value of SOLD to 0 and OUT[1] takes the place of OUT[2], OUT[2] that of OUT[3], etc. In this way the modeller describes the processes of the living components. These descriptions are added to the model and the software takes care that all processes are simulated simultaneously.

Be aware that the concepts of system theory do not burden the physical constraints of computers. It just tells us to describe the behaviour of components one by one, regardless whether our computers are able to or not. That is the problem for the software producer; the modeller should be free to use the concepts.

5.4 Combined modelling

If we want to model real world systems concerning food processing, manufacturing or logistic elements, the possibility of handling multiple components simultaneously is far more important than bothering about whether processes are continuous or not. In the fundamentals of the system theory the terms discrete and continuous do not even occur.

So, if we model the milk factory as described above, we are dealing with a network of tanks, pipes and machines. All these components are working simultaneously, causing products to flow.

What are you thinking of? Continuous modelling? Discrete modelling? It is too ridiculous that a choice like this should be made at all. Just start modelling according to the concepts of system theory, finding out what the components are. A class of tanks for example. What are the attributes of a tank? A tank can hold a product, so the tank needs an attribute referring to that product. That attribute is discrete because the product in the tank can be replaced by another one; for instance a cycle of milk, air and water. The tank will have the attribute FILLING LEVEL, which value will vary continuously depending on the in and out flow. A differential equation will be attached to the attribute FILLING LEVEL specifying that relation. In other words, real world systems should be described most naturally using combined modelling tools.

When a model is running several data streams will be generated. Most of these streams are needed for output purposes as graphics, animation and statistics. Other data streams are created to be used as functions in the model, such as tabulated functions and distributions like the arrival pattern of customers in the supermarket example.

Nowadays those different kinds of data sets are covered by just one common concept called *point stream*. Every ‘point’ in the stream is an equally sized data

structure containing a code indicating how the information in the structure must be interpreted. For instance, as a table or as a discrete step function or as a continuous function holding zero or more derivatives. The uniformity enables point streams to be moved to and from hard disks (virtual memory) with a minimum of accesses. In fact the integration mechanism is just one more generator of point streams; one for each continuous attribute. Due to the possibility of derivatives in the 'points' the history of each continuous value is available at all levels without loss of accuracy. So, in models of dynamic control systems historical values of continuous attributes can be used directly as coefficients in differential equations.

In conclusion: combined modelling is just modelling. Practically, the feature of continuous attributes fits perfectly in the concepts of modelling tools based on system theory.

Some final remarks:

- As a consequence of the combination of discrete and continuous processes a tool for combined modelling offers the concept of 'continuous state event'. If, for instance, mother starts the oven in the kitchen she has to wait while the temperature in the oven reaches 180 degrees before putting in the chicken. Due to the point stream concept, the moment that mother will be triggered to proceed can be found with any desired accuracy.
- The integration mechanism will never create an integration step passing the next discrete event. Therefore, discontinuities caused by discrete processes will not harm the integration stability.
- The method of finite elements for solving partial differential equations fits extremely well in the modelling concept. Just consider elements to be components and their connections as a network.

5.5 Pitfalls

There are many ways to make a mess out of a model and there are no exceptions for models on food and food processing. Some sins are rather stubborn. Based on my experience I will show you the top four.

1. **Violating the goal of the model**
   It is impossible to create a virtual duplicate of a real world system. We have to limit ourselves to special aspects of the system according to a well-defined goal. Our supermarket model was intended to investigate the influence of logistic rules on specific aspects of the system. So, there are no customers in the model and no special storage for deliveries. By adding such a storage place, the shelves in the shop and the people filling these shelves (including their working schedules) have to be added as well. If these people are not alert there will be more missing sales. But that cannot be a reason for changing logistic rules. Why is this rule violated so much? Mostly by the influence of outsiders. The boss wants an animation for

political reasons. That would be terrible of course if you need to process millions of sales in a few minutes. Another wants to extend the model to investigate the influence of missing goods (either by error or theft). If there is no detection it will be obvious what will happen. An investigation of detection methods asks for a quite different model. If we include extensions like that, the model results will become cloudy.

My advice would be never to extend the model beyond what is necessary. If you are forced to do so, create a duplicate for each outside wish and adapt it for the specific goal.

2. **More levels of aggregation in one model**
We can model a packing machine in different ways:

- As a device that will pack a given amount of a product in a time that is, for instance, normally distributed with an average and deviation being attributes of that product.
- As a component with the following process description: take a pack from a pile, unfold it, bring it to the outlet of a tank, etc.

The aggregation level of the first way of modelling is higher. We should use the first way if we create a model of the factory in order to solve the yoghurt problem of paragraph 1. In case the distribution function mentioned is not available and cannot be obtained by measurement, just create a model of the package machine first to find that distribution.

Why is this rule violated so much? The packing machines are the most striking components in the factory. The rest of the factory looks rather dull and normally causes less trouble. So, these packing machines look important. However, not for the purpose of the model. Certainly not if such lower aggregated part takes 95% of the computer power.

3. **Too many conditions in a process description**
If there is a series of 10 not nested conditions in a process description, the component involved has 1024 ways to do its job. Would you ever verify so many routes? Murphy is waiting! If the process is really that complex, go for it. However, if these conditions are there to reach data needed to describe an activity, we commit a sin. Conditions like these are caused by missing or wrong directed references defining the network of relations of the components, and/or by attributes being attached to other components than they should be attached to. So, the conditions are the result of clumsiness in step 2. Assume, an operator has to select a packing machine to pack an amount of some product, but not all machines are suitable for each product. Therefore step 2 must contain information about possible assignments of products to machines. Because there is a *direction* in the activity of the operator (from a product to a machine) that information should be attached to the products. If you give each product an attribute specifying the suitable machines, no conditions are needed for the assignment. But, if you provide each machine with an attribute stipulating the suitable products for it then you are asking for trouble.

4. **Introducing false deviations**
   Suppose we have to model the arrivals of vehicles at the start of a road and we know that 20% of the vehicles are trucks. An example like this occurs in any modelling textbook. The almost trivial solution is: sample a number R from the uniformly 0,1 distribution and if $R < 0.2$ the next vehicle will be a truck. If 100 vehicles are generated, we get an average of 20 trucks with a deviation of 4. Nothing wrong so far.

   However, if we model the arrival of trucks early in the morning at the milk factory in the same 'almost trivial' way, knowing that the milk factory owns 32 small trucks and 8 big trucks, we commit this sin. By applying the sample approach we get an average of 8 big trucks with a deviation of 2.5 *instead of 0, as there are only 8 big trucks, nothing more, nothing less*. To avoid this sin, we have to schedule the arrivals of the 8 big trucks and the 32 small ones beforehand. See paragraph 3 for an example.

5.6 Conclusions and trends

Computers more and more control our world. The supply of water and electricity are controlled by computers already. A next step will be food supply. Nowadays, many production processes are (partially) controlled by computers and we are working on the logistics as well. It makes no sense to be afraid of automated control. It just happens and we can only hope that it will be done well. Modellers become more and more important in this process. All control starts with models. We need to know how real world systems work before we can control them. In the case of electricity (or water) supply, the modellers are raised in similar technological environments. They are able to communicate in terms of their models. That will be different in food supply. The modellers involved generally come from totally different disciplines and the confusion will be sky high. It even happens that similar terms are used for different concepts. Modellers have to be aware of that. The principles of system theory and a common attitude about the way nature is composed may form a base for mutual understanding.

In the first place models are used to explain the world. More than 50 definitions of the term 'model' are based on that goal. Afterwards models are used for control purposes as components of an information system (being a network of course). The function of a model is different then. It will amplify the value of data needed for decision making. However it is not a matter of 'either or'.

As shown in the example of the supermarket, models must interact with databases, other models and people. As a consequence, facilities for using a model in a network are at least as important as facilities for creating models.

The trends are clear in this respect. Computers will grow better and faster, however still based on one single extremely powerful but expensive processor. Trusting that nature always strives after efficiency there must be possibilities for computers with many cheap processors that can be freely connected to form

networks. People are working in this direction and some results look promising but will take time.

The development of software tools for modelling and using models will grow slower because of the variety of users, which has everything to do with the worst trend: the education of modellers. Courses on modelling and simulation are still not settled in universities. What has to be modelled is still considered to be more important than what modelling is about. Maybe, this book will be helpful this way.

Part II

The principles of modelling: empirical approaches

Introduction

It is not always feasible to unravel complex problems into their constituting processes, as, quite often, the current level of knowledge is not adequate. In other cases, decomposition could result in a complex model definition that cannot be parameterised in full, or which complexity is not warranted given the intended application of the model. In these cases, modellers may revert to the use of empirical approaches.

The development of this type of model is typically data driven and requires minimal knowledge of the products or processes involved. The resulting model converting model inputs into outputs is often referred to as a 'black box' as there is no relation whatsoever with the real underlying mechanism.

This complete absence of expert knowledge is accounting for both the main advantage and disadvantage of empirical approaches; models can be developed relatively quickly and give good results even when there is insufficient understanding of the processes involved, but at the same time, empirical models will not be able to increase the understanding or generate new knowledge on the underlying mechanism.

Chapter 6 offers a practice driven introduction into developing inductive models (models induced by the data). It outlines the general approach followed by most inductive modelling techniques; how to select and represent your data, how to select and search the space of possible models, and how to validate the final model to assess its accuracy. Several inductive techniques are described in more or less detail and applied to four case studies ranging from recipe planners for the process industry, to postharvest physiology. Chapter 7 covers the same subjects in a less formal way providing a broad

facial overview of inductive modelling techniques, referred to by the authors as data mining.

The predictive microbiology has developed a large number of inductive black box models with parameters that could be, *a posteriori*, interpreted in a biological meaningful way. The microbial environment is, almost by definition, prone to all kind of stochastic elements. Chapter 8 describes how these stochastic elements can be introduced into predictive models, generating predictions in terms of probability density functions rather than absolute values. The authors thoroughly discuss the statistical technique of accurate parameter estimation including their uncertainty.

Maarten Hertog

6

The power and pitfalls of inductive modelling

F. Verdenius, ATO BV, Wageningen and L. Hunter, University of Colorado, School of Medicine

6.1 Introduction

Modelling a process can be accomplished with varying degrees of background knowledge. If the underlying physical principles of the system to be modelled are well understood, it is possible to build mechanistic models of the process, which require little additional data (perhaps a few parameters to be estimated). The starting point for this kind of model is a hypothesis about the underlying process. This hypothesis, the basic understanding of that process, is translated in a, possibly parameterised, formalism that maps hypothesised mechanisms onto model elements. Parameters in the model represent qualitative or quantitative aspects of the hypothesised processes (in food processes, e.g. reaction rates). Actual parameter values are obtained from theory, from experts or they can be derived from real-world data by curve fitting.

However, in food processing as elsewhere, we are often confronted with a situation where there is very little relevant pre-existing background knowledge about a process to be modelled. In this situation, it may still be possible to use data describing the process and its various possible outcomes to induce a good model – perhaps even leading to the formulation of a testable hypothesis about the mechanisms involved. Traditional statistical approaches test a single hypothesis, focusing on linear and non-linear regression for continuous output variables, and discriminant analysis for classification problems. These approaches are powerful and widely used, but require as a prerequisite a general understanding of the causal relationships underlying the process to be modelled in order to generate the particular hypothesis to be tested.

Recent advances in various disciplines, including statistics and machine learning, have loosened the required prerequisites for data-driven modelling. It is

now possible to not only test particular hypotheses about a process, but to search among an entire family of related hypotheses to find the one that best describes the process. The innovations that have led to this ability relate to the development of less severe (and more appropriate) penalties for testing multiple hypotheses, which are not independent of each other. Despite the mathematical framework, which allows the testing of multiple hypotheses, and the computer power to automate the otherwise impossibly tedious repetition, we can never have enough data to test all possible hypotheses for the explanation of a phenomenon of interest. By restricting our search of the space of possible hypotheses to the areas that make the most sense, we can take full advantage of the induction process.

Practical induction methods require the modeller to make decisions about three general classes of issues:

- the *selection* and *representation* of the data used to induce the model
- the *class of models* to be considered, and the *search method* deployed to find the best particular model
- the *validation plan* that is used to generate an accurate assessment of the accuracy of the final model.

We will consider each of these tasks in more detail below. Briefly, the selection of training data is important because no induction method is capable of telling an investigator that some particular relevant aspect of the process was left out. The investigator must ensure that all of the relevant factors (and not too many irrelevant factors) are presented to the induction program. The term 'data representation' alludes to details of how data is presented to the program. The particular choices an investigator makes in representing data have an important effect on the outcome of the induction process, and unfortunately there is no easy method to ensure that the best possible choice has been made. Finally, the creation of a validation plan at the outset of the induction process is crucial in ensuring that the accuracy (and other aspects) of the induced model are accurately assessed at the end. Errors in the validation plan can easily lead to overly optimistic assessments of model accuracy and disappointment later on when the model is deployed.

The development of an inductive model is more than simply binding the three aspects of representation, search and validation. An investigator must also make good choices about the selection of processes to be modelled, and how a resulting induced model will be used in the field. Subsidiary considerations here might include breaking down the process to be modelled into several sub-processes and modelling each of those individually, or deciding that a mechanistic model might be appropriate for some aspect of the overall task. We call this process the *application methodology*. Major steps in the design of inductive models are:

- the decomposition of the required total application functionality into appropriate (sub)tasks
- selection and application of one or more of the above-mentioned techniques to create accurate models for each required task. (Often, sub-tasks find other

than inductive implementations, e.g. deductive models or expert heuristics. Designing these functions, far from trivial, falls outside the scope of this chapter.)
- Integrating the resulting models into process control or otherwise taking full advantage of the predictive abilities of the new model.

Inductive models find their origin in a wide variety of disciplines. The origin of the techniques manifests in these aspects. Specific representations indicate the origin of techniques, e.g. decision rules (cognitive science and artificial intelligence), neural networks (neuro-psychology, neuro-physiology and physics), non-parametric and bayesian techniques (statistics) and evolutionary techniques (biology). Similarly, search methods and estimation criteria can be indicative of the roots of some techniques. Inductive techniques form a main research topic of machine learning, a sub-field of Artificial Intelligence.

Induction is generally taken to mean the prediction of a particular outcome (the 'dependent' variable) from a set of inputs (the 'independent' variables). Classically, the prediction of continuous dependent variables is called regression, and the prediction of discrete dependent variables is called discrimination (in the binary case) or classification (in the multiple value case). These tasks are grouped together under the rubric 'supervised learning', since the program is always provided with an outcome. For completeness, we point out that machine learning can also be used for induction in other circumstances, for example, when there is no particular outcome and we just wish to discover patterns and relationships among a set of independent variables (e.g., association mining or clustering; see Mitchell[1] for a discussion). Inductive models can be used for a variety of tasks. The MLNet ontology of machine learning[2,3] defines eleven different learning tasks where machine learning can play a role. Part of the task structure is shown in Table 6.1. One of the subtasks is classification. In classification, all records in a data set belong to one of a (predefined) set of classes. Learning a classification function means that the function maps a data record onto a single class, being one out of a finite set of classes. When we talk about inductive models in this chapter, we address models for supervised learning of classification tasks, unless explicitly stated otherwise.

Table 6.1 A non-exhaustive overview of learning tasks and techniques that can be deployed for these tasks. The terminology stems from the MLNet ontology[2]

Learning tasks	Techniques
Characterisation (descriptive setting)	Decision tree learners, rule learners, …
Clustering (unsupervised learning)	k-means clustering, Kohonen maps, LVQ, …
Concept learning	Decision trees, rule learners, AQ, …
Function approximation	Neural networks, regression functions, …
Reinforcement learning	Q-learning, …
Time series analysis	Hidden Markov models, …
…	…

In this chapter we will elaborate the two aspects of inductive techniques (section 6.2.1) and the methodology for applying inductive models (section 6.2.2). Then we will attend the pros and cons of inductive modelling (section 6.3). This will result in some guidelines that help to recognise opportunities for applying these kinds of models and to avoid the pitfalls. The chapter ends with sketching some future trends in section 6.4.

6.2 Key principles and methods

6.2.1 Inductive techniques

Often, inductive techniques are grouped according to the originating technical paradigm. This grouping of techniques, however, ignores common characteristics of techniques. We will organise our discussion in terms of representations, search and validation, noting differences among the various inductive techniques only when they are significant. For a complete introduction in inductive techniques, the reader is referred to Mitchell,[1] Michie *et al.*[4] and Langley.[5]

Data selection and representation

Data selection

It is often the case that an investigator building a model must work with whatever data is available. However, the quantity and quality of available data has a significant impact on the quality of the induced model. Just as statisticians are generally consulted in the early stages of the design of a hypothesis test, it is best if the investigator building an inductive model is involved in the data gathering. The goals of data selection are to:

- capture all of the relevant features of the process to be modelled,
- gather enough data to ensure that the search and validation processes are not constrained by lack of data, and
- ensure that the data gathered is both unbiased and representative of the important aspects of the process being modelled.

The first goal is obvious: if relevant aspects of the process being modelled are not included in the inputs to the induction system, the quality of the resulting model will be suboptimal. Since by assumption we are working in a situation with incomplete background knowledge, it is hard to know what aspects are relevant. Fortunately, unlike classical statistical modelling, current inductive approaches are relatively insensitive to irrelevant data. That makes it incumbent on the investigator to use all potentially relevant data whenever possible.

We speak of one set of independent variable values and the associated dependent variable value as an instance or example. In general, an investigator should collect as many instances as is practical. Many of the pitfalls of inductive modelling involve the use of too few instances. However, there is generally a cost to gathering instances, so it is important to know how many instances is enough? There is no simple answer to this question. We can offer a few heuristics; note that there is no good mathematical support for these suggestions;

they come primarily from experience. In the simplest situation, there are a set of binary input variables and one binary output variable. In that case, a bare minimum would be three times as many instances as there are input variables. As the data representations become more complex (discrete values, continuous values), the number of instances required goes up. In general, ten times as many instances as there are input features is a reasonable minimum in most situations.

These suggestions take note of the theoretical results showing that there is a trade-off between the number of input variables used and the number of instances required.[6] A natural consequence of this relationship is that when the number of instances available for induction is small, the investigator must use other means to focus in on the input variables most likely to be relevant. Fortunately, a variety of methods have been developed for identifying relevant features before the induction begins; see Liu and Motada[7,8] for more details.

It is rarely the case that an investigator has too many training instances, but since the computer time (and memory) required for induction is generally proportional to the number of instances used, it is possible. In that case, there are simple methods that can be used to get the most out of the plethora of available data. One obvious approach is random sampling. Various kinds of biased sampling can also be valuable; for example, if the outcomes are unequally distributed (e.g. one interesting outcome is very rare), one may want to sample so that the rare outcome is over-represented compared to the raw data. A variety of more sophisticated 'resampling' approaches have been developed recently. For example, one might want to induce an initial model, and then resample the training data to include more of the kinds of data upon which the model is not performing well. A more elaborate version of that basic approach, called 'boosting', has found wide acceptance.[9] Another popular type of resampling is known as Gibbs sampling.[10]

Data representation

Knowledge representation has been a central issue in Artificial Intelligence over the past decades. Much of the general work in knowledge representation has been concerned with deductive inference and efficiency in storage. The representational criteria involved in induction are somewhat different. We will first discuss some general concerns in data representation, and then look at the interaction between particular representational choices and particular inductive methods.

Most inductive techniques operate on attribute-value representations (see Table 6.2), basically a single data table. Each column in the table represents an attribute. Each row in the table represents an instance. The simplest attribute type is binary (or Boolean). In fruit classification, the presence of specific diseases, such as spots on the fruits may be indicated as a *true/false* or *spots/nospots* value. Nominal attributes allow several mutual exclusive values; a nominal attribute *soil type* may take values like *sand*, *clay*, or *sandy clay*. Numeric attributes can take values from ordinal, integral or real domains; attributes like *days to maturity* or *number of days with humidity > 90%* may take values such as 24 and 30. An attribute-value representation for a hypothetical crop now may look like the one shown in Table 6.2.

Table 6.2 Data representation and the concept to learn. The table shows (a) an attribute-value of representation in the domain of plant breeding, with data types numeric (e.g. days to maturity), nominal (e.g. soil type) and binary (spots). Part (b) shows part of the table in a relational representation. Part (c) shows the concept that underlies the 'spots' column in the table. It can be learnt from the relational representation (b) with inductive logic programming, but will never be the outcome when learning from the attribute-value representation (a).

(a) Attribute value of representation

Product	Soil type	Days to maturity	# Days Humidity > 90%	Parent 1	Parent 2	Spots
1	s	23	33	73	74	s
2	c	23	37	56	64	n
3	sc	21	36	52	63	n
4	s	21	29	52	77	s
5	sc	20	28	68	96	s
6	sc	22	29	54	85	n
7	s	21	33	2	1	n
8	s	22	29	4	3	s
9	sc	21	31	6	5	n
10	sc	19	32	8	7	s
11	c	19	30	10	9	s

(b) Relational representation

Soiltype(1, s); DaysToMaturity(10, 19); DaysHumidityAbove90 (9, 31);
Parent1(1, 73); Parent2(7, 1); Spots(8);

(c) Represented concept

Spots(I):–
Parent1 (I, IP1), Parent1 (IP1, IGP1),
Parent2 (I, IP2), Parent2 (IP2, IGP2),
Spots (IGP1), Spots (IGP2).

There often are several alternative ways to represent a particular feature; some features can be reasonably represented as binary, nominal or numeric! The particular choice for data representation should be determined in two steps. In acquiring the data it is important to collect data in a format that is sufficiently informative for solving the problem.

Consider the issue of how to describe the colour of an object to a program. Objectively, colour is best represented by a high-resolution spectrum indicating light intensity for a large number of frequencies. Many measurement devices deliver data of lower resolution, represented in Red-Green-Blue (RGB) of Hue-Saturation-Intensity (HSI) measures (both 3D numerical formats). For many problems, however, the required data can have a much lower resolution. In assessing colour development of ripening fruits, 5-, 7- or 9-point scales may be applicable. Alternative representations for the nominal *soil type* attribute mentioned above may be a detailed chemical analysis, resulting in a high dimensional numerical representation (e.g. pH, fractions for different particle sizes, compound concentrations, etc.). Or it may be summarised in silt fraction, being one numerical. Or it may be from the nominal domain mentioned above.

When designing the actual inductive solution, the data representation may be adapted optimally to suit the inductive technique to be used. For example, when there is sufficient training data to support a complex neural network, in general the performance is best using representations that spread the input variables over several binary input nodes. Various schemes, such as *grey coding* are used to ensure that the binary values used to represent semantically nearby attribute values are close to each other in Hamming distance. Many symbolic rule learning and decision tree learners, on the other hand, perform best on nominal domains. For such techniques, discretising data before learning may pay off. Here again, there are few clear rules, and it is often worth experimenting with alternative choices. In one medical application that we are familiar with, the re-representation of age from a continuous variable to a nominal variable with five-year bins resulted in a dramatic increase in predictive accuracy.

Some inductive techniques, generally called 'relational learners' can take advantage of logical assertions and relational database tables. As is generally the case, learning from these more expressive representations requires both more training data and more computer time. However, relational representations are strictly more expressive than attribute-value representations, so there are some concepts that can be expressed relationally that cannot be expressed in attributes and values.

Suppose that a classification label spots is determined by the following relation: 'A crop has spots when both its grandmother from mother's side and grandfather from father's side had spots'. When this relation is represented in an attribute-value representation as shown in Table 6.2, there is no way an inductive technique can learn that relation. When, however, the same data content is represented in a relational format, inductive techniques that operate on logical representations are able to induce the correct relation (see Table 6.2). This type of complex relation often occurs in agricultural, biological and industrial domains. Examples are plant breeding,[11] and structure analysis of large molecules.[12]

Class of models and search methods

The quality and representation of data is only one of the important factors for the quality of induced models. An equally important aspect of the induction process is the types of models that are tested, and how the induction process proceeds.

Although modern techniques allow us to soften the statistician's injunction against 'fishing expeditions', it is still the case that we cannot test all possible models effectively. The set of models that our learning programs will consider must be constrained *a priori*, and we must determine how to search through the set of possible models we have decided to consider. By looking at the class of possible models that can be induced, and by considering the search methods, we can provide an ordered framework for considering the many alternative machine learning technologies that are available.

The space of possible models that is searched by any of these methods is most easily defined by considering the possible outputs of the modelling process. This is why the space of possible models is also sometimes called the hypothesis space – it is the complete collection of hypotheses that it is possible for the

induction method to consider. For example, if the induction process produces additive combinations of individual input features, we can see that it is searching through the space of possible linear splits of the data space; this is what the classical discriminant analysis approach accomplishes. If the output of the induction process produces hierarchical decision trees, the space of possible models that it searches consists of axis-parallel splits of the data space. This difference is illustrated in Figs 6.1, 6.2 and 6.3.

Some methods, such as Mitchell's 'Version spaces'[1] explicitly manipulate the hypothesis space. Others, such as artificial neural networks do not, and sometimes it is difficult to envision their hypothesis space. However, it is possible to characterise the set of functions that such networks can learn, which in effect defines their hypothesis space. In addition to discriminant analysis and classical regression, naïve Bayes and perceptron neural networks also learn linear models (in higher dimensional spaces, linear models effectively identify hyperplanes that divide the space). Discrimination trees and hierarchical regression (e.g. Breiman *et al.*[13]) explore a somewhat more complex hypothesis space that allows multiple splits. Nonlinear regression, association mining, support vector machines and hidden layer neural networks all work in hypothesis spaces that admit curved and/or multiple discriminations, and therefore explore the largest hypotheses spaces.

For best results, the space of possible models should include a model that fits the data reasonably well. Consider the differences between the space of linear discrimination models and the space of univariate decision tree models. If the true process underlying the data creates a linear discrimination that is not axis-parallel, then no univariate decision tree model can capture that very well; see Fig. 6.3 for an illustration. This illustrates the need for assessing class distributions in the data set in order to select a matching inductive technique. Sometimes the class distribution can be derived from available domain knowledge; more often it cannot. In those situations, the underlying distribution may sometimes be obtained from the available data. Note that the illustrations in Figs 6.1, 6.2 and 6.3 are very simple two-dimensional examples. In the more general case with many input features which may be correlated, it is not easy to tell what the class distribution looks like.

The more different kinds of models that a space can admit, the more difficult it is to find the best (or even a good) one during a search. For that reason, it often makes sense to start with relatively simple model spaces and move progressively to more complex ones until adequate accuracy is reached. Weiss and Kulikowski[14] develop a stepwise approach for solving problems with unknown distributions by exploring technique performances in order of increasing computational complexity (and increasing expressive power). As a first step, they advise linear discriminant methods. As long as the classification performance is not satisfactory, they apply as next steps more powerful techniques: subsequently linear discriminants, *k*-nearest neighbour,[4] decision tree[45] and back propagation neural networks.[15] As this is a blind strategy, success is not guaranteed. In their words, if you want the best performance, you should try all available techniques.

Another useful approach is to be aware of the signs of a mismatch between a model space and the training data other than just poor predictive performance.

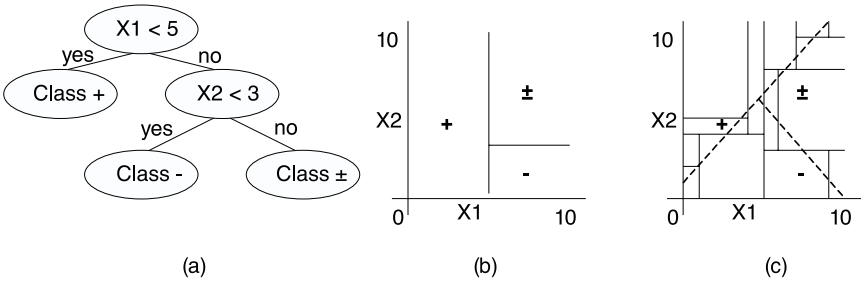


Figure 6.1 A simple decision tree and the partitioning of the data space that is represented by this tree. A branching node in the tree (i.e. $X1 < 5$) implements a test on a data tuple. In case of a positive test result, the left branch is analysed, in case of a negative test result, the right branch is analysed. Analysis continues until a node is reached that predicts a class membership.

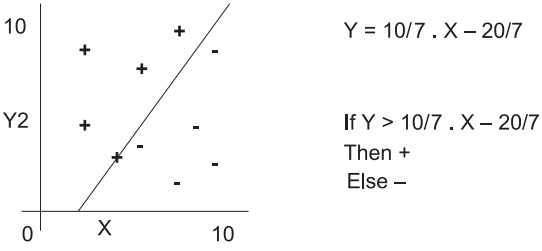


Figure 6.2 A concept space (left) with a linear discriminant function to separate the classes $+$ and $-$. The definition and rule interpretation of this discriminant function is shown on the right.

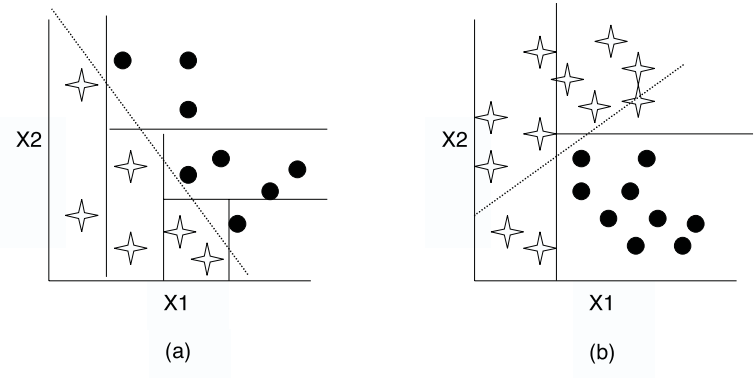


Figure 6.3 In panel (a), a linear discriminant is the best way to divide the light pluses from the dark circles; as shown with a dashed line; an axis parallel decision tree must make repeated splits in the same dimension to approximate a linear split (solid lines). In panel (b), an axis parallel decision tree easily divides the two classes of data, but no accurate linear discriminator exists.

For example, in decision trees, if the same variable is used repeatedly at various places in the tree, that is a sign that an approach not requiring axis parallel splits would be more appropriate. Or, when training a neural network, if the test performance is significantly worse than the training performance (see validation section below), then that is a sign of overfitting.[1] Accidental peculiarities of the training data, not meaningful for the concept to learn, have been learned as meaningful by the neural model. This suggests that the investigator should change the training procedure, or reduce the number of hidden nodes being used.

Searching the model space

Induction involves not only the definition of the possible hypotheses, but also the selection of one of these hypotheses based on the available training data. No learning method employs an exhaustive search through all the possible hypotheses defined by the model space, since that would be prohibitive in terms of computational time. The alternative to looking at every possible hypothesis is to try to identify the best (or at least a good one) by searching.

Search methods define how, and with what bias, models are searched. Fayyad *et al.*[16] distinguish between model search and parameter search. Model search is the selection of the detailed model configuration within a formalism. Within a representation and with an estimation criterion, a learning technique performs a structured search through the space of possible models (within the representation). The result of that search is a model that performs best according to the estimation criterion and the available data.

In the previous section we discussed the representation formalism the model is expressed in. But within a formalism model design can be expressed in model variants. Let us take the neural network as an example. Yang and Batchelor[17] use a neural network to predict the severity of soybean rust as a number between 0 and 100%. The network takes seven continuous inputs, and uses three layers. Their final network configuration counts seven input nodes and one output node. In the model search phase, they experimented with different numbers of hidden nodes in the middle layer, to determine the best configuration of the hidden layer. In Fig. 6.4(a) and (b) two model configurations with three and five hidden nodes, respectively, are indicated. Given the configuration of the hidden layer, the parameter search searches in the weight space of the configuration, to find the optimal weight values (in terms of the estimation criterion; see below) for the connections between nodes.

The model variants of Fig. 6.4(a) and (b) only differ in the configuration of the hidden nodes. Model search could have been more fundamental. Different input-output configurations could also have been considered during model search, e.g. transforming the continuous output into a nominal output (e.g. 0–30%, 31–60%, 61–100%), mapping each range on a separate output node (Fig. 6.4(c) and (d)). Doing so of course also influences the data representation and underlying distributions, consequently influencing the mappings to be induced.

In practice, model selection is not often realised as a systematic and rational search process. There may be constraints on resources, such as available

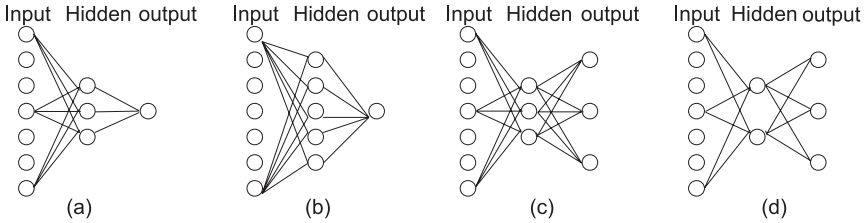


Figure 6.4 Some possible model configurations (not all connections are represented) for neural networks for predicting soybean rust. Models search iterates over these configurations, parameter search estimates optimal weight values for links within one configuration. Configuration (a) and (b) fall within the data representation that Yang and Batchelor used in their work, configuration (c) and (d) would have required a change in data representation.

software implementations or computational resources, or limitations on the experience or comfort of the system designer which constrain the model search process. However, it is generally worthwhile to be as systematic and comprehensive as is practical in any particular situation.

Each particular induction approach embodies its own search through the space of possible hypotheses that might explain the data. Some general search strategies in inductive modelling are illustrated in Fig. 6.5. In greedy search, alternative next steps are locally assessed, and the best one is selected. No decision is ever revisited, so that if in retrospect it might have been better to make a different choice, greedy methods will fail to make the optimal set of

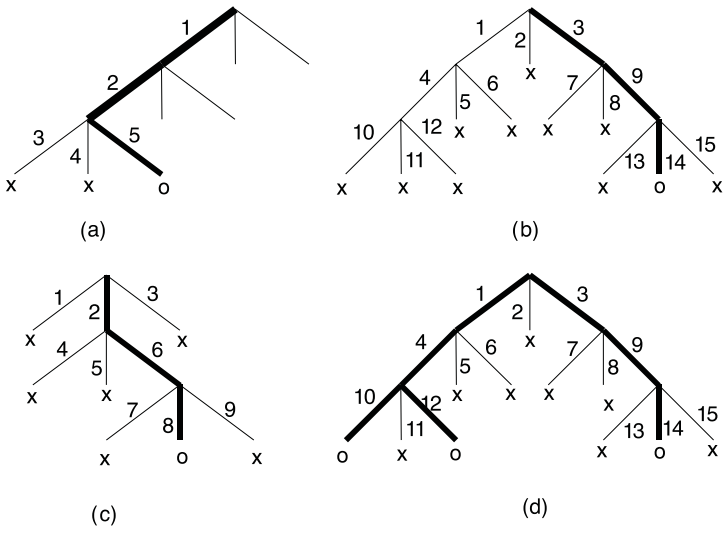


Figure 6.5 Some often used search strategies. (a) and (b) show two different concepts for exhaustive search: depth first and breadth first. In general, these methods are not applicable because of the size of the search space. Greedy search (c) expands the locally best solution in each search step. Beam search (d) expands a group of best solutions.

choices. The most popular types of decision tree induction generally use greedy search: each input variable is considered, and the one that seems to make the best division of the training data (often evaluated by 'information gain', an entropy-like measure from communications theory) is selected. Then the training data is split based on that variable, and then for each subset, the next best variable is selected recursively. Hill climbing is a continuous variant of greedy search where a series of adjustments to a current state are evaluated to see which gives the greatest improvement in outcome. A step in the best direction is then taken, and the process is repeated. The gradient descent methods used to train neural networks do this sort of optimisation of an error function to select the best weights for the neural network. As with climbing real hills, sometimes the route to the top requires going down a valley, and if one only heads upward, then one gets stuck on a subsidiary hilltop (a local optimum) rather than the final peak (the global optimum). It is also the case that greedy and hill climbing methods require some reasonable measure of progress (or, equivalently, error) so that reasonable choices can be made along the way.

Another general search method that is widely used is called 'beam search'. The idea of beam search is to avoid the local optimum problem by keeping around a small number of good candidate decisions during the search process, rather than just a single best candidate as greedy search does. Beam search is frequently used in genetic algorithms.[18] Another approach to avoiding local optima is to use *simulated annealing*.[19] Simulated annealing is a global optimisation method that is similar to hill climbing, but every so often a step is taken in a 'downward' direction. The downward steps are taken fairly often at the beginning of the search, but their frequency decreases exponentially as the search continues (hence the metaphor with annealing). Simulated annealing is very expensive computationally, but within certain limits can be guaranteed to provide the best possible answer.

In addition to these particular search methods, there are alternative approaches to generating the next subset of hypotheses that the search methods might select among. We mentioned the decision tree induction approach that uses information gain to pick the next feature to add. The step for neural networks being trained with gradient descent is to calculate the partial derivative of each weight in the network with respect to the overall output error. The weights are all moved a little bit (how little is set by a 'training rate' parameter) in the direction that reduces the contribution to the error. In genetic algorithms, the next set of hypotheses considered is generated by two random processes. One process takes a current hypothesis (recall, that due to beam search, there are always several active hypotheses being considered) and *mutates* it, making a small random change. The other process is to take two active hypotheses and *recombine* them to make two new hypotheses that each embody some of each parent hypothesis. Then the newly generated hypotheses are evaluated by a user-supplied fitness function, and the best are kept for the next round.

Other inductive techniques

The previous sections described several specific learning techniques (neural networks, decision tree induction, genetic algorithms, linear discriminants) and mentioned several others, but the main goal was to describe the common underlying techniques of model space representation and search. For completeness sake, in this section, we briefly introduce several other induction methods of interest. For more details on each particular method, consult the references above, or the documentation that comes with a particular system.

- *Lazy learning techniques*: most inductive techniques generalise a classification model from the learning data. Lazy learning techniques simply store the training data. In other words the knowledge representation is the same as the data representation. Consequently, learning time reduces to storing a record in the database. When classifying an unseen instance the database has to be searched for matching cases on the basis of a similarity criterion. The most commonly used similarity criterion is based on an Euclidean distance measure between two instances X and Y:

$$\sqrt{\sum_{f \in Features} (X_f - Y_f)^2}.$$

  In principle all cases in the memory have to be assessed, which makes the classification time grow with the size of the training set. Smart indexing mechanisms can reduce this time requirement. Knowledge based indexes, defining a weak domain theory onto the search space, have been developed in case-based reasoning.[20] Other indexes to reduce the search space are based on statistical properties of the data, e.g. by deploying a clustering mechanism.[21] To obtain a class for a new case, either the class of the k best matching solutions can be used as-is (k-nearest neighbour[4]), or one or more cases are interpreted and used for further processing (instance-based learning;[22] case-based reasoning[20]).
- *Support vector machines.* Support vector machines are related to nearest neighbour techniques, but they do some computation ahead of time. They require the user to supply a kernel function (similar to a similarity metric, but with a loosening of some mathematical constraints), and the induction method finds the maximum margin hyperplane (one that does the best job dividing the positive and negative examples) using the kernel. Only the training examples near the dividing hyperplane need be considered when classifying a new example, saving some time in the classification step. Also, these methods have good statistical properties, such as the ability to estimate accuracy on unseen data with empirical techniques like cross-validation (see next section).
- *Naïve Bayesian learning.* A classical statistical method called Bayes rule[1] can be used to calculate the probability of the hypothesis given a set of data relevant to it. This can be used as a method to search some particular types of hypothesis spaces for the most likely hypothesis given a set of data. The probability of a set of data given a hypothesis $p(D|H)$ is a straightforward calculation for many kinds of hypotheses, such as those that assume mixtures

of Gaussian distributions of independent inputs. Bayes rule says that the probability of a hypothesis given a set of data $p(H|D)$ is proportional to the product of the data given the hypothesis $p(D|H)$ times the prior probability of the hypothesis $p(H)$. We can select a space of hypotheses for which such probabilities are computable, define a reasonable set of priors on those hypotheses (e.g., hypotheses that involve fewer input variables are more likely than ones that involve more), and then search for the most probable hypothesis. The 'naïve' approach assumes that all of the input variables are independent of each other (e.g. they have no significant correlations) which greatly simplifies the calculations and reduces the amount of data required. Although this assumption is almost always false, this approach, particularly when combined with boosting (see above), often performs quite well.

Clustering, grouping data on the basis of some similarity measure, is another important task in inductive modelling. In this case learning data comprises instances (e.g. product characteristics) and the learner identifies groups in that data. In this case, only the input data is given; the instances are ungrouped and unclassified. Therefore this type of learning is called *unsupervised learning* (in contrast to *supervised learning*, see above). However, many other tasks exist.[2] Some of these tasks and matching techniques can be found in Table 6.1.

Validation of induced models

A model is never the truth. All models are incomplete and at least partially incorrect. So, for any model we are going to put to effective use, we must have a reasonable characterisation of its limits. Generally, the most important evaluation of a predictive model is an estimate of its accuracy. We will address this issue in some detail below, but we want to start by taking a broader view, including other factors that might make a difference in the practical effectiveness of an induced model.

Performance of inductive model

In addition to raw predictive accuracy, there are several other factors that can make a difference in the quality of an induced model:

- The *learning time*: for many techniques, the learning time depends on the size of the dataset to learn from, and the model complexity (e.g. neural networks); for others, it is independent thereof. An induction system that takes a very long time to train may exceed the resources available to a project using inductive modelling, or it may provide a correct answer that is too late to be of any use. A more subtle problem with induction methods that take a long time to learn is that they discourage experimentation with alternative input representations or parameter settings. In general, induction methods that are fast (like greedy decision tree induction) are preferable to ones that are slow (simulated annealing to train neural networks).
- *Classification time*: some techniques, especially techniques that store large models, require substantial classification times. This is particularly a problem

with nearest neighbour methods like support vector machines, although these classification times are almost always a good bit faster than most training times. Nevertheless, if real time performance (say, for process control) is a planned use of the induced model, then this should be considered.

- *Comprehensibility of the learned model*: it is often the case that the results of an induced model must be accepted by people other than the ones who created the model. The ability to explain how the model works is often helpful in getting support from others in an organisation. A 'black box' model with higher predictive accuracy may be less acceptable to decision makers than a slightly less accurate predictor that is more comprehensible. In general, genetic algorithms and decision trees produce models that are more comprehensible than neural networks and discriminant methods.

Estimating classification accuracy

Although the above considerations are important, classification accuracy is a crucial issue. A model that is fast and comprehensible is of no use if it is not reasonably accurate. Statistical techniques (such as linear regression) have underlying theory that allows the expected error to be estimated directly from the input data (e.g. R^2). However, most induction methods described here do not have the equivalent formal properties, and their accuracy must be estimated empirically. Generating unbiased estimates of the accuracy of an induced classifier using empirical methods is fraught with potential pitfalls. For example, it is trivial to generate a learning method that is always perfectly correct on the data it was trained on: just build a lookup table of the training data and *memorise* the correct answers. The goal of induction, however, is to generalise on the training data and to be able to make accurate predictions on future instances.

The memorisation problem is formalised in the idea of overfitting the training data. In order to get the best possible predictive accuracy on future instances, the training method should fit a model to the data, but not overfit the model too specific of the training data that will not be true of future instances.

When we estimate the accuracy of an induced model, we want to know how well it will do on future instances. Since we, by definition, do not have 'future' data, we can use various techniques to make unbiased estimates of that performance. The idea underlying all of these estimation methods is to hold some data out of the training process, and to test the induced model on the withheld (or 'unseen') data. The simplest way to achieve this is the train-and-test approach. In this situation, the original data set is randomly divided into a training set and a testing set. A model is induced from the training set. To estimate the model accuracy, the model is used to predict the classes for the records in the testing set. The performance of the model on the test data gives an unbiased estimate of the performance on the model on future data. Train-and-test does not make use of all training data. This might be considered as an inefficiency of data usage. More serious is that if the data set is small, the variation in error estimate for different samplings increases. Michie *et al.*[4] advise using the train-and-test approach only when the size of the data set is large enough ($>>1000$ records).

Table 6.3 Schematic set-up for *n*-fold cross validation. A data set is randomly divided in *n* parts, each containing $1/n^{th}$ of records of the data set. In cv-trial I, set I serves as test set, all others serve as train set

Data set

with n samples

1	2	…	…	…	n

CV-trial	Test set	Train sets
1	1	2..n
2	2	1,3..n
..	..	…
..	..	…
n	n	1..n-1

But what if the number of records is smaller then 1000? Cross validation is a good alternative. Cross validation is basically a repeated variant of train-and-test. However, the test sets of the different trials are mutually exclusive, and cover the total data set (see Table 6.3). The model to apply can now be obtained from the total data set, with its error estimate being the average of the error rates of the cross validation runs. A special case of cross validation is obtained when each record is used as a test-set (leave-one-out). The statistics of estimating classification performance have been well explored (see, for example, Mitchell[1] and Michie *et al.*[4]).

A common method of selecting among various induction methods (e.g. between neural networks and decision trees, or among different numbers of hidden nodes in a neural network) is to run cross-validation on each alternative, and then select the method that gives the best cross-validation results. Although this is a fine method for selecting among alternatives, the estimate of accuracy for the best method is likely to be overoptimistic, since the estimate of accuracy was itself used to make a choice. If an accurate estimate of accuracy is required in such a situation, then the best approach is a second set of unseen data (often called the 'validation set') that is used only to report on the estimated accuracy (and never to make decisions about the inductive hypothesis).

6.2.2 Application approach

Inductive techniques generalise models from data. The role of learning may differ from situation to situation. On one side, in stable domains where abundant data is available, a model can be generated in a one-time event. In Example 6.1

Example 6.1 Product treatment support system[23]

Fruit selection with decision rules and prediction of treatment requirement with a neural network.

System functionality

The Product Treatment Support System supports a fruit-ripening expert in designing a recipe for fruit treatment. Fruit treatment is the process of transferring agricultural produce from an unripe stage into a stage of ripeness that the client wants. Applying the proper temperature and gas conditions enforces product ripening. It is accepted practice to apply standard recipes, and react during treatment on the actual development of the produce. However, regular quality variances combined with the reactive nature of this approach cause produce losses to be fairly high. It was the goal of the project to reduce the losses and increase the quality of the treatment process, by applying a more proactive recipe generator.

Product selection

Given a product batch, an expert manually selects products that represent the average physiological development of the batch. These products are further assessed as part of the product treatment planner (see Fig. 6.6). The sample size needs to be small in order to be feasible in practice. Random sampling requires a samples size of approx. 80–200 samples. Comprehensible selection guidelines are required to be applicable for the expert.

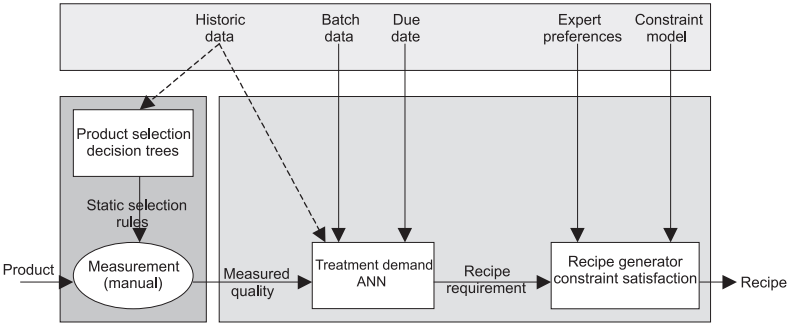


Figure 6.6. Functional overview of PTSS. A product batch enters the treatment facility. Human expert select products on the basis of static selection rules generated with a decision tree learner. The selected products are measured to assess the product quality. This measured quality serves as input to predict the recipe requirement with a neural network, which is transformed into a recipe in the recipe generator that deploys a constraint satisfaction model. The light gray area represents the data set, the dark grey area represents a manual process, and the medium gray area represents the planning module.

Product selection is a simple binary function, indicating whether a product is suitable for measurement or not. Data is obtained from 105 products from 8 different batches. The data set includes easy to assess external product characteristics, such as shape, size, maturity, etc. Class labels indicate whether a specific measurement value indicating ripeness is close to the batch mean. Given product characteristics, decision rules are derived from a decision tree. Application of these rules allows a sample size of approx. 5–10.

Recipe requirement
Given batch data, measured batch quality and the due date, a recipe requirement is demanded for the constraint satisfaction module. The required function is to provide a treatment recipe (prescription of condition values for a number of conditions, e.g. temperature, ethylene concentration, gas conditions, relative humidity). The process is decomposed in several steps. Global recipe specification are obtained from a neural network module.

decision trees are used in a one-time learning event to provide a human expert with knowledge on selecting products for fruit quality assessment.[23]

At the other extreme, in some domains every new fact generates the need to re-learn the model. This occurs not only in evolving domains, but also in domains where scarce data trickles in slowly over time, and where decisions cannot await availability of a complete data set (e.g. in the case of control problems and robot learning). Then it is not possible to learn a model in a one-time event, but an incremental or adaptive strategy is appropriate. By using new data facts to refine or retrain the model, performance may improve over time, and a model can gradually adapt in accordance with a drifting concept. Example 6.2[24] describes a situation where learning occurs under data limitations such as in the agro-industrial environment of mashing. Mashing recipes depend on a multitude of raw material and process parameters. Every new mashing run provides new information that may lead to adjustment of models. In the planning of mashing recipes, recipe planning is a case-based process. New cases are stored for later reference. New cases at the same time may lead to adjustments of the knowledge model that is used to adapt cases. Verdenius and Broeze discuss a control problem in water purification with concept drift.[21]

In between is a mixed setting where one learning process is rerun with different data sets (cf. knowledge acquisition strategy[25]). This is shown in Example 6.3[26] when analysing the functional divergence of homologous proteins. Another example of this setting can be found in marketing applications, where the learning goal is to identify and characterise prospects for a specific product from a client database (e.g. MLNet[3] and the COIL[27] and KDD[28] competitions). This is presumably a repetitive process that occurs every few

Example 6.2 Planning of mashing recipes[24]

Batch specific planning of a bio-process.

Introduction

Aarts[24] describes a knowledge-based planner for planning mashing recipes. Mashing is the process where proteins and starch components from the malt are converted into fermentable sugars and smaller proteins to facilitate the fermentation process. In his planner, Aarts combines case-based reasoning[20] with qualitative reasoning to plan complete recipes (including temperature settings, recommendations for the amounts of adjuncts, the malt (blend) to be used, etc.).

Case-based reasoning

The case-based reasoning approach realises mashing in four steps:

- *Process specification* (situation assessment). The malt analysis (raw material) and the wort specification (end-product quality) specify the mashing process; represented in a structured data format, this situation assessment serves as the functional specification of recipe.
- *Finding the ballpark solution*. The recipe specification is used for a fuzzy search in the case base, a database containing annotated recipes that have been realised in the past. Realised batches with similar recipe specification are identified and ordered according to their degree of success and the similarity with the current recipe specification. The most similar case is then selected as the ballpark solution.
- *Adapting the ballpark solution*. The new recipe specification and the recipe specification of the ballpark solution will normally differ in some aspects. The applied recipe of the ballpark solution serves as a starting point for the new recipe. Differences between the new process specification and the ballpark solution + its annotations (see below) are analysed with a qualitative model of the mashing process. Based on the identified differences, the ballpark recipe is adapted such that, according to the qualitative model, the new process results in the required specification. With the adapted process, the process is run, and the process results are fed back into the planner.
- *Evaluation of the realised recipe*. Process results, i.e. the wort analysis and process performance data, are used to annotate the applied recipe. Annotations serve as guideline to accompany cases in the case base, e.g. a guideline that a certain temperature phase was too long in the applied recipe and should be shortened. When such an annotation is encountered in a ballpark solution during planning, the adaptation module takes it into account when adapting a recipe.

Qualitative reasoning
The qualitative process model contains the available understanding of the mashing process at the enzyme level. Knowledge on enzymatic reactions was obtained from experts, from the literature, and from experiments. Knowledge chunks relate concepts such as concentrations, enzymatic reactions, etc. A detailed kinetic model nor numerical procedure for deriving process settings are present (as available knowledge does not allow such accurate calculations). Instead, advice such as 'increase α-amylase activity to increase the concentration of fermentable sugars' are 'guessed' on the basis of the case base content. During operational use of the system, the qualitative model can improve due to detailed feedback on qualitative effects in the actual running of the process. Moreover, the qualitative knowledge can also be used to refine human insight in enzymatic behaviour during mashing.

Effects
Applying this planner, a more constant quality of wort is realised in the realm of variable quality of the raw material. The system supports proactive planning. This improves the traditional approach, where constant recipes are used until the process quality becomes beyond specification. Consequently, proactive planning reduces the number of processing problems.

months on the same type of data. Consequently the process is standardised, but the data set is different for every iteration, and so is the learned model.

When inductive models are considered for solving a modelling problem, many questions arise. How can inductive models best be applied? Do different types of applications require different approaches? Does the kind of application influence the process of constructing it, or can we distinguish generic elements? Much work in machine learning and statistics concentrates on the technical aspects. The underlying assumption is that inductive models are suitable for the problem at hand.

Early work on applying machine learning techniques explored a technical view of the application process, e.g. in the Machine Learning Toolbox project.[29] The toolbox brings a large number of inductive (machine learning) techniques together. It is accompanied by a set of taxonomies, relating techniques to problem characteristics.[30, 31]

With the rise of data-mining (searching for patterns of interest with a particular inductive technique) and knowledge discovery in databases (KDD: '... the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data ..',[16]), a more process-oriented view on developing inductive models has gained interest. In the literature there is a growing inter-expert agreement on the phases in the development of KDD

Example 6.3 Identification of divergent functions in homologous proteins by induction over conserved modules[26]

Knowledge refinement in microbiology.

Introduction

The proteins that underlie biochemical functions are related to each other by evolutionary descent, i.e. are homologous. Homologous proteins do not necessarily exhibit identical biochemical function; in fact, functional divergence is required for organismal evolution. In practice, however, local or global sequence similarity between proteins is used widely as an indication of functional identity.

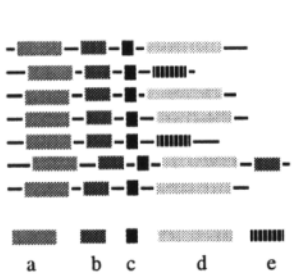
A general problem in mapping sequence to functional class is that of false positives, that is, sequences which are similar to a query, but have a different functional class. Inspection of specific errors suggests that in some cases it may be possible to identify automatically regions of a sequence that must be conserved in order for that function also to be conserved. Supervised machine learning was used to identify protein modules, which can be used to discriminate among functional classes of similar proteins. The target of the study was the set of protein groups that had significant sequence similarity to at least one protein outside its group.

Data design

A protein class is defined as a group of homologous proteins within a class of the enzyme commission classification. For each protein class an experiment is conducted. Positive class labels are assigned to all members of a group. Negative class labels are assigned to all non-positive proteins that have any significant sequence similarity to any positive example. Totally 251 training sets could be composed for the total domain. Homologous proteins are represented as binary vectors as indicated in Fig. 6.7.

Inductive approaches

Two different inductive techniques were used to generate classifiers able to recognise positive and negative examples: decision trees (C4.5,[45]) and naïve Bayesian discrimination (Mitchell,[1] chapter 6). For each protein class a tenfold cross validation experiment was conducted. Figure 6.8 shows the typical results of the Bayesian classifier and the decision tree for learning to classify the short chain alcohol dehydrogenases. Table 6.4 totalises the results for both techniques. From this table it shows that a combination of modular representations of proteins with sequence similarities improves the ability to infer function from sequence over similarity scores alone.



(i) conserved modules

a	b	c	d	e	b′
1	1	1	1	0	0
1	1	1	0	1	0
1	1	1	1	0	0
1	1	1	1	0	0
1	1	1	0	1	0
1	1	1	0	1	1
1	1	1	1	0	0

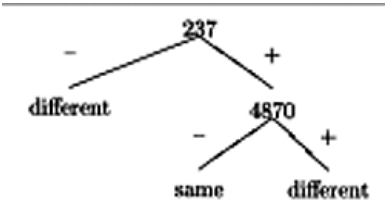
(ii) representation

Figure 6.7 Representing protein sequences on conserved modules. (i) A set of (hypothetical) homologous proteins with conserved modules shown as boxes. There are a total of five modules: a, b, c, d and e. (ii) A representation of the proteins on the basis of the modular attributes. Each protein is shown as a vector of attribute values. Modules which occur more than once in a sequence are called repeats. The second occurrence of module b is treated as a separate attribute b′.

Marginal probability of same function $P(+) = 0.474$
Marginal probability of different function $P(-) = 0.526$

Module	36	12	201	4870	237	238	36	201
Value	-	+	-	-	+	+	+	+
LP_c^+	0.778	0.778	0.778	0.778	0.778	0.778	8.444	8.499
LP_c^-	1.632	1.632	1.523	0.972	0.834	0.778	1.222	1.699

(a)



(b)

Figure 6.8 The probabilities used by the naive Bayesian classifier (a) and a decision tree (b) for a specific data-set EC 1.1.1.1. In (a), the relative contribution to classifying a protein as a member (LP^+) or a non-member (LP^-) of a specific group. The top rows indicate the module number, and their presence or absence. So, the absence of module 36 gives 0.778 support for membership of the group, and 1.632 for non-membership. In (b) the numbers at the nodes refer to ProDom modules, the branches are labelled '+' for module present and '−' for module absent. The leaves are labelled 'same' if the function is the same, and 'different' if the function is different.

Table 6.4 Summary of learning performance for 251 data sets. The performance results are summarised by learners from top to bottom, and accuracy range from left to right. The first column labelled L shows the learners; naive Bayes (NB) and C4.5. The next column shows performance (P):P_t (total), P^+ (same function), and P^- (different function). Each cell shows the number of data sets with performance P, for a given measure and learner, and the percentage of data sets that represents.

L	P	Accuracy range			
		< 0.9	≥ 0.9	≥ 0.95	1
NB	P^+	71 (28%)	180 (72%)	155 (62%)	140 (56%)
	P^-	8 (3%)	243 (97%)	225 (90%)	160 (64%)
	P_t	11 (4%)	240 (96%)	223 (89%)	132 (53%)
C4.5	P^+	85 (34%)	166 (66%)	145 (58%)	129 (51%)
	P^-	1 (0%)	250 (100%)	236 (94%)	177 (71%)
	P_t	10 (4%)	241 (96%)	227 (90%)	134 (53%)

applications. Although differences exist in the detail, some steps are often encountered, as shown by a comparison of the work by Fayyad *et al.* and Adriaans and Zantinge.[32]

Fayyad *et al.*	Adriaans and Zantinge
Data definition	Data selection
Data selection	Cleaning
Pre-processing and transformation	Enrichment
Data mining	Coding
Interpretation and evaluation	Reporting and application

Both approaches state that a data-mining problem requires proper deployment of a data-mining algorithm. But how do we know that the problem at hand is at all suitable for applying an inductive technique? In the case of supporting an expert in fruit selection (see Example 6.1), this may seem straightforward (but even here it is not!), and possible design decisions may seem obvious. But what if a complex classification system for functionality of homologous proteins (Example 6.3[26]) or a system for planning complex plant breeding processes[11] have to be designed? Such systems can be designed in many ways, some with and some without inductive components, some with one-time induction modules and some with full adaptive functionality. Consequently, an important extension to these approaches offers a systematic analysis of the functional requirements of the total application, including other than learning solutions.

MLNet[3]	
Problem definition:	unambiguous definition of input, output, and their relation.
Data design:	definition of the input and output data, including data typing and value ranges
Data cleansing:	data-set with meaningful data, a set of procedures to transform raw data into meaningful data
Implementation:	definition and configuration of machine learning technique that solves the defined problem
Evaluation:	assessment of the performance of the learning technique from step 4

By including an explicit problem definition step in the development process, where the full extent of the task at hand is analysed,[3] this problem can be partially overcome. Even then, however, the solutions under study are predominantly machine learning/inductive solutions. Hunter[25] introduces knowledge goals as a leading principle of the functional definition of the application. Several authors have proposed more elaborate approaches to applying inductive techniques.[33] Verdenius and Engels,[34] Engels[35] and Verdenius and van Someren[36] have emphasised the importance of task analysis during the functional definition of the application, a process guided by the problem solving methods and techniques that can be made available. Starting with the functionality of the total system, opportunities for applying inductive components are derived from available or acquirable data sets in a top down process.[36] Verdenius and van Someren sketch this process for the fruit treatment planner of Example 6.1.[23] Subsequently functional and non-functional requirements for the inductive components are defined. Functional requirements describe input and output attributes for the inductive component, and define constraints on the relations between input and output. Non-functional requirements include (but are not limited to) representation, accuracy, train and response times, required training set-up (e.g. one-time, incremental, etc.).

Another aspect in designing inductive solutions is that of technique selection. For each type of problem numerous techniques will be suited for the job in principle, while others seem less well suited. In many real-world projects the deployed inductive techniques were elected on the basis of heuristics, personal preferences and circumstantial availability (see also Verdenius and van Someren[37]). Moreover, tools and instruments to support the inductive system developer in a rational selection process are scarce. An often-used approach is to test the performance of several techniques on the same data set. By using cross validation unbiased estimates for the performance of the techniques can be obtained. The best performing technique is then used for the actual application. This approach may give suitable results but it has several disadvantages. First, it is time-consuming to test a large number of techniques. In principle, as there is no guidance on which technique is best, all available techniques should be included in such an approach. Second, testing a large number of techniques and selecting the best performing gives little insight as to the underlying problem and data distributions. Consequently, when a solution based on inductive

techniques has to be reused for a slightly different problem, the technique selection problem should be repeated. After all, because no insight has been gained in the justification of the selected technique, one cannot be sure that the same technique will perform best on this new problem. Finally, the experimental approach is no answer to the question whether a slightly different parameterisation of the technique would have improved the results.

Experimental comparisons (e.g. StatLog,[4] and also Lim *et al.*[38]) and problem solving competitions[27, 28, 39] suggest that the competence of the learning technique for a specific problem determines the quality of the inductive solution. This implies that selecting the best technique for the job is a crucial step in maximising the performance of the solution, but does not reveal what criteria should be used for selection.

During rational technique selection the characteristics of the data space and the classification function are surveyed, and the technique that is known to perform well for the found characteristics is selected. It is not always easy to determine the type of distribution in the data. In the StatLog project and adjacent studies,[40, 41] statistical descriptors of data sets (e.g. the numbers of examples, attributes, classes; mean, minimum, maximum skewness of attributes, etc.) have been defined and correlated to the performance of inductive techniques. This results in applicable relations that give some guidance to the process of technique selection. On the other hand, the resulting guidelines display apparently arbitrary relations between the descriptors and the performance, and only partially explain the relations found.

Representation determines the class of underlying concept distributions that can be expressed. Univariate decision trees can correctly represent concepts with concept boundaries that are orthogonal towards the attribute axis. Linear discriminant functions allow linear combinations of attributes to form the class boundaries, while neural networks impose no constraints on the shape of class boundaries. Search methods determine characteristics of the concept space shape. Hill climbing works well in a smooth error landscape under absence of local extrema, but performs poorly in bumpy or flat error landscapes (e.g. the kangaroo metaphor[32]). Statistical descriptors of the data set as indicated before do not (necessarily) relate to these aspects.

This illustrates the need to relate concept characteristics to properties of representation search methods and, maybe also, evaluation criteria, instead of data descriptors. Recently it has been suggested[42] that the applicability of machine learning techniques that apply orthogonal partitions, such as univariate decision tree learners, can be assessed at the data set level. In this approach distribution characteristics in the data set are matched in the wavelet domain against prototypical patterns in orthogonal distributions. A satisfactory match supports the application of decision trees. This matching process plays the role of a guard for univariate decision tree techniques. When the guard confirms applicability, the actual technique can be applied (see Fig. 6.9).

Although in theory similar approaches can be developed for other combinations of machine learning techniques/data distributions, in the current

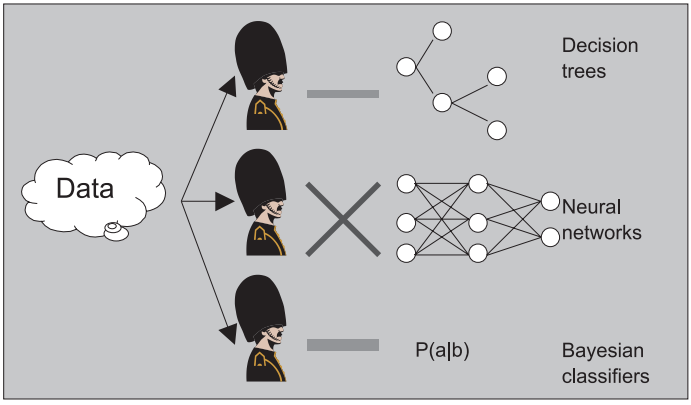


Figure 6.9 Rational technique selection by guards. A guard is a small software procedure that assesses the suitability of one inductive technique for analysing a data set. The guard results in a quantitative assessment of suitability. For decision trees, such a guard has been descibed in Verdenius.[42]

practice the assessment of the type of distribution in the data set is still based on craftsmanship. In specific situations, domain knowledge enables the inference of the type of data distribution.

6.3 Pros and cons of inductive modelling

Currently, inductive techniques do not play a major role in the modelling of food processes. Many of the modelling efforts in this domain are initiated from fundamental or applied science. Researchers in this area emphasise the structural accuracy of their models, i.e. they want their models to fit onto (possibly hypothetical) phenomena in the real world. Many inductive techniques, or more accurately, the data representations where inductive techniques are applied, cannot easily be interpreted in terms of these real-world phenomena. In several biological and food domains however, inductive techniques can offer insight in real-world phenomena. Shah and Hunter[26] (Example 6.3) apply inductive techniques to bring to light the relations between protein structure and protein functionality, while Aarts' data driven approach for planning mashing recipes (Example 6.2) results, as a side effect, in insight into the roles of enzymes during the mashing process.

Also, in scientific environments, data is mostly obtained from experiments on small populations. This results in relatively small data sets. Putting data sets together is hindered as every data set may come from an experiment with a different set-up; consequently, data design differs for each set. Mechanistic modelling is known to be less data consuming, and can therefore be applied on the smaller data sets resulting from experiments. In real-world environments data is generated in routine processes, for instance in production, distribution

and growth. It is, however, good to realise that many modern analysis techniques that are applicable in food processes are capable of generating huge amounts of data, for instance in multi-spectral analysis, micro-arrays, on-line process measurements and tracing and tracking data. Environments where such analysis equipment is deployed are promising for the deployment of inductive techniques.

Inductive techniques are not the panacea for solving every modelling problem. Still, when facing a new modelling problem it pays off to consider inductive techniques as an option. Often inductive techniques bear clear advantages. In many domains, detailed knowledge is scarcer than exemplary data. Consequently, mechanistic and qualitative modelling approaches are not applicable, as these require a hypothesis to support such models. When both are missing, collecting appropriate data in many cases is less costly than acquiring the required process knowledge. Moreover, an explicitly formulated model with structural accuracy does not add value in many situations. Note, however, that an inductive approach depends on the identification of data attributes; in other words, a basic understanding of the domain is required, even if no full domain theory is available.

When the functionality of a model has been defined, exemplary data is the main requirement for inducing a model. When working on real-world applications, the required functionality does not always have a theoretical counterpart that may serve as the basis for a mechanistic approach. In the case of Example 6.1, fruit selection was a complete new function that did not exist before. Moreover, the classification question of separating prototypical fruits from atypical ones on the basis of external appearance had never been posed, not in post-harvest sciences and not in practice. Therefore expert knowledge was not available. Similarly in Example 6.3, the ruling practice in analysing functional identity of proteins is based on local or global sequence similarity. Again there was not an off-the-shelf domain theory available.

Not only will the inductive approach often be cheaper, it is also more flexible. If the process set-up changes, a new data acquisition phase will suffice for deriving a new model. Also, it is easier to include new insights into the application. A further development of the planning system of Example 6.1, developed for the support of a product distribution chain, uses a similar set-up of the planning process as illustrated for the system of Fig. 6.6. However, the planning process aims at a different product; consequently the models represent completely different mappings. Also, different inductive techniques have been used to realise the different tasks for both functional and non-functional requirements. Had a mechanistic approach been followed, then a new hypothesis would have been needed from fundamental product research. In the followed approach, renewing the inductive models and adaptation to the new product domain proved sufficient.

The flexibility of inductive models also appears from the ease of re-instantiating models. In numerous food processes, including raw produce, food processing and biotechnology, permanent innovation as well as development of

crops, logistics, feedstock and raw materials may lead to evolving behaviour of products. The quality development of fruits today is not the same as that of thirty years ago. Inductive models in general can easily adapt to such evolving behaviour. Moreover, several inductive techniques have adaptive variants available, where the learning set-up allows for permanent adaptation to evolving domains. This is, among others, the case in the so-called lazy learning techniques.

It is important to note that all the advantages of inductive systems in some domains will turn into disadvantages in other domains. Scientists may strive for insight in fundamental processes, and models are a useful means to get and refine such insight. This is illustrated in the application on cucumber quality development (Example 6.4),[43] where the main aim of the project was to understand mechanisms responsible for quality decay. The NN model as such would not have offered a sufficient structural explanation for the observed behaviour (although a cucumber grower would be satisfied by the result). Moreover, when models are to be applied in a domain with highly educated experts (e.g. biotechnology), structural accuracy of models may be of uttermost importance for the acceptance of the model outcome by experts (even if it can be proven that predictive accuracy of inductive models is as good).

6.4 Future trends

Some recent developments in agriculture and food processing will change the assessment of inductive approaches in modelling these processes. In science, new analysis techniques such as micro-arrays, DNA analysis, and techniques such as visual image processing generate bulk data.[44] Several modern logistic and management trends such as tracking and tracing, total quality management, and HACCP also generate bulk data. For both the scientific and application aspects, the theoretical understanding is scarce and not always applicable to support modelling.

Furthermore, there are several technical developments in the world of inductive and mechanistic modelling that may help the proliferation of inductive techniques in food process modelling. First of all, as argued in section 6.2.2, research interest in inductive modelling is shifting from a technology orientation towards a real-world application. One consequence of this shift is an increasing interest in the application methodology. Early experiences with application of inductive techniques illustrate the importance of structuring the application process. This means, among others, a complete design of the solution in which inductive techniques are to be applied, the definition of an interface between the inductive technique and the embedding system and rational technique selection based on data factors and application requirements.[33] Here, we see an emerging integration of inductive modelling techniques with principles from software engineering and knowledge acquisition.

Take all these developments together, and a further proliferation of inductive

Example 6.4 Modelling cucumber quality development[43]
Verification of the underlying data assumptions of a scientific model with a neural network.

Project motivation
A main quality indicator of cucumbers is the colour. When stored, enzymatic degradation of chlorophyll in cucumbers results in a colour change from dark green to light yellow. In most cases colour is the limiting quality factor. A research project tested the hypothesis that photosynthetic parameter measurement can help describe the process of colour change. Photochemical and photo system measurements are used as indicators of the quality. The keeping quality for cucumbers is defined as the time for a cucumber to reach a certain predefined colour limit. In an experiment 2000 cucumbers of different cultivars and growing conditions were stored and colour development was monitored. A simple enzymatic model was developed. To assess the hypothetical power of photosynthetic colour measurements, the cucumbers were divided into four keeping quality classes. Two different neural network models were developed, using both initial colour and photosynthetic measurements, respectively.

Model design
After initial data pre-processing, the data of approximately 500 cucumbers was suitable for neural network analysis. Two neural models were developed, one using initial colour only (NN-colour), and one using initial colour and photosynthetic measurement (NN-photosynthetic). For both models, the optimal design was determined in a ten-fold cross-validation experiment. This resulted in a two-layer network for NN-colour, with four input and four output nodes, and a three-layer network for NN-photosynthetic, with four input nodes, two input nodes, four output nodes. The neural network model NN-photosynthetic outperforms other models (both the neural and enzymatic models) in terms of the realised R^2_{adj} (0.84/ unknown v. 0.88). This indicates that photosynthetic measurements include the required information for proper classification. Also, it indicates that the enzymatic model needs further refinement.

techniques to new areas in food process modelling is to come. As indicated elsewhere in this chapter, potential application areas are found in the total production and distribution chain for agricultural produce, i.e. in growth, processing and distribution, but also as a tool to support and direct research. Relatively young areas such as microbiology, biochemistry and genetic techniques heavily rely on data processing. These areas are therefore especially interesting from an inductive technique perspective.

Finally, proliferation of inductive techniques also extends to standard development environments and analysis tools. Software environments for micro-array analysis offer neural network techniques as a standard option, statistical tools (e.g. SPSS) migrate towards data-mining, integrating inductive techniques from neural networks and machine learning, etc. In a sense, inductive techniques are subject to a paradox: the more developed the area becomes, the more it is integrated in standard available tools, and the less visible it is as a separate discipline. Being a benefit for the proliferation of inductive techniques, for the time being inductive modelling will have its own position within food process modelling, and its own merits.

A very promising development is the combination of inductive and mechanistic techniques in one single application. Several modes of application are possible. In the example on cucumber quality modelling, inductive and mechanistic models have been used in parallel. The mechanistic model implements a domain theory. Testing this mechanistic model results in a predictive accuracy. When the accuracy is reasonable, it is hard to assess whether the remaining inaccuracy is due to data quality problems, or that the model lacks quality. By applying an inductive model in parallel, the potential accuracy as present in the data set is assessed. When the performance of the mechanistic model stays substantially behind the performance of the inductive model, the quality of the mechanistic model can be questioned.

6.5 References

1. MITCHELL T, (1997), *Machine Learning*, McGraw Hill, New York.
2. MLNet, MLnet II Network of Excellence in Machine Learning – Training Committee, http://kiew.cs.uni-dortmund.de:8001/mlnet/index.htmlt
3. MLNet, Application Guide on Machine Learning, http://swi.psy.uva.nl/projects/MLAG/
4. MICHIE D, SPIEGELHALTER DJ, TAYLOR CC, (eds) (1994), *Machine Learning, Neural and Statistical Classification*, Ellis Horwood, on-line: http://www.amsta.leeds.ac.uk/~charles/statlog/indexdos.html
5. LANGLEY P, *Elements of Machine Learning*, Morgan Kauffman, San Mateo (CA).
6. ALMUALLIM H, DIETTERICH TG, (1991), Learning with many irrelevant features. *Proceedings of the Ninth National Conference on Artificial Intelligence* (AAAI-91) (pp. 547–52). Anaheim, CA: AAAI Press.
7. LIU H, MOTADA H, (1998), *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer Academic Publishers, Norwell (Ma).
8. LIU H, MOTADA H, (Eds.) (1998), *Feature Extraction, Construction and Selection, a Data Mining Perspective*, Kluwer Academic Publishers, Norwell (Ma).
9. SCHAPIRE RE, (1999) A Brief Introduction to Boosting, in: *Proceedings IJCAI-99*, Vol. 2, 1401–6.

10. LAWRENCE CE, ALTSCHUL SF, BOGUSKI MS, LIU JS, NUEWALD AF, WOOTTON JC, (1993), Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*, 262 (5131), 208–14.
11. KRAAKMAN A, (1998), Application of Machine Learning in Plant Breeding: The Process of Putting a Learning Technique into Practice, in: Verdenius F. and van den Broek, W.H.A.M. (eds), *Proceedings of Benelearn 98*, 148–56.
12. DEHASPE L, TOIVONEN H, KING RD, (1998), Finding Frequent Substructures in Chemical Compounds, in: *Proceedings of KDD-98*, New York, 30–6.
13. BREIMAN L, FRIEDMAN JH, OLSHEN RA, STONE CJ, (1984), *Classification and Regression Trees (CART)*. Monterry, CA: Wadsworth & Brooks.
14. WEISS SM, KULIKOWSKI CA, (1991), *Computer Systems that Learn*, Morgan Kaufmann, San Francisco (CA).
15. RUMMELHART DE, HINTON GE, WILLIAMS RJ, (1986), Learning Internal Representations by Error Propagation, in: D.E. Rumelhart, J.L. McClelland and the PDP Research Group, *Parallel Distributed Processing*, MIT Press, Cambridge (Ma), 318–62.
16. FAYYAD U M, PIATESKY-SHAPIRO G, SMYTH P, (1996), From Data Mining to Knowledge Discovery: An Overview, in: Fayyad U.M., Piatetsky-Shapiro, G. Smyth, P. and Uthurusamy, R. (eds.), *Advances in Knowledge Discovery and Data Mining*, AAAI-Press, 1–37.
17. YANG XB, BATCHELOR WD, 1997, Modeling plant disease dynamics with neural networks. *AI Applications* 11(3) 47–55.
18. HEITKOETTER J, BEASLEY D, (eds) (1994), The Hitch-Hiker's Guide to Evolutionary Computation: A list of Frequently Asked Questions (FAQ), USENET: comp.ai.genetic. Available via anonymous FTP from rtfm.mit.edu:/pub/usenet/news.answers/ai-faq/genetic/ About 90 pages.
19. INGBER L, Adaptive Simulated Annealing, http://www.ingber.com/
20. KOLODNER J, (1993), *Case-Based Reasoning*, Morgan Kaufmann, San Mateo (CA).
21. VERDENIUS F, BROEZE J, (1997), Resource Allocation in Waste Water Plants, in*: Proceedings of ESIT'97*, Aachen, Germany, 55–9.
22. AHA D, KIBLER D, ALBERT M, (1991), Instance-Based Learning Algorithms, *Machine Learning 6*, 37–66.
23. VERDENIUS F, (1996), Managing Product Inherent Variance During Treatment, in: *Computers and Electronics in Agriculture* 15, 245–65.
24. AARTS RJ, (1995), Computer Planned Recipes for Effective Production Management, in: *Proceedings of EBC-95*.
25. HUNTER L, (1993), Planning to Learn about Protein Structure, in: Hunter, L. (ed), *Artificial Intelligence and Molecular Biology*, AAAI/MIT Press, ISBN 0-262-58115-9.
26. SHAH I, HUNTER L, (1998), Identification of Divergent Functions in Homologous Proteins by Induction over Conserved Modules, in: *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*, 157–64.

27. COIL competition, http://www.dcs.napier.ac.uk/coil/challenge/
28. KDD-Cup, http://www.ecn.purdue.edu/KDDCUP/
29. Machine Learning Toolbox, ESPRIT project 2154, ftp://ftp.gmd.de/machinelearning/public/
30. KODRATOFF Y, MOUSTAKIS V, GRANER N, (1994), Can Machine Learning Solve My Problem? in: *Applied Artificial Intelligence*, 8 1–31.
31. SLEEMAN D, (1994), Towards a Technology and a Science of Machine Learning. *AI Communications*, 7(1) 29–38.
32. ADRIAANS P, ZANTINGE D, (1996), *Data Mining*, Addison & Wesley Longman Ltd, Harlow (GB).
33. BRODLEY C E, SMYTH P, (1997), Applying Classifications in Practice, in: *Statistics And Computing*, 7, 45–56.
34. VERDENIUS F, ENGELS R, (1997), A Process Model for Developing Inductive Applications, in: W. Daelemans, P. Flach and A. van den Bosch (eds), *Proceedings of Benelearn-97*, Tilburg University (NL), 119–27.
35. ENGELS R, (1999), *Component-Based User Guidance in Knowledge Discovery and Data Mining*, PhD thesis, University of Karlsruhe.
36. VERDENIUS F, VAN SOMEREN M W, (1999), Top-down Design and Construction of Knowledge-Based Systems with Manual and Inductive Techniques, *Proceedings of KAW 99*, Banff, 4-15-1–4-,15–20.
37. VERDENIUS F, VAN SOMEREN M W, (1997), Applications of Inductive Techniques: a Survey in the Netherlands, in*: AI Communications*, 10 3–20.
38. LIM T S, LOH W Y, SHIH Y S, A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms, *Machine Learning*, 40 (3) 203–28.
39. Benelearn '99 competition: http://www.swi.psy.uva.nl/benelearn99/comppage.html
40. BRAZDIL P, GAMA J, HENERY B, (1994), Characterising the Applicability of Classification Algorithms using Meta-Level Learning, in: *Proceedings of ECML 94*, 83–102.
41. ENGELS R, THEUSINGER C, Using a Data Metric for Offering Preprocessing Advice for Data Mining Applications, in: H. Prade (ed*), Proceedings of ECAI 1998*, John Wiley & Sons, Chichester, 430–4.
42. VERDENIUS F, (1999) Entropy Behavior for Selection of Machine Learning Techniques, in: H. Blockeel and L. DeHaspe, *Proceedings of Benelearn 1999*, 113–20.
43. SCHOUTEN R E, OTMA E C, VAN KOOTEN O, TIJSKENS L M M, (1997), Keeping Quality of Cucumber Fruits Predicted by Biological, in: *Postharvest Biology And Technology* 12 175–81.
44. VAN DER BROEK W H A M, NOORDAM J C, PAULI A, (2000), Multivariate Imaging for Automated Process Control in the Agro Industry, in: *Proceedings of Agricontrol 2000*, Wageningen (NL), July, IFAC, 231–5.
45. QUINLAN J R, (1993), *C4.5:Programs for Machine Learning*, Morgan Kaufmann, San Mateo (CA).

7

Data mining

G. Holmes and T. C. Smith, University of Waikato, Hamilton

7.1 Introduction

7.1.1 What is data mining?

Data is often recorded in spreadsheets and databases as a set of facts. In food process modelling those facts could come from an experiment designed to determine if a hypothesis holds or to find out if certain measurements contribute to a given outcome. Either way, researchers record data and attempt to extract information, a set of patterns or expectations that underlie the data.

Sometimes you may know what you are looking for and as such you will not be doing data mining because data mining involves the automatic extraction of implicit, previously unknown, and potentially useful information from data. It is particularly relevant, therefore, when applied to very large databases, as evidenced by its growing popularity with database researchers and large institutions with terabyte size collections of data.

In order to discover and evaluate patterns in data, data mining employs a wide range of techniques from machine learning, statistics and databases. In what follows we will attempt to present some of the most widely used techniques from these areas.

7.1.2 Why do data mining?

Serious applications of data mining involve thousands or millions of examples but in order to demonstrate some data mining techniques it is instructive to work with a small fictitious data set (adapted from Quinlan[1]).

Consider a series of subjective and objective measurements related to mushroom quality. Weight and firmness measured with appropriate instruments

Table 7.1 The mushroom data

Weight	Stalk damage	Dirt	Firmness	Quality
heavy	high	mild	hard	poor
heavy	high	mild	soft	poor
normal	high	mild	hard	good
light	medium	mild	hard	good
light	clear	clean	hard	good
light	clear	clean	soft	poor
normal	clear	clean	soft	good
heavy	medium	mild	hard	poor
heavy	clear	clean	hard	good
light	medium	clean	hard	good
heavy	medium	clean	soft	good
normal	medium	mild	soft	good
normal	high	clean	hard	good
light	medium	mild	soft	poor

and dirt, stalk damage and quality assigned by expert mushroom inspectors. These measured variables constitute features or attributes of the data. The complete data set is presented in Table 7.1 and we will make use of it throughout this chapter. Our aim is to construct a model (set of patterns) of the data that accounts for each of the fourteen examples, called instances, in the table. One such model might be comprised of a set of rules describing relationships between attributes in each instance in the data set. The first line of Table 7.1, for instance, could be expressed with the rule

> if the mushroom is heavy and has high stalk damage and is mildly dirty and is firm then its quality is poor

and a comprehensive account of the entire data set could be formulated as a list of rules describing each instance. The goal of data mining, however, is to try and capture relationships between attributes in a more general and useful way. It may be that the quality of a mushroom in these instances depends only on, say, its weight and firmness, and that one or two very general rules are sufficient to characterize the complete set of instances.

The objective of data mining is thus to derive automatically a model that will tell us things about the data, such as which variables are more important than others in determining the quality of a mushroom. We could find out if the objective measurements are as good as the subjective measures in determining the quality. Finally, we could use the model to make predictions about the quality of mushrooms whose weight, stalk damage, dirt and firmness are known.

7.1.3 How is data mining done?

A key concept in finding a model that describes and generalizes some data is the notion of search. If we produce rules as the set of patterns for Table 7.1 then

there are $4 \times 4 \times 3 \times 3 \times 2 = 288$ possibilities for each rule. That is, a rule can involve any of the three possible values for Weight (heavy, normal or light) or it may be an irrelevant attribute, thus there are four ways it might factor into a rule. Similarly, there are four ways Stalk damage might be used, and three each for Dirt and Firmness, giving 288 permutations for combining these attributes within each rule.

Ideally, a single rule may be sufficient to characterize all fourteen of the instances, but it may turn out that no generalization is possible and each instance requires its own rule. This means there are 288^{14} (about 10^{34}) possible rule sets (i.e. models) that must be tested in an exhaustive search for the best one. Even the fastest computers available cannot perform searches of this magnitude in a reasonable amount of time, making it necessary to use heuristics as a guide to find a good, but very likely not optimal, solution.

Because choices have to be made in searching for the best description a bias will inevitably be introduced. This search bias is one way in which data mining tools differ. Further, different data sets may be more suited to one method over another. Consequently, many data mining tools offer a suite or workbench of methods for processing the same data.

7.1.4 Data mining in food process modelling

Data mining is an emerging technology that has not penetrated the agricultural sciences to any great extent. As this book demonstrates, there are a number of modelling techniques that can be used to analyse data and so with education and the availability of software it seems likely that these techniques will form part of the agricultural researcher's armoury.

In machine learning, one of the earliest successes was the identification of rules for diagnosing soybean diseases by Michalski and Chilausky.[2] Since then there have been some applications, for example, Holmes *et al.*,[3] who have applied inductive techniques to determine the factors that influence apple bruising and more recently a study of grower and consumer perceptions of mushroom quality (Bollen *et al.*[4]). However, the uptake of this technology in the agricultural domain has not matched the plethora of applications in business.[5]

7.1.5 Benefits for food process modelling

Data mining techniques offer scientists in food process modelling a number of benefits. It should be emphasized that the technology is not a panacea, but can be used to determine the relative importance of features (perhaps measurements from an experiment). It is possible to determine if an attribute is redundant and if there is noise in data which could assist experimental design. The models induced from data often quantify a relationship quite precisely, especially when dealing with numeric attributes (see Chapters 2 and 3). Again this can provide valuable insight.

7.2 Input characteristics

7.2.1 Raw data

Table 7.1 contains fourteen instances which each contain characteristic values from four attributes (independent variables) and a special attribute called the class attribute (dependent variable). From these instances we hope to induce a concept description which implies that in some sense the data defines collectively a concept to be learned.

The attributes in Table 7.1 are all nominal in that their values come from a finite set. Each attribute could have been numeric, corresponding to perhaps the mushrooms' real weight and firmness with a percentage figure for the amount of dirt and stalk damage. Similarly, the class attribute is nominal in this case and as such we define the concept learning task to be classification learning. Had the class attribute been an integer or real value then the concept learning task would have been numeric learning. Both forms of learning will be discussed in this chapter.

7.2.2 Data preparation

For most agricultural scientists their experimental data will be stored in either a spreadsheet or database. Spreadsheets encourage exactly the format that most learning algorithms require; rows of instances and columns of attribute values. Databases are more work in the sense that they typically store data across several files. In order to apply a learning algorithm it is often necessary to first apply a query to retrieve a single file of data to learn from. This is where database research comes into the data mining picture – extending existing database query languages to include data mining techniques and developing new data models that provide better support for data mining.

Whatever the source of the data it is often necessary to attempt to *clean* it up before trying to learn from it. This might mean detecting and removing parts that are missing or erroneous. For databases, the process of gathering data together from different sources can be challenging, time consuming and prone to error. Great care must be taken at this stage because the quality of the data has a high correlation with a successful outcome.

7.2.3 Data problems

Data is not always collected perfectly. Measurements can be made by faulty equipment, the design of the experiment can change during its course, or fields within a survey may not be filled in by all respondents. In this last case we are faced with the problem of missing values. Most data mining tools anticipate that data will be missing and each deals with such occurrences differently. They may, for example, replace a missing value with the most commonly observed value for that attribute, or they may treat the missing value as a separate value altogether. Either way it is important to acknowledge the reasons why data is missing and reflect on any models produced under such circumstances.

Data can be erroneously entered into a spreadsheet or database and this generates 'noise' that can lead to problems. The noise can be 'attribute noise' due to incorrect values being recorded for attributes or 'class noise' where two instances with the same attribute values receive different class labels. Data needs to be checked for typographical errors, values outside the reasonable range for an attribute, duplicate entries or attributes in the data that may have no connection with the concept that is being learned. Data mining tools often provide graphical means, such as histograms and sorted plots of numeric ranges, for examining data – typically on an attribute or attribute pair basis, and as the next section demonstrates, there are data engineering tools that can prepare data automatically.

7.3 Data engineering methods

Techniques to improve the chances of a successful outcome from a data mining application by manipulation of the data prior to learning are commonly described as data engineering methods.

7.3.1 Attribute selection

If some of the recorded attributes are irrelevant or redundant, or the data is noisy and unreliable, then the concept learner can become confused and may struggle to determine the true underlying concept. Attribute selection is the process of identifying and removing as much of the irrelevant and redundant information as possible (see Chapter 2). Some learning algorithms perform attribute selection as part of the learning process but even then it is still possible to improve learning by first removing attributes.

Attribute selection methods can be classified into two groups: wrapper methods and filter methods. Wrapper methods[6] use a learning method and a statistical re-sampling technique to evaluate the usefulness of attributes. A search technique is employed to guide the selection of attributes. One wrapper method, called forward selection, starts with an empty set of attributes then generates all single attribute models with the learning method and evaluates their performance. The attribute with the best performance is chosen and then combined with all the rest, and the procedure is repeated to find the best pair of attributes, then the best three, and so on. If there is no improvement from the best combination at any stage, then the search can be terminated or the next best can be revived and the process continued from there. While useful, forward selection is extremely slow and does not scale well to larger problems.

An alternative and less expensive approach to attribute selection is to use a filter. Filter methods[7] operate independently of the learning algorithm by dealing with undesirable attributes before learning commences. They work in a variety of ways: using heuristics to rank attributes according to relevancy; eliminating attributes whose information content is subsumed by others; calculating the correlation between an attribute and the class and the correlation

between the attributes themselves; and so on. Most filter methods work quickly through the data, typically returning a substantially smaller subset of the attributes for the learning algorithm to work with.

7.3.2 Discretization

Numeric data can present problems for some data mining tools. Some tools do not deal with numeric data while others make assumptions about the distribution of the data that may not hold. In such cases it is necessary to discretize the data into a number of ranges such that each range can be treated as a nominal value.

Discretization can be unsupervised, that is, the correlation between an attribute value and a class label cannot be taken into account – this is the case when clustering data. For all other forms of learning the class labels are available and so supervised discretization can be employed. Unsupervised methods include dividing the range of values into a fixed number of 'bins', or on the basis of having the same number of instances in each range. Both methods are problematic but can still yield good results in practice.

Successful discretization is more likely when the class label is available and several sophisticated entropy-based techniques have been developed to take advantage of this situation.[8]

7.4 Output representations

An understanding of how knowledge is represented and interpreted is a useful insight into how data mining works. Below we discuss the most popular forms of output representation for those methods that produce output in section 7.5. We assume that the data from Table 7.1 has been used as input and show the various forms that the results of learning can take.

7.4.1 Decision trees

One method of learning is to determine a test on an attribute that will discriminate between the instances in a data set. This test has the effect of splitting the data into two (or more) smaller data sets. By repeatedly splitting these smaller data sets we can obtain a series of attribute tests that represent the original data as a tree (Fig. 7.1). Note that although the mushroom data has only nominal values, numeric values can be handled with tests on ranges of values, for example, Weight >3.0. The output representation of Table 7.1 has achieved data reduction and covers all fourteen instances. The figure shows that weight is the most important attribute for discriminating between samples. All mushrooms with normal weight are classified as good mushrooms. The classification of heavy mushrooms depends on how dirty they are, mild are poor mushrooms whereas clean are good. Light mushrooms, on the other hand, require their firmness to be determined before a classification can be made.

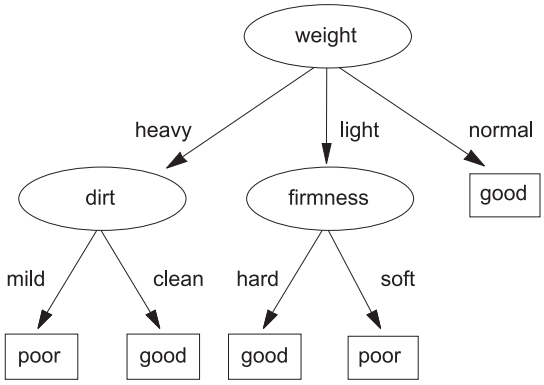


Fig. 7.1 Decision tree for the mushroom data.

This output representation encapsulates knowledge and can be used to classify 'new' or 'unknown' mushrooms. Suppose we were given values for the four variables for a mushroom which was not one of the original fourteen used to build the model and asked to classify it, say, for example: light, high, mild, soft. The model predicts that this is a poor mushroom (light and soft). Decision trees can contain the repeated use of the same attribute further down in the tree – all variables are considered at every level of the tree.

7.4.2 Classification rules

An alternative to a decision tree is to represent knowledge as a set of rules. Rule sets can be either induced directly or produced from a decision tree. For example, the tree in Fig. 7.1 can be readily transformed into the rule set in Fig. 7.2 by generating a rule for each of the leaves of the tree. The conditions of the rule are the tests encountered from the leaf to the root of the tree (combined as conjunctions) and the classification given by the rule is the class label of the leaf.

Decision trees form well-behaved rule sets in that they are mutually exclusive. Each rule will only fire once for each instance. In general, rules may overlap (provide multiple classifications) requiring a vote to be taken on the classification or taking the classification given by the first rule that is triggered. These strategies will clearly lead to different results. Rule sets can establish default rules that are applied when all others fail. These rules can make the rule representation more compact than, say, a decision tree because the default rule 'tidies up' large numbers of instances not covered elsewhere.

7.4.3 Association rules

So far we have assumed a situation where a number of independent attribute values are combined in a model involving tests on those values to establish a

```
If weight = heavy AND dirt = mild Then Quality = poor
If weight = heavy AND dirt = clean Then Quality = good
If weight = light  AND firmness = hard Then Quality = good
If weight = light  AND firmness = soft Then Quality = poor
If weight = normal Then Quality = good
```

Fig. 7.2 Rule set for the mushroom data.

```
dirt=clean firmness=hard 4 ==> quality=good 4 (1)
stalk_damage=clear 4 ==> dirt=clean 4 (1)
weight=normal 4 ==> quality=good 4 (1)
stalk_damage=clear quality=good 3 ==> dirt=clean 3 (1)
weight=light firmness=hard 3 ==> quality=good 3 (1)
weight=light quality=good 3 ==> firmness=hard 3 (1)
weight=heavy dirt=mild 3 ==> quality=poor 3 (1)
weight=heavy quality=poor 3 ==> dirt=mild 3 (1)
stalk_damage=clear firmness=hard 2 ==> dirt=clean quality=good 2 (1)
stalk_damage=clear dirt=clean firmness=hard 2 ==> quality=good 2 (1)
```

Fig. 7.3 Apriori associations for the mushroom data.

relationship with a class label. Association rules generalize this idea to include relationships between single or composite attributes. This notion gives rise to a very large number of possible attribute combinations. From this large set we are interested in those associations that cover or support a large number of instances.

Accuracy or confidence is established for an association by calculating the proportion of the total number of instances that a rule predicts correctly. The Apriori algorithm[9] is the most commonly used association rule learner. Figure 7.3 shows some of the 'best' sets of associations using Apriori for the mushroom data of Table 7.1. The number before the implication symbol ==> is the support (in the first rule, the number of instances covered by dirt = clean and firmness = hard). The first number at the end of the rule is the number of instances that also satisfy the consequent (quality = poor) and the number in brackets is the confidence (the number after the consequent divided by the support). The rules are ordered first by confidence and second by support. The first association rule of Fig. 7.3 is a 'standard' classification rule whereas the second associates stalk damage with dirt. The final two rules show how composite associations can be drawn.

7.4.4 Numeric prediction representations

Let us suppose that the mushroom data quality was determined on an integer scale rather than a binary scale. Methods for predicting numeric quantities perform regression. The simplest establish a linear relationship between the attributes of a problem and a numeric quantity by estimating coefficients for each attribute. This idea can be combined with a decision tree to produce a model tree.[10] Model trees have standard decision tree tests at interior nodes of the tree and linear models at the leaves.

```
IF Energy Level = 1 OR Energy Level = 2 OR Energy Level = 3 THEN
        IF Energy Level = 1 THEN LM1
                IF Energy Level = 2 OR Energy Level = 3 THEN LM2

IF Energy Level = 4 OR Energy Level = 5 OR Energy Level = 6 THEN
        IF Location = 2 THEN LM3
                IF Location = 1 OR Location = 3 OR Location = 4 THEN
                        IF Energy Level = 4 THEN LM4
                        IF Energy Level = 5 THEN LM5
                        IF Energy Level = 6 THEN LM6

LM1:  Bruise = 0.251 + 0.193 (Location=1 or 3) + 0.0936 (Location=1)
LM2:  Bruise = 0.88 + 0.119 (Location=1 or 3) + 0.277 (Energy Level=3 or 4 or 5 or 6)
LM3:  Bruise = 1.62 + 0.2 (Energy Level=6)
LM4:  Bruise = 1.75 + 0.101 (Location=1)
LM5:  Bruise = 2.18 – 0.181 (Location=1 or 3)
LM6:  Bruise = 2.11 + 0.162 (Location=1 or 3) – 0.0853 (Location=1)
```

Fig. 7.4 Rules generated from a model tree for an apple bruising data set.

Using the technique described earlier it is possible to generate a set of regression rules from the model tree. An example set of rules together with their linear models is shown in Fig. 7.4 for an apple bruising problem[3] – unfortunately, there is insufficient data in Table 7.1 to induce a useful set of rules. Interpreting the first rule with linear model LM1, if the energy level is one (equivalent to a drop height of 10 mm) the apple is predicted to bruise to a maximum of 0.251 when dropped on locations 2 and 4 (calyx shoulder and large cheek, respectively). Apples bruise to a maximum of 0.251 + 0.193 if dropped on location 3 (small cheek), and when dropped on their stem shoulder we get the largest degree of bruising (0.251 + 0.193 + 0.0936).

7.5 Data mining methods

Many algorithms have been developed for data mining, and what follows is a survey of some of the basic methods. More advanced techniques are available, along with optimization schemes to deal with possible complications that might arise when analysing real data sets, but it can be surprising how often even the simplest approach produces a good result.

7.5.1 Rule induction

One quick and easy way to formulate classification rules for a given set of instances is to identify the attribute with the highest co-occurrence rate with each class and generate a corresponding one-level decision tree. This simple algorithm, called 1R,[11] produces a very small set of rules that is often quite good at making accurate classification decisions for even quite large data sets, presumably because underlying relationships in real data are frequently uncomplicated. The

Table 7.2 1R rule sets for the mushroom data

Attribute	Rule	Error rate	Total error rate
Weight	heavy → poor	2/5	4/14
	normal → good	0/4	
	light → good	2/5	
Stalk damage	high → poor	2/4	5/14
	medium → good	2/6	
	light → good	–	
Dirt	mild → poor	3/7	4/14
	clean → good	1/7	
Firmness	hard → good	2/8	5/14
	soft → poor	3/6	

algorithm proceeds by considering each value of each attribute and counting how often it co-occurs with each distinct class attribute value. A tentative rule is formed which predicts the most frequent class on the basis of the attribute value in question, and then an error rate is calculated for that rule. After all values for all attributes have been trialled, the rules with the lowest error rates are chosen.

A good way to see how 1R works is to apply it to the example data in Table 7.1. When the value of Weight is 'heavy', the class 'poor' occurs three times while 'good' occurs twice; thus a rule is formed to say 'if Weight is heavy then Quality is 'poor' and its probability of error is calculated as 2/5 for the two times it is wrong out of the five times it can be tested. The procedure then iterates using the next value for Weight to produce the rule 'if Weight is normal then Quality is good' with an error rate of zero.

Table 7.2 summarises the complete 1R rules for the mushroom data. Note that the rules for high Stalk damage and soft Firmness have fifty percent error rates, in which case the complement class can be predicted with equal accuracy. When two rules perform equally well then either one can be chosen. The same goes when it comes time to select the final rule set. Decision rules for both Weight and Dirt have equally low error rates so either can be retained. Thus the very simple set of rules

Dirt: mild → poor
clean → good

is sufficient to predict the quality of a mushroom correctly more than 71% of the time.

7.5.2 Statistical models

1R derives a rule set which makes classification decisions based on just one attribute. It is quite possible to factor all attributes into the decision-making process using a simple statistical approach based on Bayes' rule of conditional probability. The idea is to assign probabilities to each class in light of the

Table 7.3 Independent probabilities for the mushroom data

Weight			Stalk damage			Dirt			Firmness			Quality	
	poor	good		poor	good		poor	good		poor	good	poor	good
heavy	3/5	2/9	high	2/5	2/9	mild	4/5	3/9	hard	2/5	6/9	5/14	9/14
normal	0/5	4/9	medium	2/5	4/9	clean	1/5	6/9	soft	3/5	3/9		
light	2/5	3/9	clear	1/5	3/9								

combined evidence afforded by all other attributes, then select the most likely classification.

The algorithm proceeds by counting how many times each value of each attribute occurs with each class value. The independent probability for a particular attribute-value/class-value pair is estimated as the number of times the attribute value is observed with the class value divided by the number of times the class value is observed.

Table 7.3 summarizes the independent probabilities for the mushroom data of Table 7.1. We observe that poor Quality occurs for five of the fourteen instances, and for three of those instances Weight is heavy, thus the probability of a mushroom being of poor Quality when it is heavy is 3/5. Note that the probabilities under the Quality attribute are simply the number of times each value is observed out of the total number of instances.

Given Table 7.3 it is now possible to estimate the likelihood that a novel instance belongs in any class. Consider the situation where a new mushroom instance is encountered with the following attribute values:

Weight = light AND Stalk damage = high AND Dirt = mild AND Firmness = soft

The likelihood that this instance is of poor Quality is estimated as the product of the independent probabilities for each of its attribute values co-occurring with poor Quality, or

$$\text{likelihood of poor Quality} = 2/5 \times 2/5 \times 4/5 \times 3/5 = 0.0768$$

and the likelihood that its Quality is good is estimated as

$$\text{likelihood of good Quality} = 3/9 \times 2/9 \times 3/9 \times 3/9 = 0.0082.$$

A comparison of these two values indicates that the new instance is much more likely to be of poor Quality than good. As it is usual for all possible outcomes to have combined probability of 1, the likelihood estimates can be normalized to probabilities by dividing each by the sum, giving

$$\Pr[\text{poor}] = 0.0768 / (0.0768 + 0.0082) = 90.35\%$$
$$\Pr[\text{good}] = 0.0082 / (0.0768 + 0.0082) = 9.65\%.$$

This technique is called *Naïve Bayes* because it assumes all attributes are independent of each other and of equal importance when making a classification decision. Such an assumption is seldom valid with real data sets, but the method

nevertheless does quite well in practice. Note that there is some danger that some attribute-value pairs will not be observed in the sample data, like normal Weight with poor Quality in the mushroom data. This results in a zero probability, and because any product of probabilities that includes a zero term will itself be zero this creates the possibility that even overwhelming evidence from other attributes may be completely discounted. The problem is well known in Bayesian analysis and is avoided by assigning very small probabilities to all unseen events.

7.5.3 Decision tree induction

Decision trees can be developed directly from data using a 'divide and conquer' approach. One-level decision trees are constructed for each attribute, splitting the data into groups according to their values for that attribute. The best of these is retained and another set of one-level decision trees is constructed for each of its groups using each of the remaining attributes, and so on. Only groups with more than one class value in them are split, and the process stops when no further splitting is possible. This is the basic idea behind the ID3 induction procedure[12] and the C4.5 system.[1]

Figure 7.5 shows the initial four decision trees that result when each attribute of the mushroom data is trialled. Only one of these is to be refined further if the algorithm is to avoid an exhaustive search. The question remains then, how can we tell which is the best one? Any group that has more than one class value represented in it must be split. But the ultimate goal is to have as shallow a tree as possible, meaning as few splits as possible. Therefore we select the tree whose groups are the most 'pure' – that is, groups that require the least number of additional splits to get all class values into separate subgroups.

It is not obvious how one might measure the purity of a group, but it turns out it can be done by determining how much information is gained by each split. This is much like the game of twenty-questions, where some number of yes-no questions must be asked to determine what is unknown at the outset. Some questions are better than others and lead you to the right answer more quickly than less prudent ones. A good question is generally one that reveals the most information and therefore reduces our search space the most. So it is with decision tree induction that we select the attribute that best partitions the data so that the fewest number of additional questions must be asked to determine the class value exactly.

The exact formula for calculating information gain is the entropy function from communications theory, and it is beyond the scope of this brief survey to outline the details. For more information, the reader should consult Shannon and Weaver.[13]

7.5.4 Covering rules

The decision tree and rule induction methods outlined above attack the problem of data mining by combining attribute tests together and testing the accuracy of

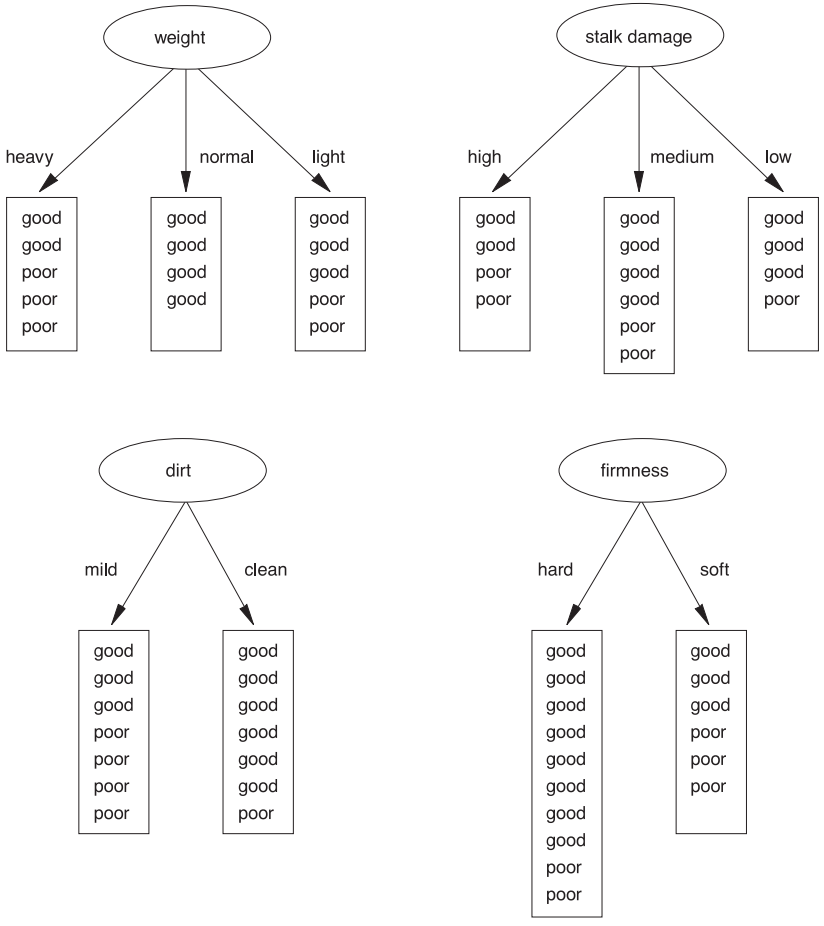


Fig. 7.5 Initial one-level decision trees for the mushroom data.

the ensuing prediction. An alternative strategy is to start with the desired prediction and add attribute tests in such a way that the accuracy of the prediction is kept as high as possible. This approach is called *covering* because rules are constructed to cover as many instances of the data as possible.

For the mushroom data, we want to form rules to cover the two possible classifications: good and poor. Thus we might begin by seeking a rule of the form

if ? then Quality = good.

There are ten possible attribute tests we could use to complete this covering rule. These are listed in Table 7.4 along with the accuracy each would provide given the evidence of the fourteen instances.

To complete the rule we simply choose the test that gives the highest number of correct classifications. If more than one test gives the best result, as

Table 7.4 Possible attribute tests to complete the initial covering rule

Attribute test	Accuracy
Weight = heavy	2/14
Weight = normal	4/14
Weight = light	3/14
Stalk damage = high	2/14
Stalk damage = medium	4/14
Stalk damage = clear	3/14
Dirt = mild	3/14
Dirt = clean	6/14
Firmness = hard	6/14
Firmness = soft	3/14

in this case, then the choice is arbitrary. Thus our covering rule can be completed as

if Firmness = hard then Quality = good.

This rule is the best we can make from a single test, but it is not all that good. It makes the correct classification for fewer than half the instances, and of the eight times the condition is true it gives the wrong classification twice. To remedy incorrect classifications the algorithm tries to find another condition that can be added to give a more accurate rule of the form

if Firmness = hard AND ? then Quality = good.

There are eight other attribute tests available to complete the rule and these are listed in Table 7.5 along with their accuracy levels for the eight instances where the first test is successful. We choose the best of these to complete the covering rule as:

if Firmness = hard AND Dirt = clean then Quality = good.

This rule applies to four of the nine instances in the sample data where Quality is good, and it gives the correct classification each time. To complete the rule set

Table 7.5 Attribute tests that can be added to the covering rule

Attribute test	Accuracy
Weight = heavy	1/8
Weight = normal	2/8
Weight = light	3/8
Stalk damage = high	2/8
Stalk damage = medium	2/8
Stalk damage = clear	2/8
Dirt = mild	2/8
Dirt = clean	4/8

we remove those four instances and repeat the process. Once an entire class is covered, a new search is started for the next class value using the complete data set.

This method of rule construction underlies the PRISM algorithm[14] but because it only generates rules that are 100% accurate it is susceptible to generating large overfitting rules when the training instances contain errors. INDUCT[15] is a modification to the algorithm that uses a statistical measure of 'goodness' to produce approximate rules from noisy data.

7.5.5 Association rules

The algorithms outlined above address the problem of finding classification rules; that is, they seek to characterize the conditions under which a particular class can be predicted based on the values of other attributes. Often the objective of data mining is more uncertain that this, and the goal is simply to discover which attributes exhibit dependencies or structural patterns of any type. In the case of the mushroom data, for example, it may be the case that heavy Weight coincides with high Stalk damage, or mild Dirt frequently accompanies soft Firmness, and uncovering associations like these can suggest to the data miner possible causative relationships.

One way to find association between other attributes is simply to ignore the class attribute and treat one of the others as the class, then use one of the algorithms given above. A more efficient method is to look for interesting *item sets*, a term that derives from the idea of items in a supermarket shopping basket. An item is an attribute-value pair and an item set is some number of items. The APRIORI[16] algorithm looks for item sets with notable frequency and translates them into association rules. How frequently an item set must be observed in order for it to be 'notable' is a parameter to the algorithm, but it must occur twice at least.

The algorithm proceeds by generating all one-item sets, then all two-item sets, then all three-item sets, and so on, where the same attribute is not allowed to appear more than once in any item set. Item sets with more than one item that occur more than once are translated into association rules where some number of items in the set predict the others, then the rules are tested against the data for accuracy.

While this sounds like it requires an exhaustive search of all possible attribute-value combinations, there is a fundamental principle that greatly reduces the number that must be considered. An n-item set can only occur more than once if all of its $(n-1)$-sets also occur more than once. Candidate n-item sets are obtained simply by taking the union of all pairs of $(n-1)$-item sets and keeping those with n items.

Association rules are often useful when seeking an initial characterization of a very large data set without class values. Large databases can lead to excessive computation, but this can be reduced by increasing the minimum frequency required for an item set to be considered interesting.

7.5.6 Mining from numeric data

The data mining techniques described above work well for data sets with nominal values, and can be applied to numeric data after discretization. But there are algorithms that are ideally suited to numeric data.

Linear regression is a standard statistical method that works well for data mining. The idea is to apply a membership function that calculates the class for a given instance using a linear combination of its attributes. For n attributes, a set of n weights is defined to produce a formula of the form

$$f(x) = w_0 + w_1 a_1 + w_2 a_2 + \ldots + w_n a_n$$

where $f(x)$ is the membership function, a_1, a_2, …, a_n are the attributes of instance x, and w_0, w_1, …, w_n are weights. The weights are calculated from the training data by finding coefficients that minimize the sum of the squares of the differences over all instances – an operation that is standard in commercial statistics software.

Another good data mining technique for numeric data is *instance-based learning*. The underlying idea is to treat instances as points in multi-dimensional space with the attribute values as their coordinates. To classify a novel instance, the algorithm calculates its Euclidian distances to the other instances and assigns it the same class as its nearest neighbour. Because some attributes may be assigned values from much greater scales than others, they have the potential to dominate the distance calculation. It is therefore necessary to normalize attribute values to between zero and one with respect to where they lie on a number line between the minimum and maximum values observed for that attribute.

7.6 Output evaluation

The objective of data mining algorithms is to develop a characterization of data sufficient to identify hidden relationships or support future decisions. Individual algorithms may vary in terms of the results they give; thus an important step in data mining is evaluating the usefulness of the result.

7.6.1 Predictive accuracy

It is quite natural to measure the performance of a classification scheme simply by calculating its rate of error or success, say, as a percentage of the number of times it assigns the correct class to novel instances. But there are a number of problems that can arise. For one thing, new instances may not be in large supply, and by the time enough evidence has been seen to indicate something has gone wrong it may be too late. Furthermore, the success rate is usually only an approximate measure of a model's accuracy. How close is close enough for a classification scheme? For example, if a model is inferred to explain the outcome of tossing a fair coin, but during testing a rare sequence of ten consecutive 'heads' is observed, should the model be dismissed as invalid? And

what if the training data is rife with exceptional instances? How can we tell the difference between 1) a model trained on poor data and tested on good data, and 2) a model trained on good data and tested on poor? To avoid this it is usual to express the estimate of success in terms of a confidence level, just like we would for any conclusion based on limited observations.

7.6.2 Cross-validation

There are two aspects to data mining that can influence evaluation of the outcome from any given learning algorithm: training and testing. In the best of all possible situations there would be a large representative data set available for training and another for testing. In practice there is just one big data set that has to be partitioned into a training set and test set.

There is always some danger that the data is not partitioned well, such that either the training data or the test data contains a high number of unusual instances, leading to poor performance either way. One way to guard against this is to train with one data set and test with the other, then swap the data sets, re-test and compare the results. If the results are significantly different the data can be partitioned another way and the experiments repeated. This procedure is called *cross-validation* and it is an important statistical technique for avoiding accidental bias in the data.

7.6.3 N-fold cross-validation

A more general technique for mitigating the possibility of bias is to divide the data into three or more separate sets. One set is put aside for testing and the remaining sets are combined to form the training set. Cross-validation tests are performed by repeating the experiments with each subset being withheld for testing. This technique is called *n-fold cross-validation* where *n* is the number of data subsets. Empirical studies have shown that tenfold cross-validation is generally sufficient to garner reliable results.

7.6.4 Leave-one-out

When data is in short supply, as is often the case, it may be desirable to keep as much data as possible for training. Leave-one-out cross-validation is *n*-fold cross-validation taken to the extreme, such that only one instance is withheld for testing and experimentation is carried out *n* times. This approach has the added advantage of minimizing the potential for bias introduced during partitioning.

7.7 Concluding remarks

There is tremendous potential for food process modelling to benefit from data mining. Computer control and monitoring have made the collection of data all

too easy, but it is nearly impossible for people to keep up with the analysis required to turn that data into information and knowledge. Data mining provides a solution by getting computers to do the analysis as well.

It is not necessary for one to become an expert in automated reasoning systems to start data mining. All the methods we have described here are readily available on the Internet ready to run. In fact the Weka System, developed at the University of Waikato in New Zealand, has combined most of the standard learning algorithms together under a single user interface, complete with built-in tools for data preparation and results visualization. Weka has been tested under Windows, Macintosh and Linux systems and is freely available through the Internet from several sites.

This chapter has necessarily been a very brief discussion of the principles and issues of data mining. For a more thorough treatment that is easy to understand, we recommend the textbook by Witten and Frank.[17]

7.8 References

1. QUINLAN J R, *C4.5: Programs for machine learning*. San Francisco: Morgan Kaufmann, 1993.
2. MICHALSKI R S, CHILAUSKY R L, 'Learning by being told and learning from examples: An experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis'. *International Journal of Policy Analysis and Information Systems*, 1980 **4**(2).
3. HOLMES G, CUNNINGHAM S J, DE LA RUE B T, BOLLEN A F, 'Predicting apple bruising using machine learning'. *Acta Hort.*, 1998 **476** 289–96.
4. BOLLEN A F, KUSABS N J, HOLMES G, HALL M A, 'Comparison of consumer and producer perceptions of mushroom quality'. *Proc. International Multidisciplinary Conference on Integrated View of Fruit and Vegetable Quality*, Georgia, USA. Lancaster: Technomic, 2000.
5. BERRY M J A, LINOFF G, *Data Mining Techniques for Marketing, Sales and Customer Support.* New York: John Wiley, 1997.
6. KOHAVI R, JOHN G H, 'Wrappers for feature subset selection'. *Artificial Intelligence*, 1997 **97**(1/2) 273–324.
7. HALL M A, *Correlation-based feature selection for machine learning*, PhD thesis, Department of Computer Science, University of Waikato, 1998.
8. FAYYAD U M, IRANI K B, 'Multi-interval discretization of continuous-valued attributes for classification learning'. *Proc. Thirteenth International Joint Conference on Artificial Intelligence*, Chambery, France. San Francisco: Morgan Kaufmann, 1993.
9. AGRAWAL R, SRIKANT R, 'Fast algorithms for mining association rules in large databases'. *Proc. International Conference on Very Large Data Bases*, Santiago, Chile. San Francisco: Morgan Kaufmann, 1994.
10. WANG Y, WITTEN I H, 'Induction of model trees for predicting continuous

classes'. *Proc. European Conference on Machine Learning*. San Francisco: Morgan Kaufmann, 1997.

11. HOLTE R C, 'Very simple classification rules perform well on most commonly used data sets'. *Machine Learning*, 1993 **11** 63–91.
12. QUINLAN J R, 'Induction of decision trees'. *Machine Learning*, 1986 **1**(1) 81–106.
13. SHANNON C E, WEAVER W, *The Mathematical Theory of Communication*, University of Illinois Press, 1949.
14. CENDROWSKA J, 'PRISM: An algorithm for inducing modular rules'. *International Journal of Man-Machine Studies*, 1987 **27**(4) 349–70.
15. GAINES B R, COMPTON P, 'Induction of ripple-down rules applied to modeling large data bases'. *Journal of Intelligent Information Systems*, 1995 **5**(3) 211–28.
16. AGRAWAL R, IMIELINSKI T, SWAMI A, 'Database mining: a performance perspective'. *IEEE Trans. Knowledge and Data Engineering*, 1993 **5**(6) 914–25.
17. WITTEN I H, FRANK E, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco: Morgan Kaufmann, 2000.

8

Modelling and prediction in an uncertain environment

J. F. Van Impe, K. Bernaerts, A. H. Geeraerd, F. Poschet and K. J. Versyck, Katholieke Universiteit Leuven

8.1 Introduction

In recent years the interest in the concepts and applications of predictive microbiology by the food industry and retailers has increased considerably.[30] Mathematical models describing the impact of processing, distribution and storage conditions on the microbial load have proved to be useful tools in the framework of *risk assessment* and *risk management*.[7, 8, 10] More precisely, applications of these models are in predicting the likely number of hazardous and/or spoilage microorganisms at a certain point in time within a food product's life cycle (in, e.g., exposure assessment, HACCP-systems (Hazard Analysis and Critical Control Points)).

At present, these models are almost exclusively deterministic. It is tacitly assumed that the kinetic parameters, the initial states and the process controlling conditions are accurately known. Hence, the predicted microbial load at a certain point in time can be calculated exactly and is therefore fully deterministic. However, predictive modelling mostly takes place in an uncertain environment; thus relating stochastic properties to predictions.[21] Predicted values must be understood as random values characterised by a probability density function. Usually, a distinction can be made between uncertainty and variability.[15] Factors of imprecision and inaccuracy which limit the ability to quantify a variable exactly are referred to as uncertainties, whereas inherent heterogeneity results in variability. Sources of uncertainty and variability in predictive microbiology include the following:[2, 15, 21, 23]

- In general, the initial microbial load or its composition is not precisely known. Moreover, initiation of growth (or inactivation) is usually preceded by an adaptation phase (lag) which is largely dependent on the previous

growth stage and the difference between the previous and the new growth conditions. Such effects may be well defined during laboratory experiments, whereas, in real life, the initial physiological state of the cells and its influence on the microbial evolution are highly uncertain and variable.

- In real operation mode, the extrinsic and intrinsic process conditions determining the microbial behaviour, e.g., temperature and product composition, usually vary within the same product, between products and between subsequent batches.
- Uncertainties may be due to the limited number of monitoring points or the inability to perform these measurements. This may concern measurements of both the independent variable(s) and the dependent variable(s). The lack of observations might be overcome by the use of other models. Schellekens *et al.*, for example, combine heat transfer models – predicting the (local) temperature – with inactivation models in order to forecast the effectiveness of a heat processing step of a lasagne.[25] Note that the model uncertainties are accumulated.
- Moreover, measurements, e.g., cell density counts, are generally corrupted by random noise, yielding model parameters with a random character as well.[24]

With respect to the practical application of predictive models, understanding and quantification of uncertainties and errors coupled with these models is of utmost importance. As stated above, sources of uncertainty and variability may enter into the modelling process at various points causing prediction uncertainties. During model building and at the level of real-life implementation, the user should be aware of the effects of uncertainties, their consequences and how to deal with them.

This chapter discusses the model building process in the context of identifying and quantifying model uncertainties. Generally, model building starts with collecting (i) data from preliminary experiments, (ii) *a priori* knowledge from the literature and experience, and (iii) model design requirements. Initial data processing may be necessary to improve the statistical properties of the data and to remove possible outliers. These issues are addressed in section 8.2. Next, an appropriate mathematical relation (i.e., the model structure) between the manipulated variable(s) (i.e., input) and the measured variable(s) (i.e., output) needs to be defined. The constants within this mathematical expression, i.e., the model parameters, are determined during parameter estimation. Note that structure characterisation and parameter estimation are usually closely entwined. Finally, the model should be tested against (i) new experimental data obtained under similar conditions (mathematical validation) and, if passed successfully, (ii) data obtained in real-life situations, e.g., within a food product (product validation).[27] Structure characterisation, parameter estimation and model validation, are repeated within an iterative process until a suitable model is obtained. Quite often additional data are to be collected (i) to distinguish between model structures, or (ii) to improve parameter estimation quality. Careful (optimal) experiment design is indispensable at this point.

Model structure characterisation is briefly discussed in section 8.3. Model parameter estimation together with the characterisation of uncertainties (including identifying parameter distributions and confidence limits) are addressed in section 8.4. In section 8.5, the uncertainty related with the predicted model output is assessed. General conclusions are formulated in section 8.6.

8.2 Data (pre)processing

Many regression techniques and statistical analysis tools assume that the measurements of a process variable consist of the true process output and some independently and identically distributed zero-mean Gaussian random error. Correct use of the classical least sum of squared errors criterion for parameter estimation, for example, is based on this assumption (see section 8.4). Consequently, the use of appropriate data transformations to stabilise the measurement error variance is strongly recommended. Similarly, outliers are to be removed as occasional erroneous observations within the experimental data may induce biased parameter estimates. General references on these subjects include Coleman *et al.*,[9] Martens and Naes,[18] and Neter *et al.*[20]

8.2.1 Variance stabilising transformations

In order to determine the stochastic properties of the measurement errors, replicate observations at various values of the dependent variable are required. Repeated sampling (or experimentation) will not yield identical measurements of the dependent variable due to experimental errors and/or (biological) process variability. The randomness of these measurement errors (defined as the difference between the *i*th replicate observation and the mean of all replicates) is characterised by some probability density function. Commonly, a normal error distribution is assumed. According to the central limit theorem, random noise will tend to be normally distributed when the sample size tends to infinity.[20] However, non-normality of the error terms together with unequal error variances frequently appear in practice. In this case, the use of variance stabilising transformations is required as explained before.

A simple method to test for the most appropriate data transformation for variance stabilisation (equally normalising the distribution) is based on the dependence of the error variance (or standard deviation) on the mean value of the dependent variable.[19,20] Some well-known data transformations are listed in Table 8.1. Note that to obtain an estimate for the original non-transformed variable, a correction for bias has to be introduced.[19] The latter is illustrated in section 8.5.

Examples of variance stabilising transformations in predictive microbiology can be found in Jarvis[16] and McMeekin *et al.*[19] The former discusses the well-known logarithmic transformation of the microbial population density N (colony forming units per volume unit) as a function of time which is based on the

Table 8.1 Variance stabilising data transformation and the appropriate back-transformation of the predicted dependent variables (partly adopted from McMeekin *et al.*)[19]

Distribution (*) continuous[c] or discrete[d] type)	Property	Variance stabilising data transformation	Back-transformation of predictions (corrected for bias)
Normal[c]	$s_y^2 = C$	–	–
Poisson[d] Binomial[d]	$s_y^2 = C \cdot \bar{y}$	$r = \sqrt{y}$	$\hat{y} = (\hat{r})^2 + s_r^2$
Gamma[c] Negative binomial[d]	$s_y^2 = C \cdot \bar{y}^2$	$r = \ln y$	$\hat{y} = \exp(\hat{r} + (1/2) \cdot s_r^2)$
Inverse Gamma[c]	$s_y^2 = C \cdot \bar{y}^3$	$r = 1/\sqrt{y}$	$\hat{y} = (\hat{r})^{-2} \cdot (1 + 3s_r^2/(\hat{r})^2)$

(*) Distribution of the output y, or equally, the error term $(y - \bar{y})$.
s_y^2: the variance on the output y; $\bar{y}$: the sample mean
C: some proportionality constant; r: the transformed output
s_r^2: the variance on the transformed output r

heteroscedastic property of the error variance. It is namely observed that the error variance increases more or less quadratically with increasing values of N (i.e., the measurement errors exhibit a negative binomial distribution). Here, the case is illustrated with some unpublished experimental results (Bernaerts K.). Replicate samples of a growing cell culture (*Escherichia coli* K12) were analysed on the cell density N (CFU/mL) by plate counting obtained via serial dilution of the sample aliquot and surface-plating of the appropriate dilution on the enumeration medium using a spiral plater. The measurement errors, i.e., $N - N_{mean}$ (with N_{mean} the mean cell density of the replicate measurements at a certain value of N) are depicted in the upper left plot of Fig. 8.1. The upper right plot in Fig. 8.1 contains the measurement error variance at each level of the dependent variable. (The measurement error variance may be approximated by the mean sum of squared residuals, i.e. $\sum_{i=1}^{n_r}(y_i - \bar{y})/(n_r - 1)$ with y_i the ith observation of n_r replicate measurements, and $\bar{y}$ the mean of the replicate measurements (see also section 8.4.1). The square root of the mean sum of squared residuals yields the error standard deviation.) Clearly, the error variance increases with increasing values of N. After logarithmic transformation, the error variance here $s_{\ln N}^2$ is stabilised as shown in Fig. 8.1 (lower plots).

If a variance stabilising data transformation is not performed, a weighted sum of squared errors should be considered during parameter estimation (see section 8.4.1).

8.2.2 Removing outliers

Observations which deviate much more from the other measured values than would be expected from the statistical properties of the stochastic value, are

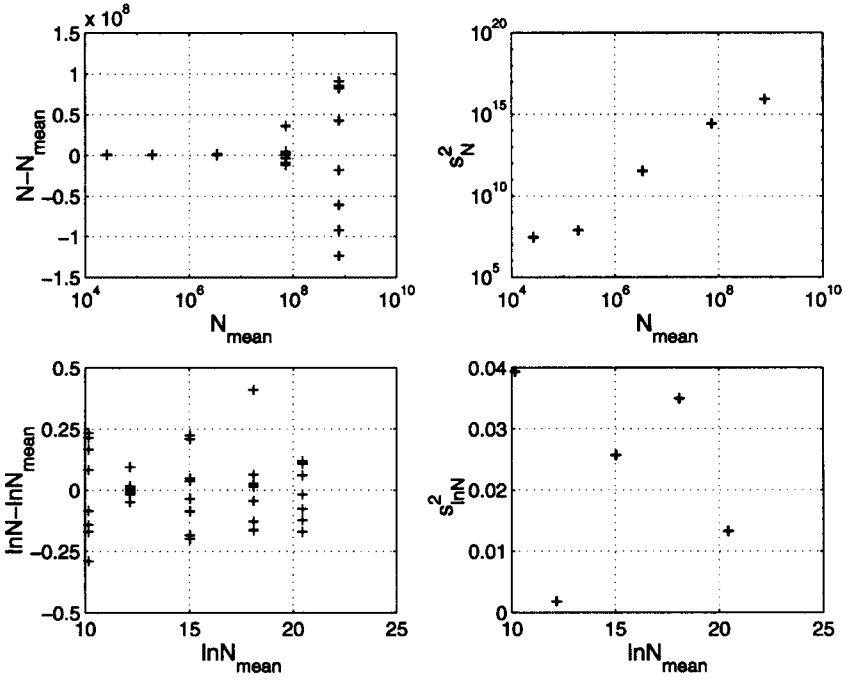


Fig. 8.1 Logarithmic transformation of cell density counts in order to stabilise the error variance (unpublished results).

called outliers and can cause great difficulty.[9, 18, 20] A major reason for discarding outliers is that under the use of the least sum of squared errors method for parameter estimation, a fitted line may be pulled disproportionately towards an outlying observation to minimise accumulative quadratic cost (see section 8.4). This could cause a misleading fit if indeed the outlier observation resulted from a mistake or other extraneous cause.

A safe rule is to discard outliers only if there is direct evidence that these represent an error in recording, a miscalculation, a malfunctioning of equipment, or a similar type of circumstances.[20] However, the more common situation is when there is no obvious or verifiable reason for the large deviation of these data points. In this case, a statistical criterion is to be used to identify points that might be considered for rejection. The selected model is then fitted to the remaining data.

In a standardised residual plot, the residual errors (i.e., the difference between the model prediction and the measured output) divided by the error standard deviation are plotted with respect to the observations.[20] Outlying points lie far beyond the scatter of remaining residues (e.g., more than three times the standard deviation). This simple graphical method, however, is not easily applied for the identification of outliers in multi-variable regression. Another

test for identifying an outlier involves fitting the model again to the other $n-1$ observations.[20] The suspect observation can now be considered as a new observation, of which one can calculate the chance to be observed in n observations. If the probability is sufficiently small, the outlier can be rejected as not originating from the same population as the other $n-1$ observations. Otherwise, the outlier is retained.

More time-consuming is the construction of a deleted residual plot.[20] The deleted residual for the *i*th data point is calculated as follows. The experimental data without the *i*th observation are fitted by the model. The difference between the model prediction and the omitted observation then corresponds to the *i*th deleted residual. If the model is appropriate, the n deleted residuals should be normally distributed. Extreme deviation from the mean of the deleted residuals could indicate the presence of an outlier.

8.3 Model structure characterisation

As there is a lack of generally applicable structure characterisation techniques for non-linear systems, the structure characterisation problem is usually formulated as a model selection problem: the most suitable model has to be selected out of a pre-specified (finite) set of candidate models.

Structural errors related with the chosen model structure contribute to the overall modelling error. A structure characterisation error results in a bias between model predictions and output measurements expressing the components of dynamic processess that are not accounted for by the model.

8.3.1 Defining a set of candidate models

The set of candidate models can be chosen based on (i) *a priori* knowledge, (ii) results of historical modelling attempts, and (iii) the aim of the model with associated model features (see also Chapter 2). For example, a model for simulation could be much more complex than a model which will be used for process control or risk analysis purposes afterwards.

Depending on the level of built-in mechanistic, biological and/or physical *a priori* knowledge, model structures can be subdivided into three classes (see, e.g., Ljung[17]). *White box models* are constructed based on physical/(bio-)chemical laws and the (full) knowledge of all underlying mechanisms. Such models commonly result from a deductive modelling approach. *Black box models* are their counterpart. They define a purely empirical mathematical relation between the input and output variables based on (informative) experimental data. Black box model parameters do not reflect physical considerations *a priori*. In this book, black box models are classified as inductive models. In between, *grey box models* try to combine the best of both worlds.

In predictive microbiology, a large number of available model types belongs to the black box models category, although the *a posteriori* biological

interpretation associated with (some of) their parameters and/or state variables facilitates their use (see, e.g., Baranji and Roberts[1] and Rosso[22]).

8.3.2 Selection criteria

Theoretical analysis

Once the candidate models are selected, it is of use to study the structural properties of the proposed models. (Model properties are termed structural if they are valid for any parameter combination within the parameter space.) Structural identifiability (or theoretical parameter identifiability) deals with the possibility to (uniquely) estimate the model parameters on the basis of the planned (perfect) measurements.[14,31] The study might reveal parameter correlation inherently associated with the (highly non-linear) model structure. In such cases, no experimental effort needs to be put into estimation of those parameters, and changing of the model structure is recommended. Furthermore, proper selection between feasible models structures can only be guaranteed when the model structures are distinguishable. Two competing model structures M and $\hat{M}$ are distinguishable if, for all feasible parameters $\mathbf{p}$ of the model structure M, there are no feasible values $\hat{\mathbf{p}}$ of a competing model structure $\hat{M}$ such that $\hat{M}(\hat{\mathbf{p}}) = M(\mathbf{p})$. In other words, competing distinguishable models cannot generate identical predictions.

Mathematical frames for checking both the conditions of identifiability and distinguishability are provided in the literature (e.g., Godfrey and Distefano[14], Walter and Pronzato[31]). However, none of the methods presented in the literature so far provides generic properties for specific classes of non-linear models, which is an important handicap in view of structure characterisation of food (bio-)process models.

Case-specific model features

Generally, it is worthwhile to define a set of model design requirements that a candidate model type should fulfill. Examples of model design requirements are: simplicity, biological significance of all model parameters, minimum number of parameters, applicability, quality of fit, minimum correlation between parameters and easy to obtain initial parameter estimations. An example in predictive microbiology is the development of the cardinal values models. An important member of this class is the CTMI-model, the Cardinal Temperature Model with Inflection Point.[22] This model, describing the maximum specific growth rate μ_{max} of microorganisms as a function of temperature, was constructed taking into account the following additional properties (based on experimental results):

1. the model should allow an inflection point between the minimum and optimum temperature for growth, (i.e., the optimum temperature for growth),
2. the value of μ_{max} should be zero at the boundaries of the temperature range of growth,

3. the model should have horizontal tangent lines near the minimum and optimum temperature for growth, and
4. the model should describe a steep decrease of μ_{max} within the super-optimal growth temperature range.

The minimum number of parameters, needed to fulfill these requirements, is four: T_{min}, T_{opt}, T_{max}, i.e., the theoretical minimum, optimum and maximum temperature for growth, respectively, and μ_{opt}, the maximum specific growth rate at T_{opt}.

Needless to say that by such an approach, the set of candidate model types will be reduced easily to a limited set of model structures fulfilling the model design requirements as defined above.

Another, more recent example in predictive microbiology can be found in Geeraerd *et al.*[13] Geeraerd *et al.* formulate a set of model design requirements by which a suitable model structure for microbial inactivation during a mild heat treatment could be selected.

Some comments on the model design requirements as stated above are the following. An important property of any model is to realise a trade-off between flexibility and parsimony. It is obvious that the capability of the model to describe different possible process conditions (flexibility) increases with the number of model parameters. However, over-parameterisation of the model is to be avoided. In such a case, the model (parameters) adjust to features of the particular realisations of the noise yielding a more or less perfect fit of the experimental data at hand (lack of model robustness or overfitting[11]), whereas good modelling practice should aim at following the general – usually quite smooth – trend hidden in the data. Furthermore, non-linear models require iterative optimisation procedures for parameter estimation which may become computationally tedious or even more unfeasible when the number of model parameters is (too) high. (The term *non-linearity* here denotes *non-linearity in the parameters*. As such, the polynome $y = ax^3 + bx^2 + cx + d$ is a linear model, whereas the exponential $y = a \exp(bx)$ is a non-linear model. For a non-linear model, some sensitivities (i.e., the partial derivatives of the model output with respect to one of its parameters) depend on at least one of the other parameters.)

Model discrimination

Further discrimination among candidate models (which all satisfy all model design requirements as discussed in the previous section) may be based on a statistical hypothesis test. The *F*-test is probably the most frequently used test.[20] Although *F*-tests are in principle only valid for linear models, they may provide some indication about suitability of the competing models.[26] Zwietering *et al.*, for example, use the *F*-test to discriminate between several microbial growth models[32] and between models describing the temperature dependence of the maximum specific growth rate.[33] Other discrimination criteria can be found in Appendix A.

8.4 Model parameter estimation

Once a suitable candidate model structure has been selected, the model is adjusted (fitted) to the experimental data for estimation of the model parameters. These parameter estimates are the result of the minimisation of some identification cost $\Im_I$ which quantifies the deviation between the model predictions and the measured output. Often model parameters are assumed deterministic implying that a unique (true) parameter vector describes the system under study. Due to the random character of the measurement error, however, the parameter estimates will be statistically distributed. Models inferring an *a priori* distribution on the model parameters are termed stochastic models. The (individual) parameter uncertainty can be characterised by the parameter variance and the confidence limits. The joint uncertainty of (correlated) model parameters can be evaluated by the construction of (joint) confidence regions.

The main source of information during parameter estimation is the available set of the experimental data. It has been demonstrated that the use of optimal experiment design for parameter estimation can reasonably contribute to an improvement of the parameter estimation accuracy (see, e.g., Walter and Pronzato[31]). In the field of predictive microbiology, promising examples of optimal experiment design for parameter estimation are reported by Versyck *et al.*[29] and Bernaerts *et al.*[6] In the respective publications, optimal temperature profiles have been designed such that the thermal inactivation and growth kinetic parameters are accurately estimated from the cell density data. The methodology of optimal experiment design for parameter estimation can also be used to optimally position sampling times (or sensors) during an experiment. Opposed to dynamic experiments applying time-varying inputs, the influencing factors are kept constant during static experiments. Here, the set of treatments (i.e., combinations of controlling factors) included in the study should be selected carefully. Typical (static) experiment designs include factorial designs, fractional factorial designs, and central composite design (see, e.g., Neter *et al.*[20]). Observe that proper experiment design may also facilitate the model structure discrimination problem.

Let us consider a time-dependent system with one measured output. The experimental data set, i.e., the (column) vector of the observations $[y_{exp}(t_1)\, y_{exp}(t_2) \ldots y_{exp}(t_{n_t})]^T$ with n_t the total number of data points, is used to estimate the vector of model parameters $\mathbf{p}$ yielding the model predictions, i.e., the vector with the model outputs $[y_{mod}(t_1, \mathbf{p})\, y_{mod}(t_2, \mathbf{p}) \ldots y_{mod}(t_{n_t}, \mathbf{p})]^T$. To illustrate the techniques of parameter estimation (section 8.4.1), the assessment of parameter estimation quality (section 8.4.2), and the quantification of the model output uncertainty (see section 8.5), Example 8.1 is considered.

Example 8.1

Growth of *Escherichia coli* K12 at 35ºC has been determined experimentally. The culture was grown in a flask containing 50 mL Brain Heart Infusion (initial pH 7.3). The cell density at different time instants, $N(t)$ (colony forming units per mL) was determined by surface-plating of appropriate serial dilutions of sampled aliquots on Plate Count Agar. As mentioned in section 8.2.1, a logarithmic transformation of the cell density data ($\ln N(t)$) stabilises the measurement noise. The experimental data are plotted in Fig. 8.2.

The Baranyi and Roberts growth model has been chosen as a suitable model structure.[1] Under constant environmental conditions (i.e., not deliberately changed), the growth model can be written as follows

$$n(t) = n_0 + \mu_{max}A(t) - \ln\left(1 + \frac{e^{\mu_{max}A(t)} - 1}{e^{(n_{max}-n_0)}}\right) \qquad (8.1)$$

$$\text{with} \quad A(t) = t + \frac{1}{\mu_{max}}\ln\left(\frac{e^{-\mu_{max}t} + \dfrac{1}{e^{\lambda\mu_{max}} - 1}}{1 + \dfrac{1}{e^{\lambda\mu_{max}} - 1}}\right)$$

where n_0 (−) and n_{max} (−) denote the natural logarithm of the initial and asymptotic cell density, respectively, μ_{max} (h^{-1}) is defined as the maximum specific growth rate, and λ (h) represents the lag time.

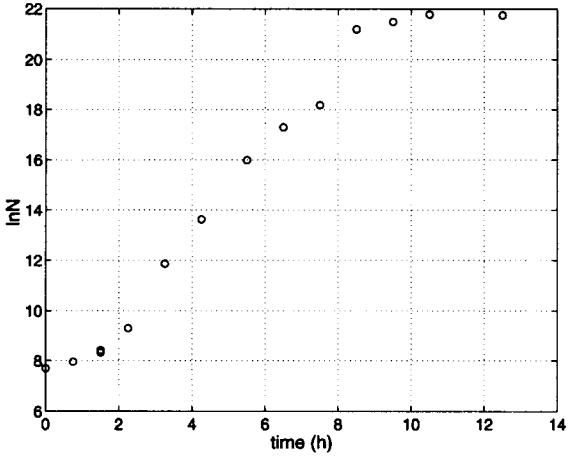


Fig. 8.2 Cell density counts measured during growth of *Escherichia coli* K12 at 35ºC.

8.4.1 Parameter estimation criteria

Probably, the most widely used identification cost is the sum of squared errors *SSE* which calculates the cumulative quadratic error between the model predictions $y_{mod}(t_i, \mathbf{p})$ and the experimental observations $y_{exp}(t_{i,j})$:

$$SSE = \sum_{i=1}^{n_{t_s}} \sum_{j=1}^{n_{r_i}} (y_{exp}(t_{i,j}) - y_{mod}(t_i, \mathbf{p}))^2 \tag{8.2}$$

with n_{t_s} the number of sampling moments, and n_{r_i} the number of replicate measurements at the sampling moment t_i. Minimisation of the *SSE* by optimal choice of the parameter vector $\mathbf{p}$, yields the so-called *least-squares estimate* $\hat{\mathbf{p}}_{ls}$. The model predictions are obtained by evaluation of the model for the *optimum* parameters $\hat{\mathbf{p}}_{ls}$, with $SSE(\hat{\mathbf{p}}_{ls})$ the corresponding least sum of squared errors, or *LSE*.

Correct use of this (unweighted) parameter estimation criterion (8.2) implies that the noise is normally distributed $N(0, \sigma^2)$ with a constant variance σ^2.[31] In this case and given the model structure is error-free, the residuals resemble the measurement noise on the experimental data, the variance of which can be approximated by the mean sum of squared errors MSE:

$$MSE = \frac{\sum_{i=1}^{n_{t_s}} \sum_{j=1}^{n_{r_i}} (y_{exp}(t_{i,j}) - y_{mod}(t_i, \hat{\mathbf{p}}_{ls}))^2}{n_t - n_p} \tag{8.3}$$

where the nominator corresponds with the *LSE* and the denominator is equal to the degrees of freedom being the total number of data points (n_t), i.e., $\sum_{i=1}^{n_{t_s}} n_{r_i}$, minus the number of estimated model parameters (n_p).

When the measurement error distribution is known, the problem can be solved in a maximum likelihood sense. A maximum likelihood estimator maximises the likelihood of the experimental data being generated by the model with the parameters $\mathbf{p}$. Under the above-stated error conditions, the maximum likelihood criterion reduces to an ordinary least sum of squared errors problem. For the interested reader, the basic principles of maximum likelihood estimators are given in Appendix B. In order to complete the overview of parameter estimation criteria, Bayesian estimators are also mentioned in this appendix.

Observe that whenever error terms are not constant, a variance stabilising data transformation as discussed in section 8.2.1, or the application of a weighted *SSE* is recommended.

$$SSE_w = \sum_{i=1}^{n_{t_s}} \sum_{j=1}^{n_{r_i}} w_{i,j} \cdot (y_{exp}(t_{i,j}) - y_{mod}(t_i, \mathbf{p}))^2 \tag{8.4}$$

The choice of $w_{i,j}$ will express the relative confidence in the various experimental data. A relative identification is obtained by choosing $w_{i,j}$ equal to the squared inverse of the corresponding output measurement, i.e., $1/[y_{exp}(t_{i,j})]^2$. Use of the relative identification cost will improve the fit at small output values, which is of interest when outputs with very different amplitude are to be fitted

simultaneously, or when the error variance increases proportionally with increasing output values (see Table 8.1).

Division of the *SSE* by the total number of experimental data allows comparison of costs obtained with different number of data points.

Example 8.2

For the current example, the model parameters have been estimated by non-linear regression in SAS (SAS Institute Inc., North Carolina, Release 6.12) based on the minimisation of the sum of squared errors. Results are stated in Table 8.2, first two columns.

Table 8.2 Evaluation of the approximate uncertainty on the growth model parameter estimates for the data of *E. coli* K12 depicted in Fig. 8.2

Parameter		Estimate	Asymp. St. Dev.	95% confidence interval
$n(0)$	$(-)$	7.541	0.412	[6.634, 8.448]
n_{max}	$(-)$	21.848	0.340	[21.099, 22.598]
μ_{max}	(h^{-1})	1.771	0.082	[1.591, 1.592]
λ	(h)	1.061	0.373	[0.239, 1.883]

Asymptotic Correlation Matrix

	$n(0)$	n_{max}	m_{max}	λ
$n(0)$	1.000			
n_{max}	−0.0474	1.000		
μ_{max}	0.323	−0.230	1.000	
λ	0.842	−0.110	0.715	1.000

8.4.2 Assessing uncertainty on parameter estimates

The quality of an estimator can be assessed by its parameter covariance matrix:

$$\mathbf{P} = E\{(\hat{\mathbf{p}} - \mathbf{p}^*)(\hat{\mathbf{p}} - \mathbf{p}^*)^{\mathrm{T}}\} \tag{8.5}$$

where E denotes the expectation value, $\mathbf{p}^*$ represents the true parameter vector, $(\hat{\mathbf{p}} - \mathbf{p}^*)$ is the vector of estimation errors, and T denotes the transpose. However, other methods can be used in assessing the parameter uncertainty as will be pointed out in the following paragraphs. The emphasis will be put on the case that measurement noise can be presented as an additive random variable which is independently and identically distributed according to a zero-mean Gaussian law with variance σ^2. The model structure is assumed perfectly known.

Analysis of parameter covariance matrix

Given data perturbated by an independently and identically distributed random noise and structural identifiability of the model, it can be shown that maximum

likelihood estimators $\hat{\mathbf{p}}$ are asymptotically efficient, asymptotically unbiased, and asymptotically Gaussian distributed (see, e.g., Walter and Pronzato[31]).

The maximum likelihood estimator is said to be an unbiased estimator of $\mathbf{p}^*$ when

$$E\{\hat{\mathbf{p}}\} = \mathbf{p}^*$$

which means that if it were possible to replicate the same experiment and estimate $\hat{\mathbf{p}}$ an infinite number of times, the mean of these estimates would coincide with the true value $\mathbf{p}^*$. Maximum likelihood estimators are unbiased when the number of data n_t tends to infinity, but are generally biased for a limited number of data points.

According to the Cramér-Rao inequality, the inverse of the Fisher information matrix $\mathbf{F}$ provides the lower bound on the parameter covariance matrix $\mathbf{P}$ for any absolutely unbiased estimator (see, e.g., Ljung[17] for the elaborated proof):

$$\mathbf{F}^{-1}(\mathbf{p}^*) \leq \mathbf{P} \tag{8.7}$$

with

$$\mathbf{F}(\mathbf{p}) = E_{y^s|\mathbf{p}}\left\{\left[\frac{\partial}{\partial \mathbf{p}} \ln \pi_y(\mathbf{y}^s|\mathbf{p})\right]\left[\frac{\partial}{\partial \mathbf{p}} \ln \pi_y(\mathbf{y}^s|\mathbf{p})\right]^{\mathrm{T}}\right\} \tag{8.8}$$

where $\pi_y(\mathbf{y}^s|\mathbf{p})$ represents the conditional probability density of the observed data set $\mathbf{y}^s$ given the parameter vector $\mathbf{p}$. An estimator that reaches the Cramér-Rao lower bound is called efficient. If n_t tends to infinity and the model is linear in its parameters, the maximum likelihood estimator is efficient, thus yielding the smallest parameter covariance that can be achieved.

Further, it can be demonstrated that the distribution of $\hat{\mathbf{p}}$ tends to a Gaussian distribution $N(\mathbf{p}^*, \mathbf{F}^{-1}(\mathbf{p}^*))$ with the true parameters $\mathbf{p}^*$ as the mean value, and the inverse of the Fisher information matrix $\mathbf{F}$ (8.8) as the covariance matrix.

Given a zero-mean Gaussian error distribution with variance σ^2, the Fisher information matrix can be written as follows.[31]

$$\mathbf{F}(\mathbf{p}^*) = \frac{1}{\sigma^2}\sum_{i=1}^{n_t}\left[\frac{\partial y_{mod}(t_i, \mathbf{p})}{\partial \mathbf{p}}\right]_{|\mathbf{p}=\mathbf{p}^*}\left[\frac{\partial y_{mod}(t_i, \mathbf{p})}{\partial \mathbf{p}}\right]^{\mathrm{T}}_{|\mathbf{p}=\mathbf{p}^*} \tag{8.9}$$

When using the matrix $\mathbf{F}$ in equation (8.9) to quantify the quality of the parameter estimates of a particular experiment, besides the assumption on the measurement error and the commonly small set of data points, further approximations are made by replacing the true parameter vector by the estimated parameters, and the error variance σ^2 by the mean sum of squared errors *MSE*. Especially with highly non-linear models, the results should be interpreted with caution.[31]

In some textbooks and statistical software packages, the approximate parameter covariance matrix is calculated as

$$\{\mathbf{J}^{\mathrm{T}} \cdot \mathbf{J}\}^{-1} \cdot MSE \tag{8.10}$$

with **J** the (n_t by n_p) Jacobian matrix containing the partial derivatives of the model output with respect to the model parameters evaluated at each measurement point, and *MSE* the mean sum of squared errors (equation 8.3).

Calculation of the inverse of the Fisher information matrix for the model parameter estimates, $\mathbf{F}^{-1}(\hat{\mathbf{p}})$, yields approximate values for the variances on the parameter estimates $s^2(\hat{p}_i)$ (main diagonal), and also the covariance between two parameter estimates $s(\hat{p}_i, \hat{p}_j)$ (off-diagonal elements), which can be used to construct confidence intervals and to calculate correlation coefficients.

- An approximate $(1-\alpha)\cdot 100\%$ confidence interval on the estimated value of the *i*th parameter can be formulated:

$$[\hat{p}_i - t_{(1-\frac{\alpha}{2}, n_t - n_p)} \cdot \sqrt{s^2(\hat{p}_i)}, \hat{p}_i + t_{(1-\frac{\alpha}{2}, n_t - n_p)} \cdot \sqrt{s^2(\hat{p}_i)}] \tag{8.11}$$

  where t represents the Student t-value for $n_t - n_p$ degrees of freedom at a confidence level $(1-\alpha)$ (which can be read from statistical tables).
- An approximate correlation coefficient between the *i*th and *j*th parameter is defined by:

$$-1 \leq \frac{s(\hat{p}_i, \hat{p}_j)}{s(\hat{p}_i) \cdot s(\hat{p}_j)} \leq 1 \tag{8.12}$$

Example 8.3
The approximate covariance matrix of the parameter estimates is calculated according to equation (8.9). Note that this matrix is a side-product of the parameter estimation (optimisation) algorithm. The results are stated in Table 8.2.

Joint confidence regions
Consider the case of a maximum likelihood parameter estimation problem. Given independent observations and an output linearised with respect to the parameters, the errors on the simultaneously estimated parameters can be assessed by the construction of joint confidence regions which are defined by the following inequality:[5]

$$SSE(\mathbf{p}) \leq SSE(\hat{\mathbf{p}}_{ls}) \cdot \left(1 + \frac{n_p}{n_t - n_p} F(n_p, n_t - n_p, 1-\alpha)\right) \tag{8.13}$$

with $SSE(\mathbf{p})$ the sum of squared errors for a parameter vector $\mathbf{p}$, $SSE(\hat{\mathbf{p}}_{ls})$ the least-squares error (corresponding to the best-fit estimates $\hat{\mathbf{p}}_{ls}$), n_p the number of parameters estimated simultaneously, n_t the number of observations, and F the value of the Fisher-Snedecor F-distribution with n_p and $n_t - n_p$ degrees of freedom and a confidence level of $1-\alpha$. The F-values are tabulated in statistical textbooks, e.g., Neter *et al.*,[20] or can be provided by statistical software.

In the case of $n_p > 2$, it is impossible to visualise correctly the n_p-dimensional hypervolume constituted by the joint confidence region (8.13). On

the one hand, a projection on the state plane of two parameters results in an overestimation of the confidence limits (marginal region). On the other hand, a two-dimensional representation of the confidence region for two parameters can be obtained by calculating the $SSE(\mathbf{p})$ only varying the respective two parameters and keeping the other $n_p - 2$ model parameters at their optimum value (conditional region). This may result in an underestimation of the confidence regions. Examples in predictive microbiology can be found in the literature.[12, 22]

Example 8.4

The 95% joint confidence regions of the parameter estimates for the Baranyi and Roberts growth model, using the experimental data of *E. coli* presented in Fig. 8.2, are depicted in Fig. 8.3.

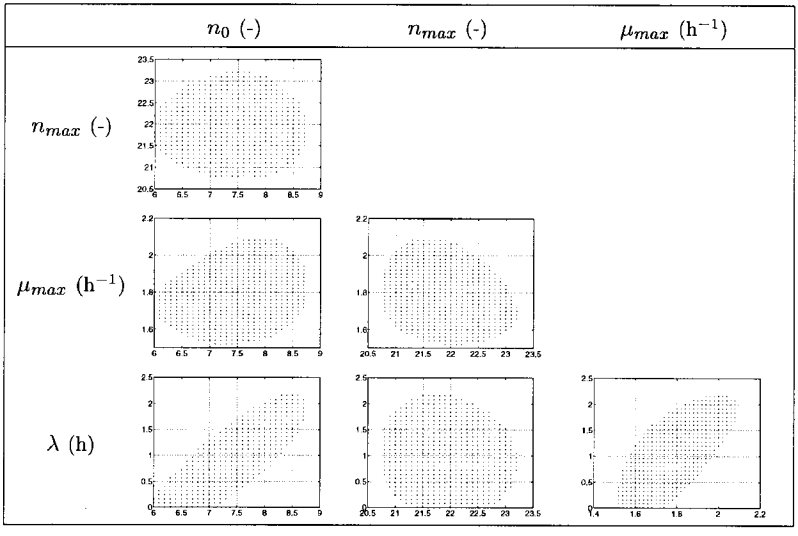


Fig. 8.3 Projection of the four-dimensional hypervolume on the two-dimensional parameter space.

Monte Carlo method

The procedure is as follows.[31]

- First, the model parameters of a model *M* are estimated based on the experimental data.
- Second, fictitious data are generated. Hereto, the model *M* is evaluated at each sampling point using the estimated model parameters (i.e., simulation) and artificially generated experimental noise is added to the model output. This procedure is repeated yielding a series of fictitious experimental data.

Note that the experimental uncertainty needs to be characterised completely *a priori*, though a normal error distribution is no prerequisite in this case. The jack-knife and bootstrap methods make it possible to avoid estimating the distribution of the noise from residuals.[31]

- Third, fictitious parameters are estimated from each fictitious data set.
- Finally, the statistical properties (e.g., probability density function, mean, covariance matrix) can be determined for the set of (fictitious) parameter estimates.

A major drawback of the Monte Carlo method is that it is computationally intensive. A large number of repetitive simulations is necessary to determine the parameter estimates distribution accurately. Moreover, a high level optimisation algorithm is required in order to converge to the best parameter set starting from other sets. This is not evident for problems with a strong correlation between the parameters.

Example 8.5

A Monte Carlo analysis is performed for the example of *E. coli* K12. A constant Gaussian noise on the natural logarithm of the cell density counts with variance 0.25 is assumed. After 10 000 runs, the frequency distribution for each model parameter as depicted in Fig. 8.4 is obtained. For each model parameter, a normal probability function is plotted on the histograms. The corresponding mean values and variances are listed in Table 8.3.

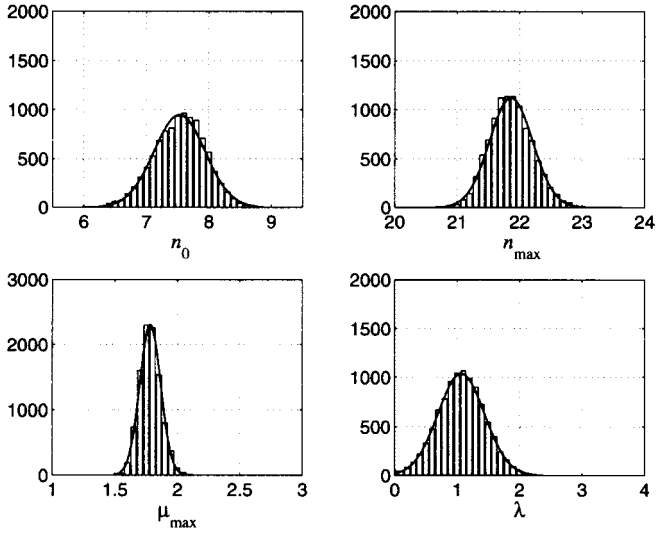


Fig. 8.4 Frequency plots for the growth model parameters determined by application of the Monte Carlo method to the experimental data given in Fig. 8.2.

Table 8.3 Mean growth model parameters and their standard deviation determined by the Monte Carlo method applied to the data in Fig. 8.2

	$n_0(-)$	$n_{max}(-)$	$\mu_{max}(h^{-1})$	λ(h)
Mean	7.529	21.863	1.782	1.066
Standard deviation	0.412	0.347	0.084	0.376

8.5 Model output uncertainty assessment

The final objective of mathematical modelling is the prediction of future responses at given conditions (or time instances). Hereto, the model is evaluated using the estimated model parameters. Owing to the random nature of the model parameters, the model outputs cannot be predicted in a deterministic sense. Instead, the predicted model output is also a random variable, characterised by means of some probability density function.

At this point, the following should be noted. To fulfil the requirement of normally distributed error terms, a data transformation may be necessary as explained in section 8.2.1. Consequently, any statistical analysis is to be established for the transformed data. The use of an appropriate back transformation (see Table 8.1) yields untransformed predictions and the accompanying confidence limits.

The Monte Carlo method is the most general way to determine the stochastic properties of the model predictions. A sample of the random parameter set $\mathbf{p}$ is generated on the computer and the model is solved numerically. After a reasonable amount of repetitions, the mean $\overline{y_{mod}(t_i, \mathbf{p})}$ and the variance on $y_{mod}(t_i, \mathbf{p})$ can be estimated using common statistical techniques.[24]

Major drawbacks of the Monte Carlo method are (a) the large number of repetitive simulations necessary to obtain an acceptable level of accuracy and (b) the fact that the stochastic parameter set must be specified completely in a probabilistic sense.[24]

An alternative approach to determine the uncertainty on the predicted model output assumes a Gaussian distributed prediction error and makes a linear approximation of the model output to derive the prediction error variance (see, e.g., Bard[3] and Walter and Pronzato[31]). In this case, the mean value of the predicted model output $\overline{y_{mod}(t_i, \mathbf{p})}$ is set equal to $y_{mod}(t_i, \overline{\mathbf{p}})$ i.e., the model evaluated at time t_i with $\overline{\mathbf{p}}$ the mean of the model parameter distribution. The latter may be replaced by the estimated model parameters $\hat{\mathbf{p}}$ if an unbiased estimator has been used. Note however that the parameter estimates of non-linear models are usually slightly biased due to the non-linearity of the model. The prediction error variance at an arbitrary time t_i can be approximated by

$$\sigma^2_{y(t_i,\mathbf{p})} \equiv E\left\{\left(y_{mod}(t_i,\mathbf{p}) - \overline{y_{mod}(t_i,\mathbf{p})}\right)\left(y_{mod}(t_i,\mathbf{p}) - \overline{y_{mod}(t_i,\mathbf{p})}\right)^{\mathrm{T}}\right\}$$
$$= \left[\frac{\partial}{\partial \mathbf{p}} y_{mod}(t_i,\mathbf{p})_{|\mathbf{p}=\bar{\mathbf{p}}}\right]\mathbf{P}\left[\frac{\partial}{\partial \mathbf{p}} y_{mod}(t_i,\mathbf{p})_{|\mathbf{p}=\bar{\mathbf{p}}}\right]^{\mathrm{T}} \quad (8.14)$$

with **P** the parameter covariance matrix which can be obtained as explained in section 8.4.2.[3, 4, 24]

Given the stochastic properties, the confidence limits on the predicted model output can be defined. These limits define a confidence interval for the expected value of the model output. In the case of a Gaussian distribution of the predicted model output, a confidence interval on the (mean) predicted model output with a $(1-\alpha)$ certainty level is given by

$$\overline{y_{mod}(t_i,\mathbf{p})} \pm t_{(1-\frac{\alpha}{2},n_t-n_p)} \cdot s_{y(t_i,\mathbf{p})} \quad (8.15)$$

where t represents the *Student* t-value and $s_{y(t_i,\mathbf{p})}$ denotes the approximate standard deviation of the prediction error which can be derived from a Monte Carlo analysis or approximated by the square root of equation (8.14).

If the outcome of a new experiment is to be predicted, measurement noise should be taken into account.[3, 20] A so-called prediction interval for the actual model output at time t_i is written as

$$\overline{y_{mod}(t_i,\mathbf{p})} \pm t_{(1-\frac{\alpha}{2},n_t-n_p)} \cdot \sqrt{s^2_{y(t_i,\mathbf{p})} + s^2_{y(t)}} \quad (8.16)$$

where the variance on the measured output $s^2_{y(t)}$ is usually given by the mean sum of squared errors (equation 8.3).

Note that here the independent variable t_i is assumed perfectly known. In general, however, these measurements are also corrupted by a random error. Inaccuracies associated with the independent variable will yield an additional uncertainty on the model parameters, as well as on the model output predictions (see, e.g., Bard[3]).

Example 8.6

Figure 8.5 depicts the 95% approximate confidence and prediction intervals on the predicted growth $n(t,\hat{\mathbf{p}})(\equiv \hat{n}(t))$ of *E. coli* K12 calculated according to equation (8.15) (making use of equation (8.14)) and equation (8.16), respectively.

To conclude, estimates for the non-transformed output can be calculated using the back-transformation formula presented in Table 8.1 (section 8.2.1). Remember that an appropriate back-transformation is necessary in order to account for the bias induced by applying a non-linear data transformation. Here, the natural logarithm of the cell density $n(t) \equiv \ln N$ has been modelled. An estimate for the cell density N at a time-instant t_i is then given by

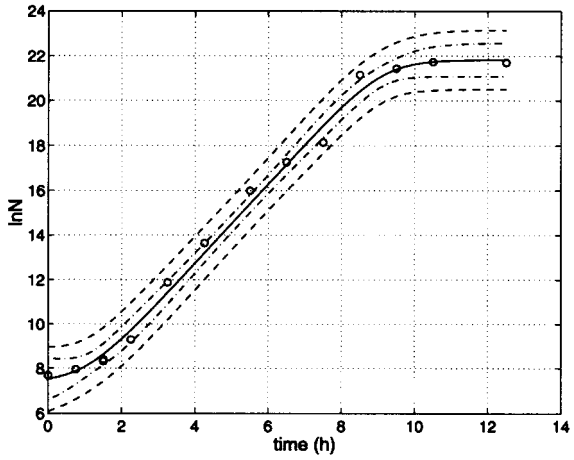


Fig. 8.5 Approximate 95% confidence ($- \cdot -$) and prediction ($--$) interval on the predicted growth $n(t, \hat{\mathbf{p}})$ of *E. coli* K12 with $\hat{\mathbf{p}}$ the model parameter estimates stated in Table 8.2. Circles (o) represent the experimental data points.

$$\hat{N}(t_i) = \exp(\hat{n}(t_i) + (1/2) \cdot s^2_{n(t)})$$

where the variance on the transformed data ($s^2_{n(t)}$) may be approximated by the mean sum of squared errors (see section 8.4.2).

8.6 Conclusions

In view of revealing and dealing with modelling uncertainties, the basic steps of model building have been discussed in this chapter. In summary, the following conclusions are to be remembered.

- Due to the incomplete (mechanistic) knowledge of the process and/or its complexity, it is difficult – if not impossible – to find a perfect mathematical relation between the process input and output (section 8.3). During model structure characterisation (or selection), the user should keep in mind some general model design requirements. The definition of case-specific features may further facilitate the model structure selection process.
- When the most appropriate model structure has been selected, the objective of parameter estimation is to find unbiased parameters, i.e., estimates that coincide with the true process parameters. Correct use of the (classical) parameter estimation criterion is indispensable here (section 8.4). Hereto, data (pre)processing involving variance stabilisation and the removal of outlying erroneous observations might be necessary (see section 8.2).
- Measurement noise superimposed on the dependent variable(s) yields parameter estimates with a random character described by some probability density function. Methods to assess the parameter estimation uncertainty are

explained in section 8.4.2. Note that the results of each method should always be interpreted with caution (as many presumptions and approximations are made).

- Because of the random nature of the parameter estimates, model output predictions are stochastic variables as well. In section 8.5, confidence limits on the expected values of the output are defined. Whenever predictions are made, appropriate confidence limits should be established. Here, it is important to note that measurement uncertainties on the independent variable(s) will augment the final prediction errors as well.[3]

Within the complete modelling cycle, the proposed model – after selection of a suitable model structure and parameter estimation – should be validated. Here, two validation steps can be distinguished.[27] During mathematical validation, the model is tested on new experimental data generated under conditions similar to those applied during the experiment yielding the identification data. If the modelling capacity is not acceptable, the model structure or the model parameters should be re-identified. An iterative process of structure characterisation, parameter estimation and model validation may be necessary to obtain the desired descriptive quality. In the field of predictive microbiology, most modelling exercises are performed based on laboratory scale experiments under conditions sometimes quite different from real food product conditions. Consequently, it is not always trivial to apply the developed model to real food products. Therefore, a product validation step is to be performed. Model predictions are compared with the outcome of experimental trials in food products under real-life conditions. As a result, product validation can be seen as a way to investigate the appropriateness of the laboratory medium as model system for the real food product conditions. An adaptation of the laboratory medium could be a first step towards product validation.

8.7 Appendix A

The Final Prediction Error, Akaike's Information Criterion and the Bayesian Information criterion are examples of model selection criteria different from the *F*-test. These criteria measure the quality of the model by means of the number of model parameter n_p and the sum of squared errors *SSE*.[17] Two basic forms can be distinguished:[28]

$$\frac{SSE}{n_t}[1 + \beta(n_t, n_p)] \tag{8.17}$$

$$n_t \log\left(\frac{SSE}{n_t}\right) + \gamma(n_t, n_p) \tag{8.18}$$

In both cases, the first term decreases with increasing n_t while the second term penalises too complex models. The Final Prediction Error is given by equation (8.17) if $\beta(n_t, n_p) = 2 \cdot n_p/(n_t - n_p)$. Replacing $\gamma(n_t, n_p)$ by $2 \cdot n_p$ in

equation (8.18) results in the Akaike's Information Criterion. Unfortunately, these criteria are not consistent, i.e., they do not guarantee that the probability of selecting the wrong model tends to zero as the number of data points tends to infinity. A consistent criterion is the Bayesian Information Criterion in which $\gamma(n_t, n_p)$ equals $n_p \cdot \log(\log(n_t))$. For more information, reference is made to Ljung,[17] and references within that text.

8.8 Appendix B

8.8.1 Maximum likelihood estimators

A *maximum likelihood estimator* maximises the likelihood, i.e., the probability density function π, of the random variable $\mathbf{y}^s$ being generated by a model with the parameters $\mathbf{p}$.[31] The random variable $\mathbf{y}^s$ is a vector containing a series of experimental observations $y_{exp}(t_{i,j})$.

$$\hat{\mathbf{p}}_{ml} = \arg\max \pi_y(\mathbf{y}^s|\mathbf{p}) \tag{8.19}$$

In practice, it is often easier to look for $\hat{\mathbf{p}}_{ml}$ by maximising the log-likelihood function

$$\Im_{ml}(\mathbf{p}) = \ln \pi_y(\mathbf{y}^s|\mathbf{p}) \tag{8.20}$$

which yields the same estimate since the logarithmic function is monotonically increasing.

Given the model output is disturbed by an additive, independent Gaussian zero-mean random noise with a known or constant variance σ_i^2, the maximum likelihood estimate $\hat{\mathbf{p}}_{ml}(\mathbf{y}^s)$ of $\mathbf{p}$ is the minimiser of the quadratic cost

$$\Im_{ml}(\mathbf{p}) \equiv \Im_{wls}(\mathbf{p}) = \sum_{i=1}^{n_{ts}} \sum_{j=1}^{n_{ri}} w_{i,j} \cdot (y_{exp}(t_{i,j}) - y(t_i, \mathbf{p}))^2 \tag{8.21}$$

with the weights $w_{i,j} = 1/\sigma_i^2$. This estimator can be used if the noise does not follow a normal law, provided that the variance is known at any time t_i. In this case, it is called the *Gauss-Markov estimator*. For models linear within the parameters, minimisation of equation (8.21) yields the *best linear unbiased estimator* for $\mathbf{p}$.[31] Choosing the weights equal to one, the maximum likelihood estimator reduces to a common least squares estimator (see section 8.4.1).[31]

Given certain assumptions, e.g., no structure characterisation error, data corrupted by an independently identically distributed noise, the maximum likelihood estimators have the following interesting properties (full details are given in Walter and Pronzato).[31]

- They are *consistent*, i.e., they converge to the true parameters when the number of data points tends to infinity.
- Moreover, the estimator is *asympotically efficient* meaning that no consistent estimator with a smaller variance exists if $n_t \to \infty$.
- A third property is given in section 8.4.2 (equation 8.7).

8.8.2 Bayesian criteria

If the model parameters are characterised by a distribution with an *a priori* known probability function $\pi_p(\mathbf{p})$, this information should be included in the parameter estimation criterion.[31] The joint probability is given by

$$\pi_y(\mathbf{y}^s, \mathbf{p}) = \pi_y(\mathbf{y}^s|\mathbf{p}) \cdot \pi_p(\mathbf{p}) = \pi_p(\mathbf{p}|\mathbf{y}^s) \cdot \pi_y(\mathbf{y}^s)$$

Hence, the *a posteriori* probability of $\mathbf{p}$, given the data $\mathbf{y}^s$, can be calculated by

$$\pi_p(\mathbf{p}|\mathbf{y}^s) = \frac{\pi_y(\mathbf{y}^s|\mathbf{p}) \cdot \pi_p(\mathbf{p})}{\pi_y(\mathbf{y}^s)} \tag{8.22}$$

which is termed *Bayes'* rule. *Bayesian estimators* will maximise the posterior probability function for $\mathbf{p}$. Like the maximum likelihood estimators, Bayesian estimators are consistent and asymptotically efficient given that $\pi_p(\mathbf{p})$ is continuous and non-zero at $\hat{\mathbf{p}}_{ml}(\mathbf{y}^s)$.[31]

8.9 References

1. BARANYI J, ROBERTS TA, 'A dynamic approach to predicting bacterial growth in food', *Int. J. Food Microbiol.*, 1994 23: 277–94.
2. BARANYI J, 'Comparison of stochastic and deterministic concepts of bacterial lag', *J. Theor. Biol.*, 1998 192: 403–8.
3. BARD Y, *Nonlinear Parameter Estimation*. Academic Press Inc, New York, USA, 1974.
4. BATES DM, WATTS DG, *Nonlinear Regression Analysis and its Applications*, John Wiley and Sons, New York, 1988.
5. BEALE EML, 'Confidence regions in non-linear estimation', *J. Roy. Stat. Soc.*, 1960 B22: 41–88.
6. BERNAERTS K, VERSYCK KJ, VAN IMPE JF, 'On the design of optimal dynamic experiments for parameter estimation of a Ratkowsky-type growth kinetics at suboptimal temperatures', *Int. J. Food Microbiol.*, 2000 54(1–2): 27–38.
7. BUCHANAN RL, WHITING RC, 'Risk assessment and predictive microbiology', *J. Food Protect.*, 1996 Supplement: 26–31.
8. CASSIN M, LAMMERDING AM, TODD ECD, ROSS W, MCCOLL RS, 'Quantitative risk assessment for *Escherichia coli* O157:H7 in ground beef hamburgers', *Int. J. Food Microbiol.*, 1998 41: 21–44.
9. COLEMAN HW, STEELE WG, *Experimentation and Uncertainty Analysis for Engineers* (second edition), John Wiley and Sons, New York, USA, 1999.
10. COLEMAN ME, MARKS HM, 'Qualitative and quantitative risk assessment', *Food Control*, 1999 10: 289–97.
11. DELIGNETTE-MULLER ML, ROSSO L, FLANDROIS JP, 'Accuracy of microbial growth predictions with square root and polynomial models', *Int. J. Food Microbiol.*, 1995 27: 139–46.
12. GEERAERD AH, HERREMANS CH, LUDIKHUYZE LR, HENDRICKX ME, VAN

IMPE J F, 'Modelling the kinetics of isobaric-isothermal inactivation of *Bacillus subtilis* α-amylase with artificial neural networks', *J. Food Engin.*, 1998 36(3): 263–79.

13. GEERAERD A H, HERREMANS C H, VAN IMPE J F, 'Structural model requirements to describe microbial inactivation during a mild heat treatment', *Int. J. Food Microbiol.*, 2000 59(3): 185–209.
14. GODFREY K R, DISTEFANO J J, 'Identifiability of Model Parameters', *3rd IFAC Identification and System Parameter Estimation*, York, UK, 1985.
15. HAAS C N, ROSE J B, GERBA C P, CROCKETT C S, 'What predictive food microbiology can learn from water microbiology', *Food Control*, 1997 51(4): 88–90.
16. JARVIS B, *Statistical Aspects of the Microbiological Analysis of Foods*, Progress in Industrial Microbiology 21, Elsevier, Amsterdam, 1989.
17. LJUNG L, *System Identification Theory for the User* (Second Edition), Prentice Hall Inc., Upper Saddle River, NJ, USA, 1999.
18. MARTENS H, NAES T, *Multivariate Calibration*, John Wiley & Sons, New York, USA, 1989.
19. MCMEEKIN T A, OLLEY J N, ROSS T, RATKOWSKY D A, *Predictive Microbiology: Theory and Application*, Research Studies Press, Exeter, 1993.
20. NETER J, WASSERMAN W, KUTNER M H, *Applied Linear Statistical Models*, Irwin, Boston, 1989.
21. NICOLAÏ B M, VAN IMPE J F, 'Predictive food microbiology: a probabilistic approach', *Math. and Comput. Simulat.*, 1996 42: 287–92.
22. ROSSO L, *Modélisation et microbiologie prévisionnelle: Elaboration d'un nouvel outil pour l'agro-alimentaire*, PhD thesis, Université Claude Bernard-Lyon I, Villeurbanne Cedex (France), 1995.
23. SCHAFFNER D W, LABUZA T P, 'Predictive microbiology: Where are we and where are we going?', *Food Control*, 1997 51(4): 95–9.
24. SCHEERLINCK N, POSCHET F, VAN IMPE J F, NICOLAÏ B M, 'Stochastic perturbation analysis of a microbial growth model', *Proceedings 3rd IMACS Symposium on Mathematical Modelling (MathMod)*, Technical University Vienna, Vienna, 2000, 619–22.
25. SCHELLEKENS M, MARTENS T, ROBERTS T A, MACKEY B M, NICOLAÏ B M, VAN IMPE J F, DE BAERDEMAEKER J, 'Computer-aided microbial safety design of food processes', *Int. J. Food Microbiol.*, 1994 24: 1–9.
26. SCHNUTE J, 'A versatile growth model with statistically stable parameters', *Can. J. Fish. Aquat. Sci.*, 1981 38: 1128–40.
27. VAN IMPE J F M, VERSYCK K J, GEERAERD A H, 'Validation of predictive models: definitions and concepts', *COST-914: Predictive Modelling of Microbial Growth and Survival in Foods*, Office for Official Publications of the European Communities, Luxembourg, 1999, 31–8.
28. VANROLLEGHEM P A, DOCHAIN D, 'Bioprocess model identification', In: *Advanced Instrumentation, Data Interpretation, and Control of Biotechnological Processes*. Van Impe J F M, Vanrolleghem P A and Iserentant D M (eds.), Kluwer Academic Publishers, Dordrecht, 1998, 251–313.

29. VERSYCK K J, BERNAERTS K, GEERAERD A H, VAN IMPE J F, 'Introducing optimal experimental design in predictive microbiology: a motivating example', *Int. J. Food Microbiol.* 1999 51(1): 39–51.
30. WALLS I, SCOTT V N, 'Use of predictive microbiology in microbial food safety risk assessment', *Int. J. Food Microbiol.*, 1997 36: 97–102.
31. WALTER E, PRONZATO L, *Identification of Parametric Models from Experimental Data*, Springer, Masson, 1997.
32. ZWIETERING M H, JONGENBURGER I, ROMBOUTS F M, VAN 'T RIET K, 'Modeling of the bacterial growth curve', *Appl. Environ. Microbiol.*, 1990 56: 1875–81.
33. ZWIETERING M H, DE KOOS J T, HASENACK B E, DE WIT J C, VAN 'T RIET K, 'Modeling bacterial growth as a function of temperature', *Appl. Environ. Microbiol.*, 1991 57: 1094–101.

Part III

Applications: agricultural production

Introduction

Nothing in life is purely black or white, as, at close inspection, a whole gradation of grey colours is revealed. In the same way, models are often a mixture of fundamental and empirical approaches. Depending on the preference of the modeller one of the sides may be emphasised, but the other side does not need to be concealed. At this stage, it should be clear that both approaches have their own benefits and drawbacks, and can be combined to strengthen each other. After having discussed the basic principles of modelling (Parts I and II) it is now time to see how these approaches can be applied. This part will deal with four examples from agricultural production. The basic principle applied throughout the development of all four models discussed was to complete the mass and energy balances of the studied systems.

Chapters 9 and 10 deal with the important principles of simulation models for greenhouse and field cultivated vegetables, illustrated using cucumber and tomato. Both authors present hybrid models, closely linked to photosynthetic activity of the crop, using varying levels of fundamental and empirical relationships. In both cases, this resulted in valuable models for use in decision support systems, for greenhouse climate control and for production planning and crop management.

Chapters 11 and 12 deal with models on dairy and beef cattle production. These models start from the same type of crop growth models outlined in Chapters 9 and 10, as grassland production forms the base for the lifecycle of cows. Grassland production is linked to cow and dairy production via intake models. The efficiency of animal and dairy production strongly depends on the digestibility of the food. To be able to assess the profitability of the production

system, economic costs and values are included. The developed models can be used to test alternative scenarios and to assess the impact of management changes on the efficiency of the production systems.

Maarten Hertog

9

Yield and quality prediction of vegetables: the case of cucumber

L. F. M. Marcelis, Plant Research International, Wageningen

9.1 Introduction

Fruit vegetables such as cucumber as well as many cut flowers and pot plants are often grown in greenhouses. Greenhouses enable growers to control climate conditions and nutrition to a large extent in order to optimize the production. Despite this high level of control, a constant production throughout the year or immediate response of production to market demand is still impossible at individual farms. There are several factors causing large variation in quantity and quality of production. First, solar radiation, which is the most important growth factor, may strongly fluctuate from day to day. In the Netherlands (latitude: 52°), the daily radiation during summertime is about ten times higher than that during wintertime. These fluctuations have a strong impact on production, which can only partly be compensated for by controlling other growth conditions. Other climate conditions may vary as well, leading to variations in production. Besides environmental conditions, the developmental stage of the crop influences the production; young crops are usually not yet productive while older crops may become less productive because of ageing. This problem could be overcome by starting the growth of a crop at different times of the year.

Nowadays, the market demands guaranteed prices, guaranteed quality and continuous supply of horticultural products. For a good price setting, logistic arrangements and marketing in advance of harvest, it is of utmost importance to be certain of supply at a specific time. Modern farm management implies that quantity and quality of produce can be predicted and controlled.

The quantity of production of fruit vegetables, such as cucumber, sweet pepper and tomato, is determined by the growth of the fruits. Growth of fruits is

the result of dry matter production, dry matter partitioning among the plant organs and dry matter content of the fruits. An important quality aspect is the fresh weight of the individual fruits with each size grading having a different market price. Besides size, important quality aspects are shape, shelf-life and taste. In cucumber both young and old fruits have a short shelf-life due to respectively softening and yellowing.[1–3] Therefore, the developmental stage of the fruit at harvest is a main determinant of shelf-life.

Mechanistic models for simulation of crop growth have been developed successfully for a number of crops.[4] Such models might be powerful tools for prediction of production and product quality. In this paper the key principles of simulation models for greenhouse grown vegetables are presented. Possibilities for applying these models, their generic nature, and some of the future trends are discussed. Most of the examples used refer to a cucumber crop.

9.2 Key principles and methods

Often, a distinction is made between descriptive and explanatory models. Descriptive models, also called statistical, regression, empirical or black-box models, reflect little or none of the mechanisms that are the cause of the behaviour of a system, whereas explanatory models contain a quantitative description of these mechanisms and processes.[5] Explanatory models contain submodels at least one hierarchical level deeper than the response to be described, e.g. crop growth is described at the level of crop photosynthesis and leaf area expansion. At the lowest hierarchical level, submodels in an explanatory model are often descriptive again. The model's ability to explain is limited by its number of hierarchical levels.[6] Although the explanatory crop growth models in horticulture do, to some extent, reflect physiological structures, they do not incorporate the complete current knowledge of the biochemical relationships that we have for many of the mechanisms at the cellular level. If they did, the models would be impossible to manage and to use for predictions and analysis at the crop level.

Descriptive models are fast and usually have few parameters, which is important if crop models are to be used in on-line greenhouse climate control.[6, 7] Furthermore, model parameters are relatively easy to estimate.[6] Although the predictive value of descriptive models can be high, because they take into account implicitly all unknown effects as well, there are important limitations. When descriptive models are used outside the range of input data on which they are based, predictions usually become unreliable. Furthermore, extrapolation of these models to other species or locations is often impossible, and adding new input factors to such a model means building a new model based on an extended data set.

Explanatory models, more than descriptive models, allow for testing hypotheses, synthesizing knowledge and a better understanding of complex systems. Most explanatory models for crop growth are photosynthesis-based models.

9.2.1 Mechanistic crop models (photosynthesis-based models)

Crop production is the result of a complex system of interacting processes with both short-term and long-term responses (Fig. 9.1). Photosynthesis is often considered as the driving force for crop production. In photosynthesis-based models the interception of light by the leaf area is calculated first to simulate the gross photosynthesis (production of photosynthates) (Fig. 9.2). The rate of dry matter production (dW/dt) can be described by

$$dW/dt = Y_g^*(P_g - Rm)$$

where Y_g is the growth conversion efficiency, which is the weight of dry mass formed per unit weight of assimilates; P_g is canopy gross photosynthesis; Rm is maintenance respiration. Subsequently, the partitioning of dry matter among the different plant organs is calculated and finally the fresh weight can be estimated from the dry weight. For greenhouse grown vegetables many photosynthesis-based models have been developed.[4]

Although vegetables are usually sold on a fresh weight basis, most of the photosynthesis-based models simulate growth in terms of dry mass rather than fresh weight yield. Moreover, quality is seldom addressed by these models. Most

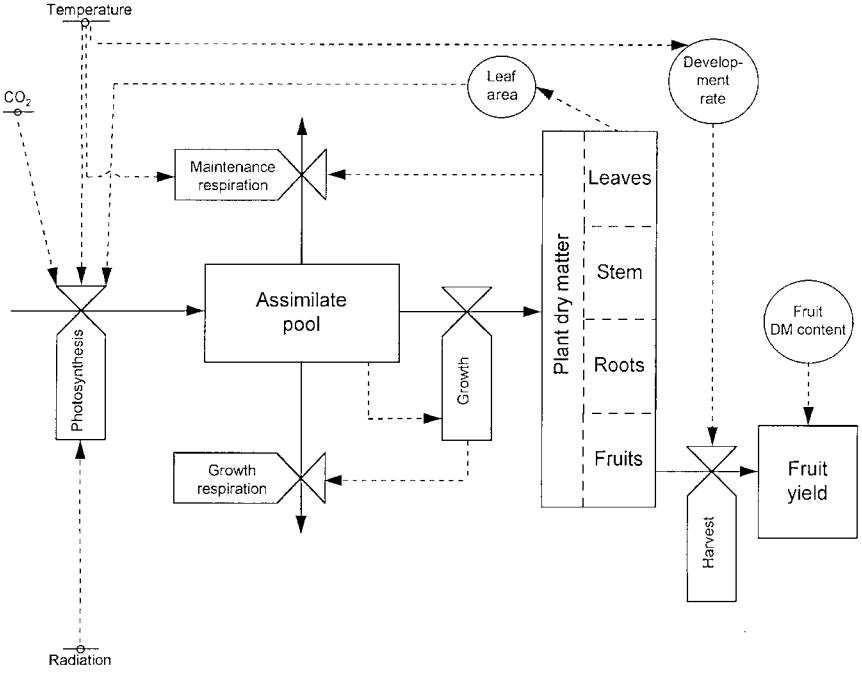


Fig. 9.1 A typical relational diagram for a growth model of a horticultural crop. Boxes are state variables, circles are parameters and valves are rate variables. Solid lines represent carbon flow and dashed lines represent information flow (DM = dry matter). From Marcelis *et al.* (1998).

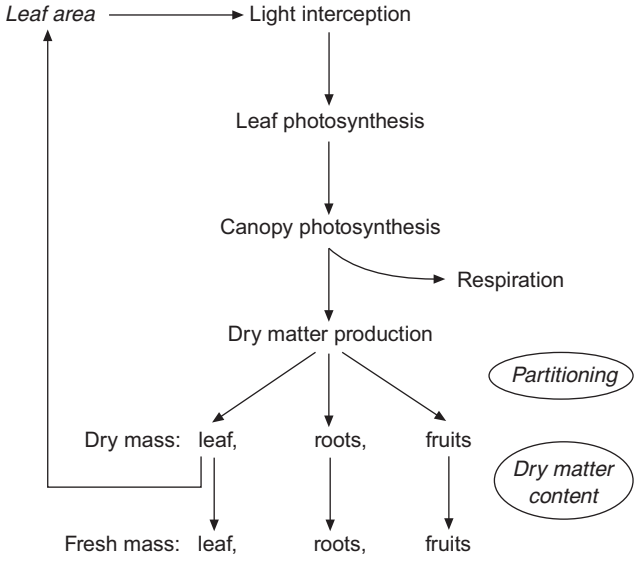


Fig. 9.2 Calculation steps of a photosynthesis-driven growth model.

mechanistic crop models have to be fed by a large number of input parameters, which is not feasible for large-scale commercial applications.

Recently Marcelis and Gijzen[8,9] developed a mechanistic model (KOSI 1.0) for prediction of the weekly fresh weight yield of cucumber fruits and the fresh weight and developmental stage of the individual fruits at harvest, the latter two being major criteria of fruit quality.

9.2.2 A model for crop growth and yield prediction of cucumber (KOSI)

The model (KOSI 1.0), as described by Marcelis and Gijzen,[8,9] is primarily based on the model INTKAM for simulation of dry matter production[10] and the model of Marcelis[11] for simulation of dry matter partitioning and fruit growth. The model consists of modules for greenhouse climate, greenhouse light transmission, light interception by the crop, leaf and canopy photosynthesis, assimilate partitioning, dry matter production, fruit growth, fruit dry matter content and fruit harvest.

Greenhouse climate and greenhouse light transmission, light interception and photosynthesis are calculated at five moments during the day. These moments are selected such that from these data the daily photosynthesis can be calculated using the Gaussian integration method.[12] The time step for the modules for assimilate partitioning, dry matter production, fruit growth, dry matter content and fruit harvest is one day.

A simple module for greenhouse climate calculates the daily temperature and daytime CO_2 concentration as a function of outside radiation, outside

temperature and set-points for temperature and CO_2. Light transmission is calculated for Venlo-type glasshouses according to Bot,[13] assuming 70% transmission of diffuse light. Light interception and canopy gross photosynthesis are calculated for a multi-layered uniform canopy according to the SUCROS model,[12, 14] while leaf gross photosynthesis is calculated for the various layers in the canopy with the biochemical model of Farquhar *et al.*[15] as described by Gijzen.[16]

Leaf area index (LAI) is calculated as a function of the daily temperature sum after planting. Net assimilate production results from the difference between canopy gross photosynthesis and maintenance respiration. Maintenance respiration is calculated as a function of fruit dry weight, vegetative dry weight and temperature, according to Spitters *et al.*[17]

Assimilate partitioning between vegetative parts and individual fruits is simulated on the basis of the concept of sink strengths, as described by Marcelis.[11] In this concept the fraction of assimilates partitioned into an organ is calculated proportional to the ratio between its potential growth rate (sink strength) and that of all plant parts. Dry matter production of the organs is calculated as the amount of assimilates partitioned into each organ divided by the assimilate requirement for dry matter production.

Flower formation is calculated as a function of temperature and radiation.[11] Fruit set is a function of the source/sink ratio (ratio between total dry matter production and total potential growth rate), according to Marcelis.[11] Fresh weight of the fruits is calculated from the dry weight and dry matter content. According to Marcelis[18] and Marcelis *et al.*[4] the dry matter content (DMC, %) is calculated as a function of temperature sum after anthesis of the fruit, temperature, fruit weight and the source/sink ratio of the past five days.

A fruit is harvested when its fresh weight exceeds a threshold value, which value depends on the season (380 g in summer and 300 g in winter) and decreases with increasing temperature sum after anthesis of that fruit. This reflects that growers are used to harvesting a fruit when it has reached a certain size, but that they harvest older fruits at a smaller size. When fruits grow slowly they attain a high temperature sum from anthesis before they reach the threshold weight for harvest. When the temperature sum of a fruit increases, the threshold weight for harvest decreases, leading to harvest of smaller fruits when growth rates are low. Moreover, at a high fruit growth rate the average weight will exceed the threshold weight more than at a low growth rate, because the weight increase between two successive harvests is bigger when the fruit growth rate is high. A random error (coefficient of variation was 0.1) is added to the threshold weight, as it is assumed that growers do not estimate the fruit weight exactly. Fruits are harvested three times a week.

Second-class fruits are fruits with undesired shape or colour or with abnormalities. The fraction of second-class fruits mainly depends on the age of the crop.[9] The model, in fact, simulates an average plant, although there is a great variation among plants. To overcome this to some extent, simulation results are the average of calculations for a plant receiving 90% and a plant

receiving 110% of the average net assimilate production. Moreover, as the simulated harvested fresh weight per week of one plant fluctuated much more than the average of a whole crop as measured by the growers, the modelled harvested fresh weight is smoothed by calculating moving averages of three weeks.

KOSI 1.0 outputs the total weekly harvest yield at individual farms (kg m^{-2} $week^{-1}$ and number m^{-2} $week^{-1}$), the weight and developmental stage of the individual fruits at harvest and the percentage second class fruits. As Marcelis and Baan Hofman-Eijer[19] have shown a close relationship between temperature sum and developmental stage of a cucumber fruit, temperature sum after anthesis (base temperature was 10ºC) is used as a measure for developmental stage of a fruit. At the planting of the crop KOSI can already predict the harvest yield for the whole growing period. It is possible to adjust these predictions every week using actual weather data of the previous week or forecast weather for the coming week.

The minimum data required by KOSI for harvest prediction are date of planting of the crop and scheduled date for the last harvest. When only this minimum data set is used, calculations are based on long-term average Dutch weather data, i.e. weekly global radiation and temperature outside the greenhouse. When the distance to the Dutch coast is provided, weather data will be corrected for higher radiation levels close to the coast. Instead of using long-term average weather data, predicted or measured weather data can be provided as input. To keep the use of the model as simple as possible, average values are assumed for all other factors, e.g. climate control (set-points for temperature and CO_2 of the greenhouse air), light transmission of the greenhouse, harvest strategy (frequency of harvesting and whether small-, medium- or large-sized fruits are harvested), plant density, number of fruits retained on the main stem, plant size at planting, cultivar properties and ratio between total and cropped greenhouse area. Depending on the objective or availability of data, the model calculations can be based on more or less input data.

9.2.3 Crop models including water and nutrient relationships

For yield predictions a time step of one day is suitable (but using a Gaussian integration method during the day) and leads to reliable results. For optimization of the greenhouse climate and water and nutrient relationships a shorter time step (minutes or hours) is desired. However, climate input data are not always available for shorter time intervals. The above described model for yield prediction assumes no limitations in plant water or plant nutrient relations. For modern greenhouses these assumptions seem to hold very well as long as we only consider the quantity of production. However when we are dealing with energy saving or preventing nutrient and water losses, the model should also consider plant water and nutrient relationships. Moreover, when more sophisticated models for product quality are developed, the simulation of water and nutrient relationships will be essential.

We have developed models comparable to the KOSI model where, in parallel to the leaf photosynthesis, the leaf transpiration is calculated using the Penman-Monteith equation,[10, 20, 21] and where water content of the plant follows from the difference in transpiration and water uptake.[20] Incorporation of nutrient relationships in growth models of greenhouse crops has, until recently, not had that much attention from modellers.[22] Marcelis *et al.*[21] described a preliminary crop growth model that included uptake and plant content of all macronutrients. In the model the nutrient demand of each plant organ is calculated as the product of dry weight growth and a demanded nutrient concentration. This is comparable to the approach of Mankin and Fynn.[23] The demanded or maximum nutrient concentration is the concentration the plant itself aims at when there would be no limitation in the nutrient supply. The uptake process of anions (NO_3, PO_4, SO_4) is actively regulated by the plant. When there is no limitation in nutrient supply, the simulated uptake by the plant will be according to its demand: for these ions, limitation usually only occurs when they are completely depleted from the root surface (see, e.g., Ingestad[24]).

The uptake of cations (K, Ca, Mg, Na) is to a great extent related to the cation-anion balance of the plant, activity (concentration) in the root medium and water uptake.[25–27] Total uptake of cations is simulated as a function of anion uptake. As a first approximation the uptake of Ca^{2+} and Mg^{2+} is modelled as a function of the water uptake and cation concentration in the solution at the root surface. Uptake of K^+ and Na^+ is a function of the calculated total cation uptake, Ca^{2+} and Mg^{2+} uptake and their concentration in the nutrient solution.

9.3 Areas of application

Models are powerful tools to increase efficiency of experiments, to test hypotheses, to synthesize knowledge, to describe and understand complex systems, to compare different scenarios and to teach students about complex systems. Moreover, models may be used in decision support systems, greenhouse climate control and production planning. Consequently, the interest in modelling of growth and yield of horticultural crops is still increasing as indicated by the increasing number of horticultural publications that are dealing with models, as listed in the CAB Abstract Database.

In general, a successful introduction of models in horticultural practice is only possible on a step-by-step basis.[28] Users of the models must be involved from the very start of the effort. Information should be represented in a simple, straightforward way. However, a step-by-step approach is not always possible, e.g. when a complete new concept is introduced. An example is the shift from traditional climate control where the grower supplies set-points to the computer, towards climate control based on controlling processes and set-point generation by a model. In such situations, the step-by-step approach may be reflected in using the new concept first as a learning tool: the model shows the grower what the alternative control system would do in the present situation and the grower

can decide to take over this strategy or to follow his old strategy. Hence, the new control concept runs in 'background mode' only, giving advice. When confidence is gained, the new control concept may replace the old one.

Model applications in horticulture have been discussed extensively by Lentz.[29] He states that in order to implement models for decision support, it is not sufficient to know the potential problems, but it is also necessary to understand the decision-making process. Furthermore, he concludes that the use of models in practice will only increase if the models deal with problems faced by the decision makers, and if it becomes obvious to the farmers that they can derive answers to their problems in a more efficient way.

9.3.1 Decision support to the grower

Three levels of decision are often distinguished depending on time scale: strategic (one–many years), tactical (months–one year) and operational decisions (minutes–24 h).[30, 31] At the strategic level, models may be used to decide on capital investments in durable equipment, e.g. energy screens, artificial lighting or heat storage tanks for CO_2 enrichment.[32] At the tactical level, before the start of a new cultivation, models can help a grower to decide on, for example, planting date, plant density and the amount of labour needed. The tactical plan should provide the framework for the operational level. At the operational level, control of climate, water and nutrient supply and operation of artificial lighting may be supported by models. Crop photosynthesis is a major (fast reacting) yield determining process and therefore it is a suitable criterion to evaluate short-term effects of climate control in greenhouses. Van Henten[33] and Van Willigenburg *et al.*[34] showed how a photosynthesis-based model can be used for optimum control of the greenhouse climate.

9.3.2 Prediction of yield and quality

To test whether the results of the model KOSI correspond to reality, the weekly cucumber production of commercial Dutch growers (ten growers in 1996 and fourteen in 1997) was calculated using actual climate data. The average harvest for all growers calculated by the model agreed closely with the actual harvest: the actual average production of ten growers in 1996 was 71.8 kg m^{-2}, while the model calculated 72.0 kg m^{-2}, an error of only 0.3%. The average weekly error was 0.17 kg m^{-2}, or 12.6%. A more detailed validation of the model when actual climate data were used as input to the model was presented by Marcelis and Gijzen.[8]

Subsequently, the accuracy of the model's predictions for the same group of growers was tested when only the first and last week of each crop cycle and distance from the Dutch coast were input. For outdoor temperature and radiation, long-term weather data (average of 30 years data from De Bilt, The Netherlands, corrected for distance to the coast) were used. Average values were assumed for climate set-points and farm management. The average production

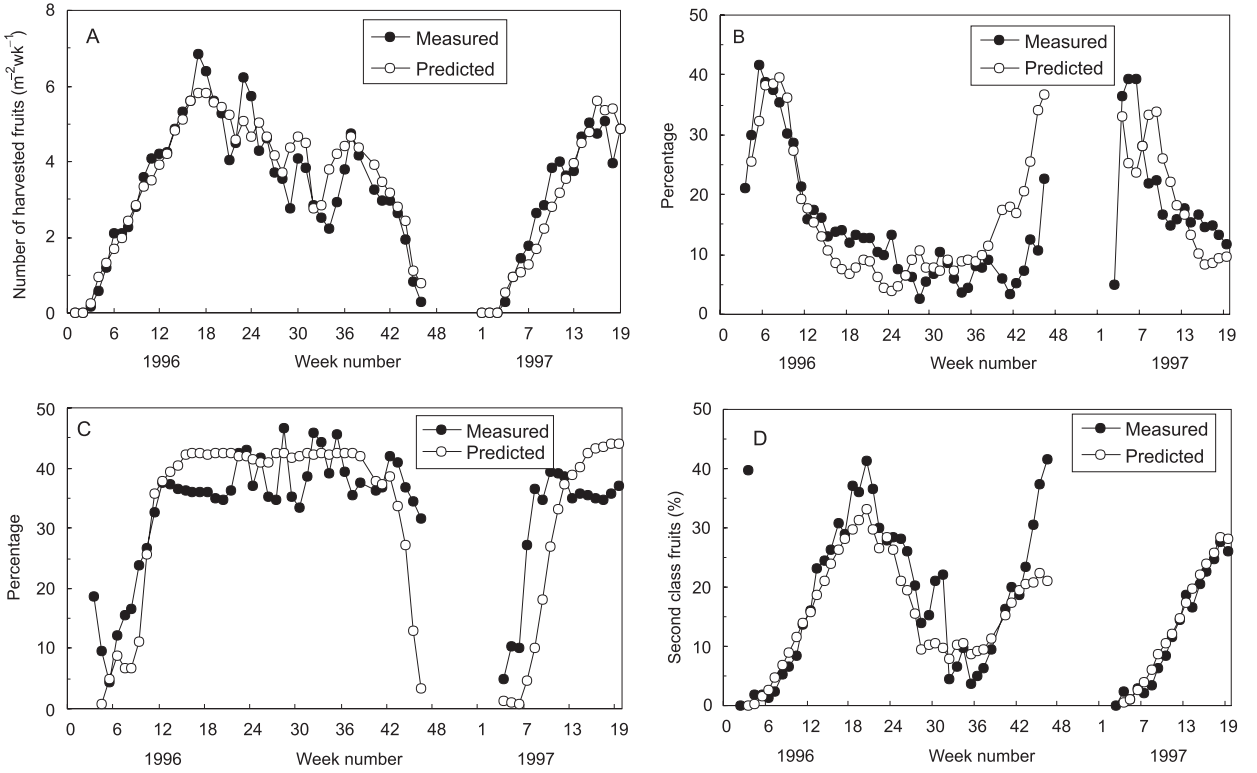


Fig. 9.3 Predicted and measured weekly harvest in 1996 and 1997. Measured data are the means of ten (1996) or fourteen (1997) commercial growers.

A. Fruit number B. Percentage of fruits in size class 310–360 g fresh weight per fruit C. Percentage of fruits in size class 510–610 g fresh weight per fruit D. Percentage of second class fruits Fig. 9.3A and 9.3D are from Marcelis and Gijzen.[9]

of number of fruits and total fresh weight by the group of growers corresponded well with the model's predictions (Fig. 9.3A). The average production in 1996 was 71.8 kg m^{-2}, while the model predicted 73.8 kg m^{-2}, an error of 2.8%. The average weekly error in 1996 was 0.19 kg m^{-2}, or 14.9%. The production between week numbers 33 and 36 was overestimated by the model, because the start of production for some growers was not predicted correctly by the model. One factor contributing to this error was that during this period of the year the initial production is quite different depending on whether the crop is planted on a Monday or a Friday, while only the week number was provided as input for the model. As discussed by Marcelis and Gijzen,[8] discrepancies between model and measurement for individual growers were larger than for the average of all growers. For a reliable prediction at individual farms more input data are necessary.

The average fresh weight of the individual fruits was predicted well by the model KOSI, except for a distinct underestimation at the end of the growing season (October–November). Probably the harvest strategy of growers changed towards the end of the year (two instead of three harvest days per week), while the model assumed the same harvest strategy throughout the year. Further analyses of harvest strategy and fruit growth in this period are needed to improve these predictions. The average weekly error of average fruit weight was 6.5%. As an example, the percentage of fruits in two of the main size classes is presented in Fig. 9.3B and 9.3C.

The amount of second-class fruits, which changes markedly during a year was predicted quite well by the model (Fig. 9.3D). The average weekly error of the percentage of second-class fruits was 5.3% in 1996. In the same period (the end of the growing season) when the average fresh weight was underestimated, the percentage of second-class fruits was underestimated.

The results show that a mechanistic photosynthesis-based dynamic growth model can be applied for accurate predictions of cucumber yield and fruit size. Such predictions can be used for optimizing the price setting, logistic arrangements and marketing of the produce. As growth as well as farm-specific conditions are variables of the model, such a model is also a promising tool to control the quantity and quality of the produce such that the production better fits to the demand by retailers or consumers. For instance, effects of future cultivation measures and climate set-points on crop growth and harvest can be calculated, so that the most appropriate strategy can be chosen.

To keep the use of the model as simple as possible, average values were assumed for most model parameters, as described before. If values of more parameters are available for individual growers, predictions could be improved. A cost-benefit analysis of gathering more input data for the model is an important step towards practical application of the model. The accuracy of the model might be improved further by combining model calculations with on-line measurements of crop status.

9.3.3 Combination of models and sensors

In specific situations, model results may deviate from actual data. This has hampered the introduction of (crop) models in control systems. By supplying the model feedback information from the actual system on some crucial plant parameters (e.g. light interception, leaf photosynthetic rate), the reliability of the model can be increased substantially.[21]

Sensor information *per se* is of limited importance for control of crop growth unless the consequences of the value of a measured parameter are quantified; for this a model is an ideal tool. Effects of a change in a value of a measured plant parameter on the final production or nutrient uptake may depend on many factors. For instance, an increase in temperature may lead to a decrease in net leaf photosynthesis. However, in young plants the simultaneous effects on leaf area expansion lead to a distinct increase in crop growth (Challa *et al*.[35]; Schapendonk, unpublished data). A crop model can be used to quantify these effects. Furthermore, measurements of crop status like photosynthetic parameters or leaf temperature may be used to indicate the occurrence of stress in the crop, by comparing the measured value to the value calculated by a crop model for the prevailing climate conditions.

9.3.4 Closed system for water and nutrient management

In greenhouses water and nutrients are supplied in excess in order to ensure no shortage. This procedure leads to large losses of water and nutrients to the environment. Therefore, closed greenhouse systems are being developed to nullify losses of water and nutrients. As described in Marcelis *et al.*,[21] several research groups at Wageningen University and Research Centre (Plant Research International, ATO, IMAG, WU-Agricultural Engineering and Physics, WU-Horticultural Production Chains) and the private companies Van Vliet Pijnacker and Hydrion Wageningen have combined efforts in an ambitious multi-disciplinary research project on an on-line monitoring and control system for process water and crop growth in closed greenhouse systems. A schematic representation of the system being developed is presented in Fig. 9.4. In this system a dynamic crop model predicts, based on environmental conditions and crop properties, both crop growth as well as the demand of the crop for water and individual ions. The model considers physiological processes and gets feedback from actual on-line measurements of crop status. For this purpose, non-invasive sensors that measure important parameters of the crop model are being developed. These parameters include photosynthetic activity and capacity, light interception by the crop, leaf area index, water use and fresh weight growth of the crop. The crop model is linked to a substrate model. The latter model predicts the strategy of fertigation to meet the demand of the roots without creating an undesired accumulation of nutrients. This model gets feedback information from ion-selective sensors in the rooting medium. For this an on-line robust multi-ion sensor unit (for measurement of Na^+, K^+, Ca^{2+}, NH_4^+, Cl^-, NO_3^-, $H_2PO_4^-/HPO_4^{2-}$ and SO_4^{2-}) is being developed. To supply the

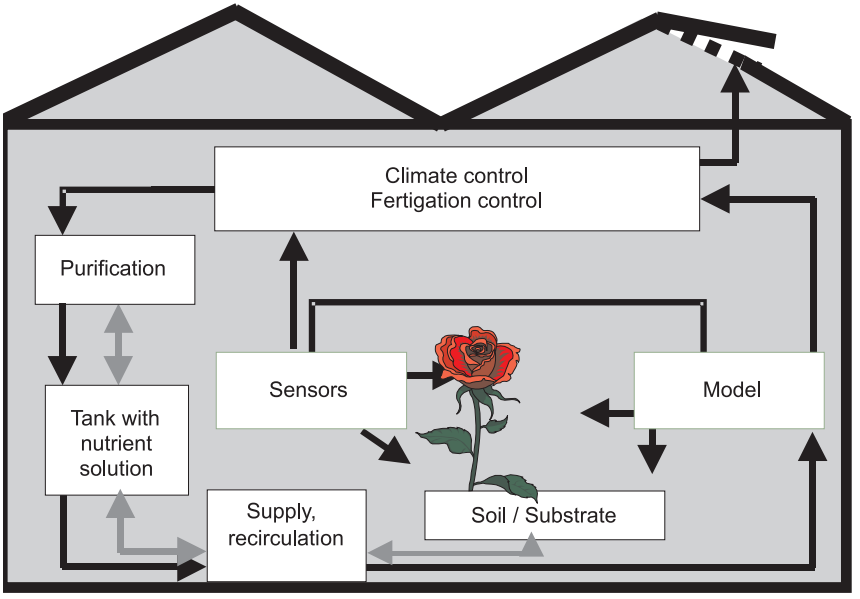


Fig. 9.4 Schematic representation of an on-line monitoring and control system for process water and crop growth in closed greenhouse systems.

individual ions in proportion to the dynamically varying crop demand, an installation is being developed that enables real-time adaptations of the supply of the different ions. The system aims at minimizing pollution by controlling the input of resources rather than by end-of-pipe measures. As a safeguard, a purification system for selectively removing ions from the recirculating nutrient solution is being developed. All aspects mentioned above will be integrated in a hardware and software platform enabling a real-time control system and an off-line data-management system.

9.4 How generic are crop models?

The photosynthesis-based models described in this paper have a modular structure and are based on physiological processes. This allows these models to be adapted to other crops relatively easily. In many cases this can be done by changing model parameters. Modelling of canopy photosynthesis is largely independent of plant species.[4] Plant type comes in at the parameter values of leaf photosynthetic properties and of radiation interception such as extinction coefficient and leaf angle distribution. In contrast to canopy photosynthesis, there seems to be a great diversity in the way a crop partitions its assimilates,[4] in developmental aspects (such as formation of organs) and in quality. Consequently, the simulation models available at the moment are rather species specific with respect to dry matter partitioning and developmental aspects.

Fruit vegetables such as cucumber, sweet pepper, tomato and eggplant have comparable growth habits; therefore only minor modifications are needed to adapt the model to each crop. In fact model adaptation is a matter of parameterization for these crops. Differences between cultivars of the same species can usually be modelled by changing a few model parameters. That similar models work very well for the different fruit vegetables has been illustrated by several authors (e.g. Gijzen[10]; De Koning[36]; Heuvelink[37,38]; Marcelis and Gijzen[8]; Marcelis *et al.*[4,21]).

A model is a simplified description of part of reality and for several variables it assumes average conditions. Consequently, in specific situations, model results may deviate from actual data. As discussed before, by supplying the model with on-line sensor information on some crucial plant parameters (e.g. light interception, leaf photosynthetic rate), the reliability of the model can be increased substantially. In addition, self-learning techniques can be adopted in order to ensure not only feedback from on-line sensors, but also from databases containing historic data sets on production at individual farms.

Almost all crop models are deterministic models simulating one average plant while there might be a large variation among individual plants of a canopy. For most purposes the average plant will do perfectly. If we need information on the variation it is obvious that simulation of the average plant is not that useful. In order to consider plant to plant variation, a population of plants with different values for some key parameters (for instance parameters on fruit set and fruit harvest) could be simulated. Then an average of these plants and confidence intervals could be calculated. Examples of models considering the plant to plant variation are still rare. Pearson *et al.*[39] considered the variation among plants in their stochastic model of fruit set in tomato. As discussed before, the model KOSI considers some aspects of plant to plant variation.

9.5 Some future trends

Techniques to provide crop models with feedback information of the actual situation will gain increasing interest. For optimal control of the production process in greenhouses, the growth conditions should be controlled such that they fulfil the demands of the plant. The demands of the plant can be quantified by the combined use of plant sensors and models. In fact, this refers to the speaking plant concept, a concept that was first discussed some years ago.[40–42] Only now have sensor technology and models reached a stage that enables actual application of the speaking plant concept. Not only on-line sensors but also manual registrations (e.g. recording dates of flowering or counting of numbers of leaves) and historic data sets will be used to provide models with feedback information. Then self-learning techniques will be adopted to improve model predictions.

The modelling of quality of horticultural products is still in its infancy, especially with respect to internal quality aspects. The problem for modelling quality is that there are very many different quality attributes, they are often

species specific, not well defined or cannot be measured easily in a quantitative and objective way. Due to our increasing understanding of quality and because of the importance of product quality, it is expected that the number of (mechanistic) models simulating quality aspects will increase rapidly.

Most of the process-based models lack a description of plant topology and geometry. Techniques to generate 3D virtual plants by morphogenetic models have also been developed.[43,44] There might be good opportunities to link these two types of models. This might be especially relevant for floriculture where morphology is an important quality feature.

Most of the photosynthesis-based models simulate growth in terms of dry mass rather than fresh weight yield, while vegetables are usually sold on a fresh weight basis. Often a constant conversion coefficient from dry to fresh weight is used. However, the water content of crops may vary substantially.[4] Future research should pay more attention to the regulation of water content as it is important for an accurate simulation of fresh weight production as well as an important quality attribute.

Production of produce is part of a chain from breeding to consuming. Growers can only produce profitably when their production is well embedded in a production chain. Consequently, models should not be limited to the growth of the crop, but linked also to other parts of the chain. It is striking that the development of models for product quality is still quite separate for pre- and post-harvest. In future, focus should be on models simulating the product quality as a continuous process throughout the production chain.

In general, Internet techniques and applications are developing rapidly. In the near future we may expect a rapid increase in Internet applications in horticulture as well. Some possible Internet applications have been described recently with respect to Internet integrated decision support systems to assist growers,[45] image databases for remote control of plant growth,[46] or virtual greenhouses for training, controlling and managing.[47]

9.6 Summary

Modern farm management implies that quantity and quality of produce can be predicted and controlled. Mechanistic models for simulation of crop growth have been developed successfully for a number of crops. Such models might be powerful tools for prediction of production and product quality. In this paper the key principles of simulation models for greenhouse grown vegetables, with special emphasis on cucumbers, are presented. Mechanistic crop models are described, consisting of modules for greenhouse climate, greenhouse light transmission, light interception by the crop, leaf and canopy photosynthesis, assimilate partitioning, dry matter production, fruit growth, fruit dry matter content and fruit harvest as well as modules for plant water relations and plant nutrient relations. The application of crop models is discussed with special emphasis on yield and quality prediction, decision support to growers and

minimizing of nutrient and water losses to the environment. The importance of feedback information from sensors or manual registrations and also of minimizing the number of input data is discussed. Mechanistic models that have a modular structure can be adapted relatively easily to different species. The modelling of canopy photosynthesis is largely independent of plant species, but dry matter partitioning, developmental aspects and quality attributes tend to be species specific. Some future trends in crop modelling are discussed including the combination of models and sensors, self-learning techniques, mechanistic models for product quality, virtual plants, Internet applications and crop models integrated in models for the entire production chain.

9.7 References

1. KANELLIS A K, MORRIS L L, SALTVEIT M E, Effect of stage of development on postharvest behavior of cucumber fruit. *HortSci*, 1986 **21** 1165–7.
2. LIN W C, EHRET D L, Nutrient concentration and fruit thinning affect shelf life of long English cucumber. *HortSci*, 1991 **26** 1299–1300.
3. JANSE J, Effect of growing methods on the incidence of rubber necks in cucumber fruits. *Acta Hortic*, 1994 **379** 281–8.
4. MARCELIS L F M, HEUVELINK E, GOUDRIAAN J, Modelling biomass production and yield of horticultural crops: a review. *Scientia Hortic*, 1998 **74** 83–111.
5. PENNING DE VRIES F W T, JANSEN D M, TEN BERGE H F M, BAKEMA A, Simulation of ecophysiological processes of growth in several annual crops. *Simulation Monographs 29*, PUDOC, Wageningen, 1989, 271pp.
6. LARSEN R U, Plant growth modelling by light and temperature. *Acta Hortic*, 1990 **272** 235–42.
7. VAN STRATEN G, On-line optimal control of greenhouse crop cultivation. *Acta Hortic*, 1996 **406** 203–11.
8. MARCELIS L F M, GIJZEN H, Evaluation under commercial conditions of a model of prediction of the yield and quality of cucumber fruits. *Scientia Hortic*, 1998 **76** 171–81.
9. MARCELIS L F M, GIJZEN H, A model for prediction of yield and quality of cucumber fruits. *Acta Hortic*, 1998 **476** 237–42.
10. GIJZEN H, Development of a simulation model for transpiration and water uptake and an integral growth model. AB-DLO Report 18, AB-DLO Wageningen, 1994, 90pp.
11. MARCELIS L F M, A simulation model for dry matter partitioning in cucumber. *Ann Bot*, 1994 **74** 43–52.
12. GOUDRIAAN J, VAN LAAR H H, *Modelling Potential Crop Growth Processes*. Kluwer Acad. Publishers, Dordrecht, 1994, 238pp.
13. BOT G P A, Greenhouse climate: from physical processes to a dynamic model. Dissertation, Wageningen Agricultural University, Wageningen, 1983, 240p.

14. SPITTERS C J T, Separating the diffuse and direct component of global radiation and its implications for modeling canopy photosynthesis. II. Calculation of canopy photosynthesis. *Agric For Meteor*, 1986 **38** 231–42.
15. FARQUHAR G D, VON CAEMMERER S, BERRY J A, A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta*, 1980 **149** 78–90.
16. GIJZEN H, Short term crop responses. In: J.C. Bakker, G.P.A. Bot, H. Challa and N.J. van de Braak (Eds), *Greenhouse Climate Control, an Integrated Approach.* Wageningen Pers, Wageningen, 1995, pp. 16–35.
17. SPITTERS C J T, VAN KEULEN H, VAN KRAALINGEN D W G, A simple and universal crop growth simulator: SUCROS87. In: R. Rabbinge, S.A. Ward and H.H. Van Laar (Eds), *Simulation and Systems Management in Crop Protection. Simulation Monographs* 32. PUDOC, Wageningen, 1989, pp. 147–81.
18. MARCELIS L F M, Non-destructive measurements and growth analysis of the cucumber fruit. *J Hort Sci*, 1992 **67** 457–64.
19. MARCELIS L F M, BAAN HOFMAN-EIJER L R, Effect of temperature on the growth of individual cucumber fruits. *Physiol Plant*, 1993 **87** 313–20.
20. MARCELIS L F M, Simulation of plant-water relations and photosynthesis of greenhouse crops. *Scientia Hortic*, 1989 **41** 9–18.
21. MARCELIS L F M, VAN DEN BOOGAARD R, MEINEN E, Control of crop growth and nutrient supply by the combined use of crop models and plant sensors. *Proceedings Int. Conf. Modelling and control in agriculture, horticulture and post-harvested processing.* IFAC, 2000 351–6.
22. LE BOT J, ADAMOWICZ S, ROBIN P, Modelling plant nutrition of horticultural crops: a review. *Scientia Hortic*, 1998 **74** 47–81.
23. MANKIN K R, FYNN R P, Modeling individual nutrient uptake by plants: relating demand to microclimate. *Agric Systems*, 1996 **50** 101–14.
24. INGESTAD T, Relative addition rate and external concentration; driving variables used in plant nutrition research. *Plant, Cell and Environment*, 1982 **5** 443–53.
25. NIELSEN N E, HANSEN E M, Macro nutrient cation uptake by plants. II. Effects of plant species, nitrogen concentration in the plant, cation concentration, activity and activity ratio in soil solution. *Plant and Soil*, 1984 **77** 347–65.
26. HAYNES R J, Active ion uptake and maintenance of cation-anion balance: a critical examination of their role in regulating rhizosphere pH. *Plant and Soil*, 1990 **126**, 247–64.
27. ADAMS P, Effect of diurnal fluctuations in humidity on the accumulation of nutrients in the leaves of tomato (*Lycopersicon esculentum*). *Journal of Horticultural Science*, 1991 **66** 545–50.
28. HEUVELINK E, MARCELIS L F M, Model application in horticultural practice: summary of discussion groups. *Acta Hortic*, 1998 **456** 533–4.
29. LENTZ W, Model applications in horticulture: a review. *Scientia Hortic*, 1998 **74** 151–74.

30. CHALLA H, Prediction of production: requisite of an integrated approach. *Acta Hortic*, 1988 **229** 133–41.
31. VAN STRATEN G, CHALLA H, Greenhouse climate control systems. In: J.C. Bakker, G.P.A. Bot, H. Challa and N.J. van de Braak (Eds), *Greenhouse Climate Control, an Integrated Approach.* Wageningen Pers, Wageningen, 1995, pp. 249–61.
32. RIJSDIJK A A, HOUTER G, Validation of a model for energy consumption, CO_2 consumption and crop production (ECP-model). *Acta Hortic* 1993 **328** 125–31.
33. VAN HENTEN E J, Greenhouse climate management: an optimal control approach. Dissertation, Wageningen University, Wageningen, 1994, 329pp.
34. VAN WILLIGENBURG L G, VAN HENTEN E J, VAN MEURS W T M, Three time-scale digital optimal receding horizon control of the climate in a greenhouse with a heat storage tank. *Proceedings Int. Conf. Modelling and control in agriculture, horticulture and post-harvested processing*. IFAC 2000 184–9.
35. CHALLA H, HEUVELINK E, VAN MEETEREN U, Crop growth and development. In: J. C. Bakker, G. P. A. Bot, H. Challa, N. J. van de Braak (Eds) *Greenhouse Climate Control: an Integrated Approach.* Wageningen Pers, Wageningen, 1995 pp. 62–84.
36. DE KONING A N M, Development and dry matter distribution in tomato: a quantitative approach. Dissertation, Wageningen Agric. Univ., Wageningen, 1994, 240pp.
37. HEUVELINK E, Dry matter production in a tomato crop: measurements and simulation. *Ann Bot*, 1995 **75** 369–79.
38. HEUVELINK E, Evaluation of a dynamic simulation model for tomato crop growth and development. *Ann Bot*, 1999 **83** 413–22.
39. PEARSON S, HADLEY P, WHELDON A E, DUNGEY N, A stochastic model of truss set in a long-season tomato crop. *Acta Hort*, 1996 **417** 33–40.
40. HASHIMOTO Y, MORIMOTO T, FUKUYAMA T, Some speaking plant approach to the synthesis of control system in the greenhouse. *Acta Hortic*, 1985 **174** 219–26.
41. TAKAKURA T, KOZAI T, TACHIBANA K, JORDAN K A, Direct digital control of plant growth. *Transactians ASAE*, 1974 **17** 1150–4.
42. TAKAKURA T, OHARA G, NAKAMURA Y, Direct digital control of plant growth III, *Acta Hortic*, 1978 **87** 257–64.
43. DE REFFYE P, BLAISE F, HOULLIER F, Modelling plant growth and architecture: some recent advances and applications to agronomy and forestry. *Acta Hortic*, 1998 **456** 105–16.
44. PRUSINKIEWICZ P, Modelling of spatial structure and development of plants: a review. *Scientia Hortic*, 1998 **74** 113-150.
45. IVEY J, KEENER H, SHORT T H, Internet decision support for hydroponic greenhouse tomato production. *Proceedings Int. Conf. Modelling and control in agriculture, horticulture and post-harvested processing.* IFAC 2000, 346–50.

46. HATOU K, TEKEUCHI T, HASHIMOTO Y, Image database for the remote control of plant factories in the internet age. *Proceedings Int. Conf. Modelling and control in agriculture, horticulture and post-harvested processing.* IFAC 2000 236–41.
47. SIGRIMIS N, ARVANITIS K G, PASGIANOS G, ANASTASIOU A, FERENTINOS K P, New ways on supervisory control: a virtual greenhouse: to train, to control and to manage. *Proceedings Int. Conf. Modelling and control in agriculture, horticulture and post-harvested processing.* IFAC 2000, 212–17.

10

Modelling and management of fruit production: the case of tomatoes

C. Gary and M. Tchamitchian, INRA, Avignon

10.1 Introduction: the contexts of tomato production

Tomato is the most popular fruit vegetable crop: a large range of cultivars enable the production of various fruit types. These can be consumed either as fresh vegetables or as processed products. The per capita consumption is increasing: its world average was 13.4 kg/pers/year in 1997.[1] The yearly world production in 1995 to 1999 ranged between 86 and 95 million metric tonnes per year.[1] On average, 28% of the crop is processed.[2]

10.1.1 Greenhouse versus field production

The double use of tomatoes (fresh and processed) has led to two major cultivation systems, one under cover and one in the field. Protected cultivation is specific to production for the fresh market. Its rationale is a gain in productivity. This goal can be achieved through application of transparent cover (plastic or glass) to reduce convective and radiative heat losses,[3] which increases the temperature around the growing crop. Productivity can be increased by extending the production period and by reducing the number of limiting factors through a better control of the physical and biological environment of crops.

A greenhouse can comprise various types of equipment to control the environment.[4] The temperature can be increased by heating, for example by burning natural gas, oil or coal, or by using thermal screens during the night. The temperature can be reduced by natural (vents) or forced (fans) ventilation, or by absorbing heat through evaporation of water applying cooling pads or fog systems, or by cooling the cover material by water sprinklers. The light level can be controlled with shading screens, by whitening the cover; by using roof

materials having a higher light transmission (light transmission of plastics decreases with ageing) and by applying supplementary lighting. Water vapour is released by crop transpiration. The air humidity can be decreased by ventilation, sometimes in combination with heating. It can be increased by evaporation of water using, for example, a fog system. The CO_2 concentration in the air can be increased (or maintained at normal level when greenhouses are closed and crop photosynthesis is active) by the injection of either industrial CO_2 or flue gases from the boiler. The latter option is only applicable if the flue gases are clean, for example, when the fuel is natural gas.[5] Some of these techniques (such as supplementary light, fog system, injection of industrial CO_2) are expensive and seldom used for tomatoes. It should be noted that the transpiration of the crop itself very effectively reduces the air temperature and increases the air humidity. In this respect, a proper management of the development of the canopy is a major contributor in controlling the climate.

In soilless culture, the root environment is continuously monitored and controlled (ion concentration, pH, no soil diseases). Roots may develop either in mineral (rock wool) or organic (coco peat) substrates or directly in the nutrient solution (Nutrient Film Technique). As the substrate can be replaced (mineral substrate can be recycled and organic substrate used as soil amendment), no soil disinfection is needed. In order to limit environmental pollution, growers are now encouraged to close the fertigation systems: the drained nutrient solution is pumped, disinfected and brought back to set point by replenishing the water and the nutrients.

Finally, protected cultivation facilitates the control of pests and diseases. The use of pesticides can be reduced or suppressed thanks to biological control. For greenhouse tomato crops, some natural enemies of the most damaging pests have been identified. The development of some diseases such as *Botrytis* can be avoided with proper control of humidity and temperature thereby limiting condensation on the foliage.

In contrast to greenhouse cultivation systems, the field cultivation system permits much less control of the physical and biological environment. The timing of operations can be adapted to allow the crop to grow at the most favourable climate conditions. The required nutrients can be provided either in one run before plantation or several times during crop growth. If necessary, water (possibly together with nutrients) is supplied by irrigation. Plasticulture systems equipped with drip irrigation allow the highest control of water and nutrient availability: a plastic cover spread out on the soil keeps rainfall off and limits soil evaporation. There is, of course, a large range of intermediate cultivation systems between the most sophisticated glasshouse and the most basic field cultivation system. For example, significant areas of tomato crops are cultivated on soil under cover. In this particular case, growers still have some control on the climate but the conditions of water and nutrient supply are close to those encountered in the field.

Different cultivars are used for the two cultivation systems. For long-season production under greenhouses (up to one year), *indeterminate* (with a vine

shape) cultivars are grown with all side shoots removed. New inflorescences continuously appear. As a consequence, irrespective of the season, mature fruits can be harvested two to three times per week and delivered to the fresh market. *Determinate* (with a bushy shape) cultivars are preferred in the field when the growing season is short or when the pest pressure is high. These plants have a grouped flowering and fruiting. This latter characteristic makes such crops suitable for mechanical harvesting.

10.1.2 The uses of crop modelling in greenhouse and field production

All the physical and biological processes involved in both cultivation systems can be formalized in different ways to carry out simulations, make predictions, or optimize their management. In practice, the modelling effort has been proportional to the ability to control the cultivation system, that is, much higher for greenhouse than for field production. In greenhouse production, models have a much wider range of applications. Yield prediction is needed to match the market requirements. Models of the greenhouse climate and of the crop carbon, water and nutrient balances are designed for the optimization of the climate and for the control of fertigation. Simulation of crop growth and development makes it possible to evaluate policies of crop management. In field production, modelling has been more dedicated to predict harvest date, to achieve a steady supply of product to the factories (organization of plantation schedules), and to estimate water and nutrient requirements (for scheduling irrigation and fertilization).

In this chapter we will review the processes that have been described and the methods that have been used for modelling tomato crops in relation to their environment. We will then consider the various areas of application of these models in the protected and field cultivation systems. Of course, a modelling approach (see also Chapter 2) is often closely linked to a specific type of application.[6]

10.2 Processes and methods of modelling tomato crops

The crop models presently available are based on two different approaches. On the one hand, new models appear with progressing knowledge as a mathematical formalization of observed phenomena and their related processes. Such models can be called *research models*. On the other hand, models can be designed to be part of procedures aimed at solving practical problems. In that case, they can be called *engineering models*.[7] Research models are evaluated on their scientific value (realism). They are explanatory or process-oriented models, as the behaviour of a simulated system at a particular hierarchical level is the result of processes described at lower hierarchical levels. The engineering models are evaluated on their operational value (effectiveness). They can be more descriptive, being built from statistical relations (‘black-box’ models) or knowledge-based (heuristic models).

In the literature, the majority of publications on modelling in horticulture report work on the processes of plant growth and development. The processes of water and nutrient uptake, of quality formation, the interactions between crops and pests or fungi have received much less attention.[6]

10.2.1 Mass and energy balances of tomato crops

Carbon

Basically, the production of biomass by a canopy relies on the net assimilation of atmospheric CO_2. The net assimilation is the balance between gross photosynthesis and respiration. It depends on the amount of available energy (light) and carbon substrate (CO_2), and on the ability of the canopy to intercept light and assimilate CO_2. In greenhouses, the assimilation of CO_2 is not only important for crop growth, it interacts strongly with the composition of the atmosphere. The daily consumption of carbon by a tomato canopy can be up to ten times the amount of carbon available in the greenhouse atmosphere.[5] It must be balanced by either ventilation or CO_2 enrichment.

Longuenesse *et al.*[8] and Gijzen[9] have reviewed extensively models on photosynthesis of horticultural species at leaf and canopy levels. Experiments on tomato have often been used to develop photosynthesis models. The leaf gross photosynthesis responds to light by a saturation-type curve. Various mathematical formulations have been proposed and tested on tomato data, for example the rectangular hyperbola,[10] the non-rectangular hyperbola[11] and the negative exponential.[12] Despite their slight difference in shape, all these functions include two important parameters: the maximum rate of leaf photosynthesis (P_{max}) and the initial (close to darkness) light use efficiency (α). P_{max} increases with CO_2 concentration and with the conductance to CO_2 transfer from the atmosphere to the chloroplasts. It is limited at low and high temperatures (see examples of parameterization for tomato in Bertin and Heuvelink[13]). Initial light use efficiency α is positively affected by CO_2 concentration and negatively by temperature. The conductance to CO_2 transfer gets lower at low light intensity, high CO_2 concentration, high vapour pressure deficit (VPD) and under water stress.[14]

Gross photosynthesis has been integrated at canopy scale in different ways. The simplest approach is to multiply the unit leaf activity by the leaf area index or by the projected leaf area ('big leaf' approach). Other models take the transmission of light in the canopy into account using an exponential law of extinction.[15] When the leaf light response curve is a rectangular hyperbola, analytical integration at canopy scale is possible (for example, in Jones *et al.*[16] for tomato crops). More sophisticated models provide a detailed simulation of light scattering and transmission through leaves, and of the distribution of diffuse and direct light based on a thorough description of the spatial distribution of the leaf area and leaf angle.[17] The specific case of row crops such as tomato has been addressed and reviewed by Critten.[15]

The respiratory efflux of CO_2 is significant: on a daily basis, it can represent a quarter to a half of the gross photosynthesis of a developed greenhouse tomato

crop.[9,5] Respiration of plants has been divided functionally in two components: maintenance respiration and growth respiration. Maintenance respiration corresponds to the energy needed to maintain the ionic gradients across biological membranes and the pools of macromolecules such as proteins. Growth respiration corresponds to the energy involved in the synthesis of new biomass from assimilates and minerals. Maintenance respiration is calculated as the product of the plant or organ dry weight times a maintenance coefficient. Growth respiration is calculated as the product of the plant or organ growth rate times a CO_2 production factor. In crop models, maintenance and growth respiration are summed up to estimate total respiration, generally on a daily basis. Respiration rate increases exponentially with temperature. Since this conceptual framework was proposed in the late 1960s, most research has been carried out on the accuracy and determination of the parameter values. The maintenance coefficient has been related to the tissue metabolic activity. For tomato, Heuvelink[18] has hypothesized that the maintenance coefficient decreases with ageing. The CO_2 production factor is proportional to the energy cost of biomass synthesis. It varies among organs and with ageing (see Gary *et al.*[19] for tomato). The modelling of plant respiration has recently been re-examined by various authors[20,21] looking for more mechanistic connections between the production of respiratory energy and ongoing processes in the growing plant (biosyntheses, translocation, ion uptake, N assimilation, protein turnover, ion-gradient maintenance).

The crop carbon balance includes the carbon exchanges between the atmosphere and the canopy (net photosynthesis), and the partitioning of carbon in the plant between one or several pools of photoassimilates and the growing organs. Gent and Enoch[22] put together simple formulations for gross photosynthesis and respiration, and provided a relation between availability of photoassimilates and growth, that is, production of the elaborate compounds of the plant tissues from the photoassimilates. With these simple formulations the 24-hour dynamics of CO_2 exchanges and assimilate pool of young tomato plants can be simulated.[23,24] Such a simple carbon balance model (Fig. 10.5) was reshaped for control purposes by Seginer *et al.*[25]

Water

The water balance in the crop is an important crop property in various respects. Water import contributes to the plant growth, as water status influences cell extension in growing organs and water flow conveys nutrients to growing or storage organs. Water status also partly controls the stomatal conductance and may therefore affect photosynthesis. At last, the evaporation of water during transpiration is connected to the absorption of latent heat: it strongly determines the temperature of the canopy and therefore, of the air in a greenhouse.[3]

The modelling of water relations of horticultural crops has been reviewed by Jones and Tardieu,[26] van de Sanden[27] and Jolliet.[28] Research in this domain has been motivated by two main concerns:

(1) simulating the water status and its relation with various physiological functions (organ extension, stomatal opening, water flux) and
(2) simulating the water flux through the canopy to estimate the water requirements of crops.

The basic framework that has generally been adopted is an analogue of Ohm's law: the water volume flux along a certain path is proportional to the gradient of water potential and to the inverse of a flow resistance. For tomato, van Leperen[29] designed a model describing the pathway of water from the root environment to the atmosphere through one root compartment and three shoot layers within a vegetative plant, and the dynamics of water potential in roots, stems and leaves. Premises of modelling the water fluxes to the tomato fruit through the phloem and xylem vessels can be found in Guichard *et al.*[30] These premises are based on Fishman and Génard's model.[31] The dominating phloem fluxes depend on the concentration of carbohydrates in the phloem vessels and on the ability of the fruit to unload these carbohydrates.[31] The xylem flux varies with the water potential in the stem, since the fruit water potential remains fairly stable in time and in different environmental conditions.[30] Due to a high resistance to water flux in its epidermis, the transpiration of the tomato fruit is limited. Recently, it was modelled as a function of irradiance and VPD by Leonardi *et al.*[32]

At the canopy scale, the transpiration in tomato crop has been modelled applying the classical Penman-Monteith approach[33] as the sum of a radiative component, proportional to the global radiation absorbed by the canopy, and of a convective component, proportional to the VPD. The canopy resistance to transfer water vapour comprises the aerodynamic resistance that depends on wind speed and air and leaf temperatures, and the stomatal resistance that depends on radiation, leaf air saturation deficit and leaf temperature (see, for example, Boulard *et al.*[34] for tomato crops). For operational purposes, the complete analytical model has been simplified to a two-parameter formula, the parameters being either derived from the complex model or identified *in situ*.[35]

Energy

A crop canopy can be compared to a solar collector. The absorbed radiation is the balance between incident, reflected and transmitted global radiation. In their study of light interception by glasshouse crops, Warren Wilson *et al.*[36] measured for a tomato canopy an average reflectance of 13% and an average transmittance of 23.5% of the incident light in the photosynthetic active radiation (PAR) waveband. Light absorption was improved by about 10% when the soil was covered with a white plastic sheet. Light absorption increased also with the foliage development to almost complete with a leaf area index (LAI) of 4 or above. It is also related to plant density and row spacing as it tends to increase when the plant distribution is more uniform.[37] The distribution of light and its absorption by rows of canopies such as tomato crops has been modelled by using several approaches (see review by Critten[15]). Among these are the exponential extinction curve, and various

models that take light scattering and the distribution of diffuse and direct light,[38] and leaf angle distribution into account.[17]

Part of the absorbed radiation is used by photosynthesis for carbon assimilation and biomass production. This proportion is estimated by the radiation use efficiency (RUE), that is, the ratio between the energy equivalent of biomass and the absorbed (or incident) global (or PAR) radiation. For a tomato crop, Aikman[39] estimated it to be about 7% when based on the absorbed PAR or 1.6% when based on the global radiation outside the greenhouse.

A significant part of the absorbed energy is actually dissipated by the crop as latent heat by transpiration. As a consequence, the temperature of a transpiring canopy is lower than the air temperature. This difference generates a flux of sensible heat from the air to the canopy. In a greenhouse, depending on the LAI, 50 to 70% of the solar energy input is used for evapotranspiration.[3] This justifies that the crop water requirements are estimated from the absorbed or incident global radiation.

Minerals

Nutrients are essential components of the plant tissues. Fertilization is a very basic cultivation technique to avoid any limitation of growth by the availability of minerals and to gain some control on yield and product quality. As for carbon and water, both mechanistic and black-box models have been designed (see the extensive review of Le Bot *et al.*[40]). The mechanistic models are research models describing specific processes like nutrient uptake, transport and assimilation. Even for nitrogen, the most studied element, the regulation and integration of these processes at a whole-plant scale are still in discussion. For tomato, two main approaches of mechanistic modelling have been proposed. According to Le Bot *et al.*,[40] the time-course of nitrate uptake is related to the translocation of carbohydrates to the roots to cover the energy cost of nutrient uptake. According to Cardenas-Navarro *et al.*,[41] nitrate uptake is related to the maintenance of a steady internal ion concentration.

More general (black-box) models link the demand of nutrients directly to the growth rate. It has been established for several elements (nitrogen, potassium, phosphorus) that a critical concentration in plant tissues should be maintained to approach the potential growth based on total intercepted radiation. For nitrogen, this critical concentration gradually declines with the accumulation of biomass during the vegetative phase.[42] Le Bot *et al.*[43] parameterized this relation for tomato plants. To explain this decline in nitrogen content, Caloin and Yu[44] suggested two compartments in the biomass, one mostly active for growth and having a high nitrogen content, and another dedicated to structures and storage having a lower nitrogen content. With crop development, the second compartment tends to dominate the first. This model was calibrated for a greenhouse tomato crop by Bellert *et al.*[45] A comparable approach of the nitrogen demand by processing tomatoes has been implemented in the EPIC model to evaluate different fertilization policies in terms of crop growth and nitrogen dynamics in the soil.[46]

Few models are presently available for other nutrients.[47] Only recently, a first model simulating the flux of calcium in pepper fruit and its relation to quality, measured as the occurrence of blossom-end-rot, was reported.[48]

10.2.2 Yield formation

Tomato has been a pioneer vegetable species for crop modelling. The formation of yield (organ appearance, dry matter production and partitioning) has been studied thoroughly and formalized with various approaches, again either empirical or mechanistic. The approach of fruit growth has been based on models of dry matter production. Water fluxes towards the fleshy tomato fruits (around 95% water) have been studied and modelled only recently.

Production of biomass

Different approaches to modelling biomass production have been developed for different crop species including tomato. In the 'photosynthesis-driven' models, integration of net photosynthesis and conversion of the resulting photoassimilates into biomass are used to compute the accumulation of dry matter. As already mentioned, net photosynthesis is the balance of photosynthesis minus respiration. The coefficient of conversion of assimilates into biomass depends on the energy value of the synthesized tissues. Gary *et al.*[19] have estimated its ontogenetic variation for the different types of tomato organs. Challa and Bakker[49] estimated the potential production of greenhouse crops in various regions of the world using this approach. It is also the first step in most of the tomato crop models.[16, 18, 50] Bertin and Heuvelink[13] compared the dry matter production estimated by Jones *et al.*'s[16] and Heuvelink's[18] models.

In the RUE approach, the production of biomass is considered as a sequence of energy conversions from the incident radiation to the energy content of biomass. Interception of (photosynthetically active) incident radiation is linked to the leaf area index by a saturation type curve; the coefficient of conversion of intercepted light into biomass is higher for C_4 (e.g. maize) than for C_3 (e.g. tomato) species and it increases at high CO_2 concentration. This approach was validated at different conditions for greenhouse tomato crops.[51–53] A similar approach has been used for different species including tomato in the STICS modelling platform.[54]

Timing of development

Development processes include the formation of new organs and their ageing and phase transitions at whole plant (e.g. vegetative vs. generative periods) or organ (e.g. fruit setting) scales. Formation and ageing of organs depend mainly on temperature,[55] following a bell-shaped curve that can be described by the Arrhenius equation.[56] Such a response curve has been calibrated for the formation of new leaves and trusses and for fruit development from flowering to maturity, and introduced in most tomato crop models (e.g. De Koning[57]). Under the hypothesis that the response of development rate to temperature can be considered

as linear in a limited range of temperature, daily temperatures can be summed to calculate a 'thermal time' expressed in degree-days that is, by definition, independent of the temperature regime.[58] (Tijskens and Verdenius[59] recently revisited the modelling of biological processes that depend on temperature.)

In tomato plants, fruit setting is the phase transition from flowering to fruit growth. It has been observed that the higher the source-sink ratio (that is, the fraction of the plant potential growth rate that can be covered by the current production of photoassimilates), the more successful fruit setting.[60] This relation was formalized in the TOMGRO model.[61] In this model, the dynamics of flowering, fruit setting and fruit ageing determines the age structure of the populations of vegetative and generative organs at any time during production.

Dry matter partitioning

The dry weight of harvested organs depends on the fraction of dry matter that is allocated to them. In the case of fruit species such as tomato, the vegetative-generative dry weight balance is a key component of crop models. This ratio can change with the plant development stage, and dynamically with the strength of vegetative and generative sinks. The sink strength of an organ or a group of organs is their ability to attract photoassimilates. It is the potential growth rate when no competition for carbon resources exists among organs.[62] It varies with the stage of development of the organ and increases with temperature. It is not affected by the availability of assimilates themselves. Heuvelink[63] demonstrated that, in tomato, the relative position of leaves and fruits on the plant does not affect dry weight ratio. In other words, all the organs of a tomato plant have the same access to the carbon resources. Consequently, (1) the vegetative-generative dry weight allocation ratio depends on the number and age structure of leaves, stem internodes and fruits, and (2) when the source activity (net photosynthesis) is lower than the sink demand, the actual growth rate of all organs is limited in the same proportions. These concepts have been implemented in the tomato crop models, designed for indeterminate cultivars. Heuvelink and Bertin[64] have compared two of them. Until now, only a few attempts[65] have been made to verify and validate this theory for determinate cultivars.

Dry matter content of fruit

Like most vegetable species, tomato fruits contain a high content of water at harvest. This water content is the result of xylem and phloem influxes and transpiration efflux during fruit growth. As mentioned earlier, the modelling of lateral fluxes within the plant (from stems to fruits) and of the fruit transpiration has only been studied quite recently. These processes will be introduced in a tomato crop model provided carbon and water fluxes can be coupled. To this end, the dynamics of water potential in the stem and of carbohydrate content in the phloem and the possible variations in water transport resistance in the fruit peduncle and epidermis have to be determined.

At present, tomato crop models are based on the assimilation and partitioning of carbon only. The dry weight of harvested fruits is calculated and converted into

fresh weight by applying a coefficient of dry matter content that is either fixed[16] or variable with the season.[57] In the latter case, the fruit dry matter content is higher in summer than in winter as the environmental conditions in summer tend to favour water stress (when radiation, VPD or salt concentration in the nutrient solution are high). The dry matter content of mature fruits is also genetically determined: it is generally higher in cultivars with small (cherry, cocktail) than with large fruits.

10.2.3 Other processes

Quality formation

The quality of tomato fruits covers a number of different characteristics. Which are the most important depends on the use of the products, whether for the fresh market or for the industry (Table 10.1) (see also Chapter 17). The average fruit fresh weight can be modelled based on the weight and number of harvested fruits. The fruit sink strength or potential growth rate is a genetic parameter. In tomato, it increases from cherry over cocktail cultivars to round and beefsteak cultivars. Within the range of genetically determined fruit sizes, the actual fruit grade obtained can be controlled in greenhouses by climate and crop management. Larger fruits can be obtained by increasing the source activity with, for example, CO_2 enrichment or by decreasing the competition for assimilates, for example by fruit pruning to a lower total fruit load. At present within the growing area, fruit grade is the only quality attribute that is properly simulated. For example, the SIMULSERRE simulator, based on the TOMGRO model, enables the evaluation of different strategies of climate and crop management in terms of yield, fruit grade, and energy and CO_2 consumption.[66]

Table 10.1 The quality variables of tomato fruits, their major sources of variability and their significance (* low to *** high) for the fresh market and the industry

Quality variable	Sources of variation				Fresh market	industry
	Genetics	Climate	Fertigation	Crop management		
Fruit grade	x	x		x	***	*
Uniform colour	x	x			***	***
Cracking		x	x	x	***	*
Blossom end rot		x	x	x	***	**
Shelf-life	x				***	*
Dry matter content	x	x	x	x	*	***
Sugar content	x	x	x	x	**	**
Acid content	x	x	x	x	**	**
Aroma content	x	x	x		**	**
Texture	x	x	x		***	*
Health value (antioxidants)	x	x		x	**	**

Ongoing research is conducted on the formation of the tomato fruit quality in terms of chemical composition (sugar, acid, aroma contents), appearance (colour, cracking, blossom-end-rot) and health promoting compounds (anti-oxidants). For these quality variables models are still largely unavailable, although some can be related to the carbon, water or mineral fluxes to the fruit. The sugar content could be linked to the carbon availability but acid or aroma contents could not.[67] The frequency of cracking of the fruit epidermis has been linked to the crop water status.[67] The occurrence of blossom-end-rot has been related to the calcium flux transported by the xylem network.[68]

Interactions with pests and diseases

Even though the effects of pests and diseases on crop behaviour is of major importance to tomato cultivation systems (especially in relation to environmental and health concerns), few simulation models of plant disease are available. Seghi *et al.*[69] reviewed some empirical models that forecast diseases (*Alternaria solani*, *Phytophtora infestans*) in tomato crop from climatic data. The effect of disease, induced by *Septoria lycopersici*, on tomato yield was estimated. In their review on the control of *Botrytis cinerea* in greenhouse tomato, Nicot and Baille[70] identified only a small number of models for greenhouse vegetables, one designed to forecast the fungus epidemics in cucumber under unheated greenhouse and another to simulate spore germination on tomato leaves. More generally, epidemiological models developed for field crop and often based on the occurrence of free water on the canopy, could be generic enough to be adapted to greenhouses. However, the development of mechanistic models on epidemiology or population dynamics in relation to the environment and crop status is a complex task. An example of a research model simulating the parasitoid-host relation between *Encarsia formosa* and the greenhouse whitefly on tomato crops was published by van Roermund *et al.*[71] Such a model can be used to evaluate release strategies under various climate conditions.

10.3 Areas of application

One of the ideas about model application most widely shared within the crop modeller's community (but maybe not so among scientists involved in control or management), is that crop or plant process models can readily be adapted to design control or management systems for decision support. Along with teaching, management and support are probably the two main areas of application for the models described in the previous sections of this chapter. Management can be defined as the sequence of three operations: planning, implementation and control. The planning operation sets up the strategy which encompasses the goals assigned to the cultivation and the means to achieve these goals. Implementation performs the translation from the strategy into actions, while control ensures the proper applications of these actions by constantly monitoring the process and revising the mode of application of the action.

Being a management task, crop management (and hence climate or fertigation control) requires identification and execution of suitable actions to obtain the desired crop behaviour and to reach the assigned goals. The decision process leading to the determination of the actions to be taken is a difficult task. It depends on:

- uncontrolled and uncontrollable external factors
- several complex interactions between the crop and its environment (both physical and economical)
- knowledge of the crop state.

While crop models can help to determine the crop state, models of the crop environment (not discussed so far) can also be used to clarify the state of the crop environment and to identify the interactions between the crop and its environment.

In view of these remarks, the first obvious application of crop models is as *information providers*, providing information that is otherwise not readily accessible to the grower, either because no measurement system is available or because the cost of obtaining the information would be prohibitive. Remarkably enough, most of the reported applications of models in the field of climate or fertigation control in greenhouses go one step further. As will be seen, models are used in optimization routines as representations of the crop processes (not of its states). This shift deserves a more detailed explanation. When using models as information providers, the interest is focused on the output, whatever the expression of the model may be. Accuracy of the results is paramount, especially when these results are *predictions*, that is, expected values of one or more crop states. Whether the model formulation represents the actual crop processes or not is irrelevant here. On the contrary, when models are used in optimization routines, and more specifically within the optimal control theory, the emphasis is on the formulation itself. For example, representing data with a parabola or a sine function may give numerical results which will differ little while the derivatives will be very different, and derivatives are used heavily in optimization procedures. Different expressions can lead to different control actions, which implies that the choice of the model to be used will not only depend on its accuracy in terms of prediction or simulation, but also on the actual equations that make the model. With respect to this use of models, they are *process representations*.

In the following subsections we present an overview of current works using models as information providers (teaching and crop management applications) and as process representations (climate and fertigation control).

10.3.1 Teaching

Crop models are interesting tools for teaching horticulture. They can fulfil different goals. They give an insight to the processes that together result in the crop behaviour. They provide a faster means than real experimentation to

demonstrate the effect of management on the crop. They can, of course, not replace the experience one gains from observing a real crop growing. In the SIMULSERRE project,[66] embedded in a user-friendly interface, the TOMGRO model was coupled to a greenhouse climate model and to a model of a simplified greenhouse climate controller. In the virtual experimentation, the user defines the climate control strategy (heating, ventilation, CO_2 set-points) and the truss pruning plan for the growing season. The results of these simulations are stored in extended details (from hourly values for all matter fluxes within the plant and between the crop and the environment to daily values of various state variables of the crop such as LAI, fruit load, harvest, etc.). Results can be viewed and compared with several types of plots (hourly, daily, cumulative values). For example, CO_2 enrichment policies can be compared and the differences in yield as well as in photosynthetic fluxes can be tracked. Demonstration of the role of key elementary processes is therefore straightforward and helps the student in obtaining an integrated view of the crop and management.

10.3.2 Yield prediction and crop management

The demand for yield prediction varies with the tomato cultivation system. In field production, determinate cultivars are selected to get fruits ripe for a single harvest. The expected time of harvest and expected amount of product are predicted to enable an integrated planning of production and processing. For example, Wolf *et al.*[72] estimated, the times of emergence, flowering, turning stage and harvesting of tomatoes for processing based on the heat sums. McNeal *et al.*[73] went a step further and predicted the mass of fruits at harvest using a greenhouse tomato crop model (TOMGRO) adapted to field conditions.

In greenhouse production, yield is planned for a long period of time. In Europe, there is strong competition between the various regions of production. In negotiations with the product buyers, growers must be able to announce their weekly production for the next couple of months. For this purpose, a simple tomato crop model named TOMPOUSSE was developed.[51] To be useful for practical operation, it had to respect a set of requirements:

- run with data commonly available in commercial conditions (plantation date, duration of the crop cycle, cultivar, weekly average values of radiation, temperature, CO_2 concentration and policy of fruit pruning and stem density);
- simulate the effect of the major cultivation techniques of climate control and crop management;
- deliver the weekly yield and average fruit weight.

Within this framework, the simplest formulations were retained in the model. The RUE approach was adopted to simulate the production of biomass. Every week, the dry matter allocated to fruits is partitioned among different age groups, according to the box-car-train technique.[74] The integration of fruit growth and development permits the estimation of the weekly number and dry weight of mature fruits. The dry weight is converted into fresh weight using a

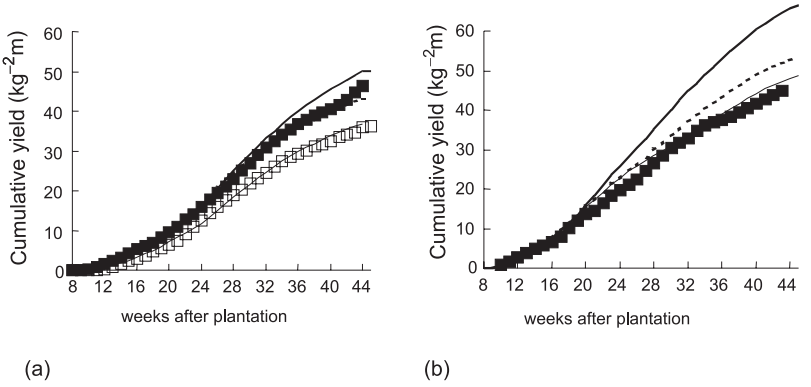


Fig. 10.1 Cumulated yield of greenhouse tomatoes in various locations and production systems. (a) Brittany (Brest, France), plantation in January 1994, observed yield without (❑) and with (■) CO_2 enrichment, simulated yield with continuous CO_2 enrichment (—), with CO_2 enrichment until week 23 only (– – – – –), without CO_2 enrichment (—). (b) South France (Alenya), plantation in January 1994, observed yield with CO_2 enrichment (■), simulated yield with continuous CO_2 enrichment (—), with CO_2 enrichment until week 15 only (– – – – –), with CO_2 enrichment until week 15 and roof whitening from week 25 (—). Note the differences between potential and actual yield in summer, due to the interaction between CO_2 enrichment and ventilation, and to roof whitening in Mediterranean areas.

dry matter content varying along the year. Such a simple model enabled the simulation of the time-course of production of greenhouse tomatoes in various regions and production systems (Fig. 10.1). It was also possible to estimate the effect on potential production of limiting operational factors (see previous sections).

In commercial conditions, yield forecast can be produced for the next weeks by running the model several times with weather inputs taken from a radiation database recorded in the region, and for the planned cultivation strategy (management of temperature, CO_2 concentration, truss pruning, stem density). Interestingly, due to the indeterminate development of greenhouse cultivars, the variation in the short-term predicted yield is low because the possible variations in global radiation affect only the last stages of fruit development (Fig. 10.2b). Another attempt to estimate the potential production of greenhouse tomato crops has been based on the calculation of the daily crop photosynthesis in response to the available radiation depending on season and latitude.[49] The difference between the potential and actual production in various parts of the world helps analysing the importance of different limiting factors.

Different parts of the TOMPOUSSE model reflect the different cultivation techniques. The efficiency of light interception depends on the leaf area index. Stem density and leaf pruning can affect the leaf area index. RUE increases with CO_2 concentration.[75] The dry weight allocation ratio to fruits responds with a hyperbolic curve to the fruit load per plant.[76] The appearance and ageing of organs is strongly affected by temperature.[57] The TOMPOUSSE model can be

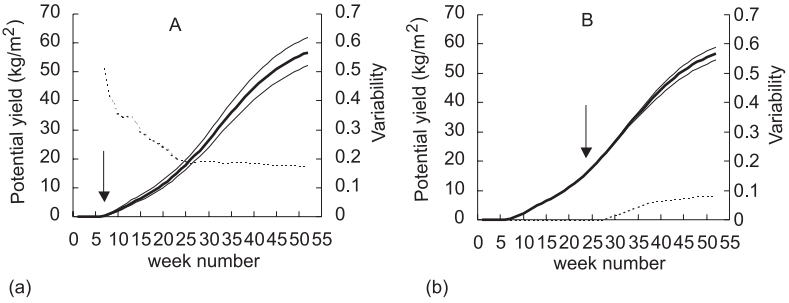


Fig. 10.2 Simulated potential yield of a greenhouse tomato crop based on the mean global radiation (measured in Avignon, France from 1969 to 1992) until week 1 (a) or 25 (b), and on the different 24 years (1969 to 1992) from week 2 (a) or 26 (b). Mean (—), maximum and minimum yields (—), relative variation ((max−min)/mean, – – – – –). The variability of yield prediction is higher when carried out from the vegetative (a) than from the generative (b) phase. It stabilises at a lower level in (b) than in (a) because, in this particular region, the weekly global radiation varies less during the second half of the year.

used as a simulator to evaluate different strategies of crop management. In the example presented in Fig. 10.3a to d, various policies of truss pruning are evaluated with respect to the fruit grade. De Koning[57] used a similar approach in a model of dry matter partitioning to optimize shoot density and number of fruits per plant.

These crop models, used to evaluate the biological consequences of policies of crop management, are still far from real decision support systems (DSS). For this purpose, the models should describe not only the dynamics of the crop and of its physical environment (greenhouse climate and/or soil), but also the decision-making process itself and its interactions with the biophysical system. For example, the GX/Sim system[77] is a greenhouse simulating platform that can specify the decision rules the grower uses to adapt the climate settings to the current climate conditions.

In the CONSERTO project,[78] a dynamic model of the greenhouse production system has been designed with three components: the decision system, the instructions-to-actions system and the biophysical system (Fig. 10.4). The decision system describes the management strategy applied over a cultivation period to realize production objectives. A management strategy consists of several conditional plans the realization of which is conditioned to the occurrence of specific events. In CONSERTO, it deals with climate and fertigation management, and manual operations such as fruit and leaf pruning, training and harvesting. A conditional plan comprises a nominal plan of instructions and a trajectory of desirable states and appropriate reactions that permits adjustments along the crop cycle. A nominal plan is a sequence of tasks assigned to a worker team.

The instructions-to-actions system converts these decisions into actions via automatons (the climate and fertigation control system) and workers. Because of

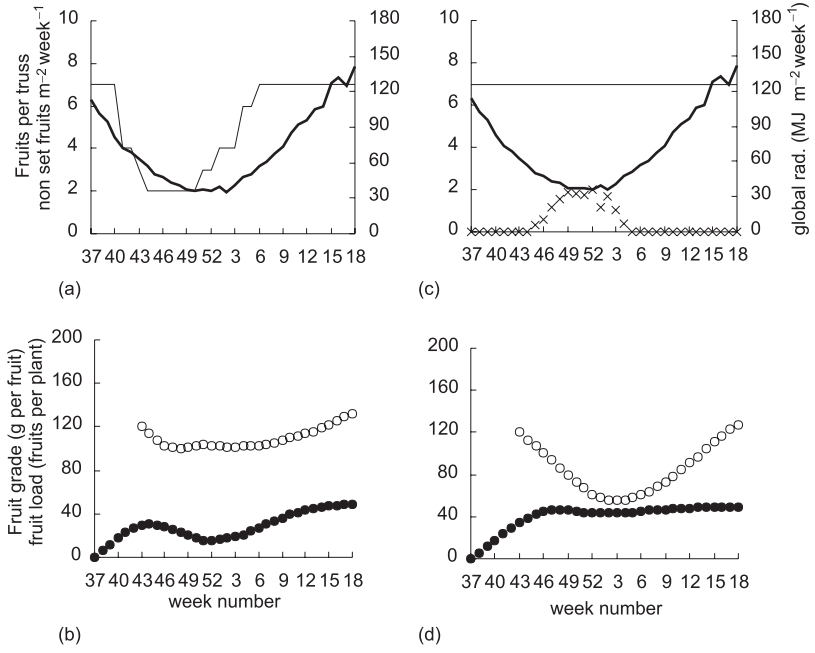


Fig. 10.3 Simulated effect of two strategies of truss pruning on the tomato fruit grade along a crop cycle. In this example, the objective is to keep the fruit grade constant after planting mid-September in the south of France, even though global radiation (—, mean of weekly values measured in Avignon, France from 1969 to 1992) goes down in winter then up in spring. (a), (b): when the number of fruits per truss (—) and consequently the fruit load (●) vary together with the radiation level, then the fruit grade (○) varies little. (c), (d): when the number of fruits and the fruit load are not regulated, some fruits do not set (x) during the shortest days but this natural regulation is not enough to control the fruit grade.

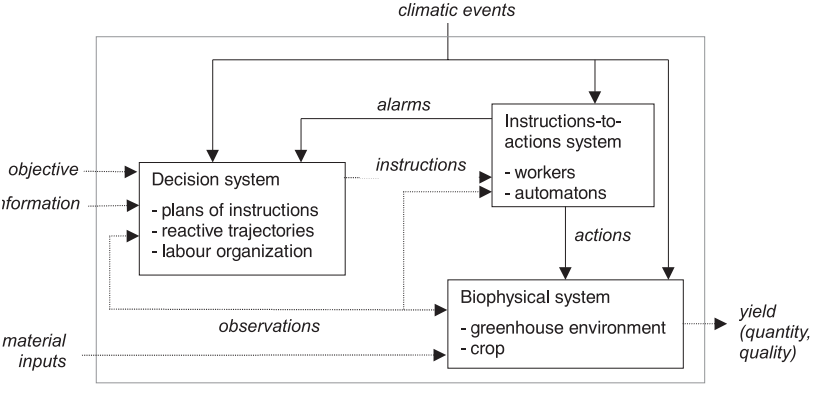


Fig. 10.4 The greenhouse production system as formalized in CONSERTO.[78] Dotted arrows represent the information each sub-system gets from the environment whereas full arrows represent control inputs.

limited resources (e.g. workforce or capacity of the heating system), what is actually realized may not fit with what was prescribed in the framework of the nominal plan, and alarms may be raised. Information on the actual use of resources is sent back to the decision system.

The biophysical system comprises a greenhouse climate and a tomato crop model. Actions and climatic events control it. The tomato crop model is a redesigned version of TOMGRO[16, 61] implemented in an object-oriented framework.[79] The outputs provide not only information on physical and biological performances of the system under a set of actions but also indicators (for example, the plant vigour or predictions of important events such as flowering or fruit maturity) useful for the decision system.

10.3.3 Climate control

Climate control is an operational management task which includes the activities of determining the crop and system states, choosing the set of goals to pursue within the more general set of goals defined by the strategic management of the crop and finally deciding upon which climate modifications are needed for the day at hand. Crop and greenhouse models can be useful in different ways for the grower to perform this task, either by giving an estimate of the crop state or by performing the decision task and offering a ready-to-use solution.

Models as information source

We are aware of very few model applications in this way for tomato, although a successful application on cotton is reported.[80] Attempts to provide information on diseases have also been reported for tomatoes, however, with only minor connections to the climate management problem.[81, 82] Finally the work of Harazono *et al.*[83] and Harazono[84] can be mentioned: a model for photosynthesis is used to provide information to a climate control system on this difficult to monitor flux. The task of the controller (based on rule inference) is to adapt environmental conditions (temperature, humidity and CO_2) to maintain an appropriate photosynthetic flux, as predicted by the model. None of these systems apply on-line parameter estimation, a technique widely used in computerized industrial process control. The goal of on-line parameter estimation is to assure that the model, through adequate parameterization, always describes the process as closely as possible.

Model-based climate control

Our understanding of model-based climate control encompasses all the approaches where new climate set-points are determined using either information output by the model or the knowledge contained in the model itself. Within this scope most approaches are designed for the ultimate goal to control directly the greenhouse climate. However, Schotman[85] argues that some of the drawbacks of such a conception of climate control are that:

- the grower has no control over the objectives assigned to the control generating routine;
- information needed by the model-based control system often has no agronomic or relevant meaning to the grower.

Optimal control is probably the most widely used method to exploit available models and determine 'optimal' crop environmental conditions. The optimal control theory, either based on Pontryagin's minimum principle[86] or on Bellman's principle of optimality[87] develops as follows. Noting that scalars are denoted with italics and vectors with bold, let $\mathrm{d}\boldsymbol{x}/\mathrm{d}t = f(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{p}(t))$ be a dynamical model of the plant (*plant* is taken here in its industrial meaning of factory, the factory being in our case the *crop*!), where dynamical model means differential equations, $\boldsymbol{x}(t)$ is the state vector of the plant, $\boldsymbol{u}(t)$ is the control vector (quantities that can be manipulated to modify the plant behaviour), $\boldsymbol{p}(t)$ is the perturbation (or uncontrolled) vector and t denotes time. Optimizing the plant behaviour implies that an objective function $J()$ has been defined which measures how well the system performs. $J()$ is classically the sum of two terms. The first, $\Phi(\boldsymbol{x}(t_f), t_f)$, denotes requirements on the end-point of the control horizon (expressed through a weighting function) which allows for optimization in time (minimum time problems where t_f is let free and must be as low as possible), or for the specification of desired final state, $\boldsymbol{x}(t_f)$. For example, a combined problem could be to produce lettuces of a given fresh weight in the shortest time possible so as to maximize the number of crop batches during the season. The second term in $J()$ is an integral which accounts for all running costs or gains relevant to the plant processes and is often written as $\int L(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{p}(t))\mathrm{d}t$. In the previous example, $L()$ would cumulate all the costs of growing the lettuces related to the controllable inputs $\boldsymbol{u}(t)$. For greenhouse tomatoes, where no final state is required because tomatoes are produced continuously, the function $L()$ would accumulate both the costs of controls $\boldsymbol{u}(t)$ and the gains obtained from the crop represented by its states $\boldsymbol{x}(t)$. The optimal control theory provides means to solve the problem of optimising $J()$ with respect to $\boldsymbol{u}$, under the constraint represented by $f()$. For a more thorough presentation of the optimal control theory, one can refer to Pontryagin *et al.*,[86] Bellman and Dreyfus[87] and Lewis.[88] Climate control application of crop models or crop processes models within the framework of optimal control also requires a model of the greenhouse climate because the control variables directly modify the climate. The plant behaviour is driven indirectly through its responses to the modifications of the environment.

In one of its simplest forms, the climate optimization problem is defined as: using a crop dry matter accumulation model and an algebraic expression of the greenhouse climate model, find the day- and night-time temperatures that maximize a cost function balancing the relative growth rate and the heating costs. CO_2 enrichment can also be included. Gal *et al.*,[89] Seginer,[90,91] Seginer *et al.*[92] and Critten[93] showed that under the hypothesis that the dry matter evolution can be written as $\mathrm{d}w/\mathrm{d}t = s(w) \times f(\boldsymbol{u})$ the optimal solution can be expressed as a direct function of the external climate conditions for each time-

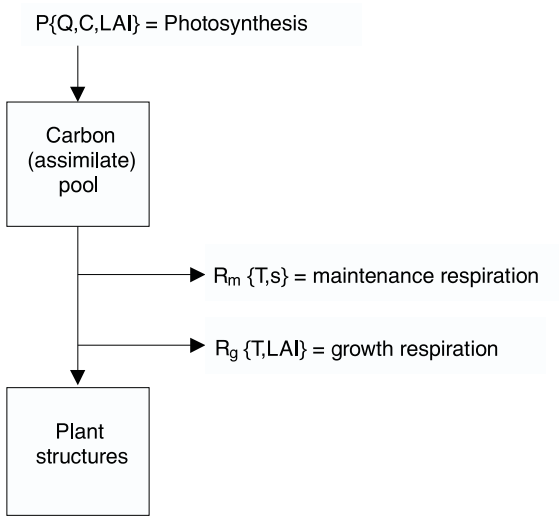


Fig. 10.5 Dynamic model of the crop carbon balance. The crop is divided into two compartments, the 'C-pool' or temporary photosynthetic assimilates and the 'structures' or the crop itself. In this model, $R_m\{T,s\}$ is a compulsory flux driven by the temperature which will take its necessary carbon in the structural compartment when the pool is empty and P{Q,C,LAI} is too low. If possible (remaining carbon in the pool, P{} high enough), $R_g\{T,LAI\}$ and growth occur. Growth is proportional to $R_g\{\}$. Q: photosynthetic photon flux density; C: air CO_2 concentration; LAI: crop leaf area index; T: plant (air) temperature; s: weight of structural dry matter.

instant independently. In practice this allows for the off-line computation of lookup tables that indicate what actions should be taken at current conditions. However, no real experiments have put these results to the test. Seginer *et al.*[25] have studied the temperature optimization problem, only based on plant need. They used a dynamic model of the carbon balance of the crop (Fig. 10.5) with a temporary carbohydrate pool to derive the day and night temperatures that maximize the relative growth rate for a given daily radiative flux. The results are that young crops need higher temperatures than old ones where maintenance respiration is more important and that for a given situation, several couples of day and night temperatures are optimal (Fig. 10.6). Tchamitchian *et al.*[94] and Tap *et al.*[95] have used a dynamical greenhouse model instead of an algebraic one to introduce the damping of temperature due to the structures in the greenhouse. Solving the climate problem, either for tomato or for lettuce, respectively, proved to be a rather difficult numerical problem.

More complex crop models have also been used within the scope of the optimal control theory to determine optimal daily climate set-points under the constraints of long-term crop production optimization. The first attempt to mention is the work of Marsh and Albright[96, 97] who tried to determine by simulation the optimal temperature set-points for lettuce production using a crop growth model. Seginer and McClendon[98] addressed the same problem but using

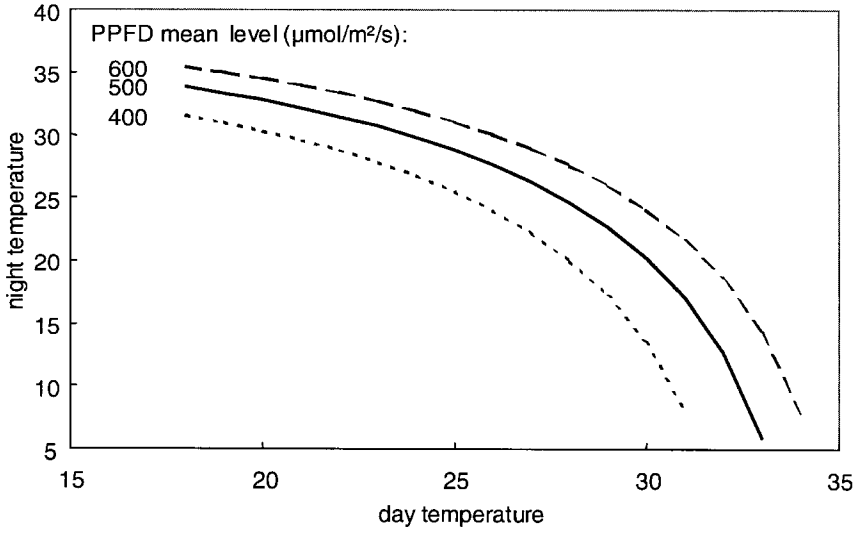


Fig. 10.6 Optimal day and night temperatures for a tomato crop, as determined from the tomato carbon pool model (see Fig. 10.5). The simulation is for a crop of LAI = 3.5, submitted to constant day and night temperatures and to a constant level of photosynthetic photon flux density (PPFD) during the day. Daylight period is 14 hours. The hypothesis for this simulation is that the C-pool level must be the same at $t = 0$ and $t = 24$ hours.

Pontryagin's minimum principle to solve the very same problem and found similar results. Later, Seginer and Sher[99] used the TOMGRO model to solve the problem for tomatoes. It should be noticed that these approaches hardly used a greenhouse model at all: the goal is to define optimal set-points trajectory along the crop cycle, namely to optimize the blue-prints already available for these crops. Coupling a dynamical model of the greenhouse climate to a lettuce growth model, van Henten[100] used the singular perturbation approach[101] to tackle the problem of models with different magnitudes of time constants. A new development in this area (Tap, pers. comm.) applies the same method to a simplified tomato crop model. Daily optimization of the climate (so-called *fast* processes) under the constraint of long-term optimisation of the crop production (so-called *slow* processes) can then be solved.

The lettuce case, being simpler to study than the tomato case, has often been used as a preliminary case study before addressing the more complex case of tomato where growth, development and harvest occur at the same time. One important remark to be made is that although many theoretical applications of models to climate control have been studied, none or very few have been put to the test in practice. The reasons might be first because control engineering groups lacking the greenhouse facilities have done these studies and second because optimal control produces time-varying set-points which cannot be implemented on commercial greenhouse climate computers.

10.3.4 Fertigation control

In both field and greenhouse production, there is an increasing pressure to improve the policies of irrigation and fertilization, that should both satisfy the plant demand for water and nutrients and avoid losses of nutrients in the environment. At present, empirical methods are used; they should be improved with mechanistic models in development.

The supply of water to the crop must fit its water requirements. In soilless culture, irrigation is usually calculated based on radiation measurements. Several relations have been established between the crop water uptake and the incident radiation for tomato as well as for other vegetable crops (formulae reviewed by Jolliet[28]). The VPD should also be taken into account when radiation and VPD are uncoupled, for example in changing climatic conditions and when using systems of climate control such as thermal screens or fog systems.[35] The water demand also depends on a crop coefficient that increases with the leaf area development. In soil culture, the availability of water in the soil compartment must also be considered. It depends on the hydraulic properties of the soil and on the root development. In the field, the rain flux must enter in the water balance.

In greenhouses, computers are used to monitor radiation and to control the quantity of water that is provided to open systems (on soil or soilless), that is, the calculated evapotranspiration plus about 25% run-off to avoid salt concentration in the root substrate. In closed soilless systems, the water input must fit the crop demand to maintain the total volume of circulating nutrient solution. In the field, new decision support systems are designed to calculate the proper water supply. For example, the IRRIGERE software, designed for field tomato, estimates the daily evapotranspiration from climate and crop development and the soil water reserve from the soil characteristics and the root depth.[102] Irrigation will not meet crop demand when water stress is needed to increase the dry matter content of fruits. In that case, the objective is to exhaust the water available in the root zone at fruit harvest. With these constraints, irrigation is proposed when the watering dose gets higher than a threshold value of 3 mm.

Few attempts have been made to build fertilization strategies using models of crop requirement, even in soilless culture. In this cultivation system, nutrients are usually supplied in excess together with water. Therefore there is no way to control the crop growth or product quality through the regulation of fertigation. Recently, Marcelis *et al.*[103] proposed the combination of models and sensors to optimize the nutrient supply in closed systems.

10.4 Discussion of the methods and future trends

In the last decades, most of the modelling effort on tomato crop has been put on the carbon fluxes and development processes in relation to the crop environment. To a lesser extent, the plant–water relationship has received some attention. It should be noted that the results of these studies for the grower are far

from proportional to the invested time: practical applications to irrigation are more numerous and more successful than yield prediction or automated climate control systems.

Fruit quality, nitrogen runoff, and sustainable cropping techniques are nowadays the focus of tomato crop modelling, using the same physiological approaches that have been successful up to now. The expected results are explanatory models that could help to estimate crop nutrient demand or harvest quality. Although this trend will widen the number of processes that will be represented, model-based plant management will still suffer from the limited scope of available models, and from the time needed to obtain these models. For example, fruit quality is defined by several criteria (see Table 10.1), but only some of them are addressed in modelling studies. Other processes are hardly addressed, such as plant architecture which determines its ability to intercept light and which is also part of the grower's perception of the state of the plant. Models on the effect of diseases or pest attacks are also largely unavailable, not to mention models predicting when and where the probability of physiological or pathological disorders is the highest.

To overcome these drawbacks, other modelling approaches can be successful. The SERRISTE project[104, 105, 106] has opted to use artificial intelligence techniques to represent both the knowledge involved in the daily climate management task and in the crop's response. Agronomical know-how, obtained from experts, is represented through a set of variables, which are constrained within a fuzzy domain and through a set of constraints relating these variables. For example, the target daily mean temperature domain is obtained by:

- computing an optimal temperature from the forecasted available radiation;
- making adjustments for the variety;
- positioning a 1ºC window around this value according to the vigour status of the crop.

A constraint is expressed as a linear combination of variables, the result of being forced to belong to a fuzzy domain. As an example, the temperature difference between day and night (a linear combination) must belong to a domain extending from 2 to 5C, values which may be changed depending on the current conditions (for example, switch from 5 to 3 if *Botrytis* has been observed). A constraint satisfaction algorithm determines the sets of variable values which satisfy all the constraints. Declarative knowledge and numerical models are mixed in what is called a *knowledge base* (see Chapter 5). Two years of experiments in extension services facilities in three different regions of France have proven the feasibility and the real agronomic success of this approach.

As proven by Guerrin *et al.*,[107] the combination of declarative and numerical models broadens the scope of the system that can be represented and thus may be a way to overcome the limitations of numerical models. Moreover, building a declarative model may on many occasions be faster and cheaper than the experimental and theoretical work that would be needed to obtain a numerical model of the same processes. However, design of hybrid models mixing

declarative and numerical knowledge and use of Artificial Intelligence techniques for crop management support is still limited.

10.5 References

1. Food and Agriculture Organisation of the United Nations, 2000. Web site: www.fao.org
2. World Processing Tomato Council, 2000. Web site: www.wptc.to
3. BAILLE A, 'The energy cycle', In: Z Enoch and G Stanhill (eds), *Greenhouse Ecosystems*, Amsterdam, Elsevier, 1999, 265–86.
4. BAKKER J C, BOT G P A, CHALLA H, VAN DE BRAAK N J, *Greenhouse Climate Control, an Integrated Approach*, Wageningen, Wageningen Pers, 1995.
5. GARY C, BAILLE A, 'The carbon cycle', In: Z Enoch and G Stanhill (eds), *Greenhouse Ecosystems*. Amsterdam, Elsevier, 1999, 287–301.
6. GARY C, JONES J W, TCHAMITCHIAN M, 'Crop modelling in horticulture: state of the art', *Scientia Hortic.*, 1998 **74** 3–20.
7. PASSIOURA J B, 'Simulation models: science, snake oil, education, or engineering?' *Agron J*, 1996 **88** 690–4.
8. LONGUENESSE J J, GARY C, TCHAMITCHIAN M, 'Modelling CO_2 exchanges of greenhouse crops: a matter of scales and boundaries', *Acta Hortic*, 1993 **328** 33–47.
9. GIJZEN H, 'CO_2 uptake by the crop'. In: J C Bakker, G P A Bot, H Challa and N J Van de Braak (eds), *Greenhouse Climate Control, an Integrated Approach*, Wageningen, Wageningen Pers, 1995, 16–35.
10. ACOCK B, CHARLES-EDWARDS D A, FITTER D J, HAND D W, LUDWIG L J, WARREN WILSON J, WITHERS A C, 'The contribution of leaves from different levels within a tomato crop to canopy photosynthesis: an experimental examination of two canopy models' *J Exp Bot*, 1978 **29** 815–27.
11. THORNLEY J H M, *Mathematical Models in Plant Physiology*, London, Academic Press, 1976.
12. SPITTERS C J T, 'Separating the direct and diffuse component of global radiation and its implication for modelling canopy photosynthesis. II Calculation of canopy photosynthesis'. *Agric For Meteorol*, 1986 **38** 231–42.
13. BERTIN N, HEUVELINK E, 'Dry-matter production in a tomato crop: comparison of two simulation models', *J Hortic Sci*, 1993 **68** 995–1011.
14. GIJZEN H, 'Interaction between CO_2 uptake and water loss', In: J C Bakker, G P A Bot, H Challa and N J Van de Braak (eds), *Greenhouse Climate Control, an Integrated Approach*, Wageningen, Wageningen Pers, 1995, 51–62.
15. CRITTEN D L, 'A review of the light transmission into greenhouse crops', *Acta Hortic*, 1993 **328** 9–32.
16. JONES J W, DAYAN E, ALLEN L H, VAN KEULEN H, CHALLA H, 'A dynamic tomato growth and yield model (TOMGRO)', *Trans. ASAE*, 1991 **34** 663–72.

17. TCHAMITCHIAN M, LONGUENESSE J J, 'Photosynthèse d'une culture en rangs de tomates sous serre. Modélisation analytique et cartographie de l'activité du feuillage', *Agronomie*, 1991 **11** 17–26.
18. HEUVELINK E, 'Evaluation of a dynamic simulation model for tomato crop growth and development', *Ann Bot*, 1999 **83** 413–22.
19. GARY C, LE BOT J, FROSSARD J S, ANDRIOLO J L, 'Ontogenic changes in the construction cost of leaves stems fruits and roots of tomato plants', *J Exp Bot*, 1998 **49** 59–68.
20. CANNEL M G R, THORNLEY J H M, 'Modelling the components of plant respiration: some guiding principles', *Ann Bot*, 2000 **85** 45–54.
21. DEWAR R C, 'A model of the coupling between respiration, active processes and passive transport', *Ann Bot*, 2000 **86** 279–86.
22. GENT M P N, ENOCH H Z, 'Temperature dependence of vegetative growth and dark respiration: a mathematical model', *Plant Physiol*, 1983 **71** 562–7.
23. GARY C, 'A simple carbon balance model simulating the short-term responses of young vegetative tomato plants to light, CO_2 and temperature', *Acta Hortic*, 1988 **229** 245–50.
24. GARY C, 'Interest of a carbon balance model for on-line growth control: the example of a daylight dependent night temperature control', *Acta Hortic*, 1989 **248** 265–8.
25. SEGINER I, GARY C, TCHAMITCHIAN M, 'Optimal temperature regimes for a greenhouse crop with a carbohydrate pool: a model study', *Scientia Hortic* 1994 **60** 55–80.
26. JONES H G, TARDIEU F, 'Modelling water relations of horticultural crops: a review', *Scientia Hortic*, 1998 **74** 21–46.
27. VAN DE SANDEN P A C M, 'Water balance'. In: J C Bakker, G P A Bot, H Challa and N J Van de Braak (eds), *Greenhouse Climate Control, an Integrated Approach*, Wageningen, Wageningen Pers, 1995, 35–50.
28. JOLLIET O, 'The water cycle', *In*: Z Enoch and G Stanhill (eds), *Greenhouse Ecosystems*, Amsterdam, Elsevier, 1999, 303–26.
29. VAN İEPEREN W, Consequences of diurnal variation in salinity on water relations and yield of tomato, Thesis, Wageningen Agricultural University, 1996.
30. GUICHARD C, GARY C, LONGUENESSE J J, LEONARDI C, 'Water fluxes and growth of greenhouse tomato fruits under summer conditions', *Acta Hortic*, 1999 **507** 223–30.
31. FISHMAN S, GÉNARD M, 'A biophysical model of fruit growth: simulation of seasonal and diurnal dynamics of mass', *Plant Cell Environ*, 1998 **21** 739–52.
32. LEONARDI C, BAILLE A, GUICHARD S, 'Predicting transpiration of shaded and non-shaded tomato fruits under greenhouse environments', *Scientia Hortic*, 2000 **84** 297–307.
33. STANGHELLINI C, 'Vapour balance', In: J C Bakker, G P A Bot, H Challa and N J Van de Braak (eds), *Greenhouse Climate Control, an Integrated Approach*, Wageningen, Wageningen Pers, 1995, 141–50.

34. BOULARD T, BAILLE A, MERMIER M, VILLETTE F, 'Mesures et modélisation de la résistance stomatique foliaire et de la transpiration d'un couvert de tomates de serre', *Agronomie*, 1991 **11** 259–74.
35. BOULARD T, JEMAA R, 'Greenhouse tomato crop transpiration model. Application to irrigation control', *Acta Hortic*, 1993 **335** 381–7.
36. WARREN WILSON J, HAND D W, HANNAH M A, 'Light interception and photosynthetic efficiency in some glasshouse crops' *J Exp Bot* 1992 **43** 363–73.
37. PAPADOPOULOS A P, PARARAJASINGHAM S, 'The influence of plant spacing on light interception and use in greenhouse tomato (*Lycopersicon esculentum* Mill.): a review', *Scientia Hortic*, 1997 **69** 1–29.
38. GIJZEN H, GOUDRIAAN J, 'A flexible and explanatory model of light distribution and photosynthesis in row crops', *Agric. For. Meteorol*, 1989 **48** 1–20.
39. AIKMAN D P, 'Potential increase in photosynthetic efficiency from the redistribution of solar radiation in a crop', *J. Exp. Bot.*, 1989 **40** 855–64.
40. LE BOT J, ADAMOWICZ S, ROBIN P, 'Modelling plant nutrition of horticultural crops: a review', *Scientia Hortic*, 1998 **74** 47–82.
41. CARDENAS-NAVARRO R, ADAMOWICZ S, GOJON A, ROBIN P, 'Modelling nitrate influx in young tomato (*Lycopersicon esculentum* Mill.) plants', *J Exp Bot*, 1999 **50** 625–35.
42. GREENWOOD D J, LEMAIRE G, GOSSE G, CRUZ P, DRAYCOTT A, NEETESON J J, 'Decline in percentage N in C_3 and C_4 crops with increasing plant mass', *Ann Bot*, 1990 **66** 425–36.
43 LE BOT J, ANDRIOLO J L, GARY C, ADAMOWICZ S, ROBIN P, 'Dynamics of N accumulation and growth of tomato plants in hydroponics: an analysis of vegetative and fruit compartments', In: G Lemaire and I G Burns (eds), *Diagnostic Procedures for Crop N Management*, Paris, INRA Editions, 1995, 37–51.
44. CALOIN M, YU O, 'Analysis of the time course of change in nitrogen content in *Dactylis glomerata* L', *Ann Bot*, 1984 **54** 69–76.
45. BELLERT C, LE BOT J, DORAIS M, LOPEZ J, GOSSELIN A, 'A nitrogen accumulation and growth of fruiting tomato plants in hydroponics', *Acta Hortic*, 1998 **458** 293–301.
46. CAVERO J, PLANT R E, SHENNAN C, WILLIAMS J R, KINIRY J R, BENSON V W, 'Application of EPIC model to nitrogen cycling in irrigated processing tomatoes under different management systems', *Agric Syst*, 1998 **56** 391–414.
47. MANKIN K R, FYNN R P, 'Modeling individual nutrient uptake by plants: relating demand to microclimate', *Agric Syst*, 1996 **50** 101–14.
48. BAR-TAL A, KEINAN M, FISHMAN S, ALONI B, OSEROVITZ Y, GÉNARD M, 'Simulation of environmental effects on Ca content in pepper fruit', *Acta Hortic*, 1999 **507** 253–62.
49. CHALLA H, BAKKER M, 'Potential production within the greenhouse environment' In: Z Enoch and G Stanhill (eds), *Greenhouse Ecosystems*. Amsterdam, Elsevier, 1999, 333–48.

50. KANO A, VAN BAVEL C H M, 'Design and test of a simulation model of tomato growth and yield in a greenhouse', *J Jpn Soc Hortic Sci*, 1988 **58** 406–16.
51. GARY C, BAILLE A, NAVARETTE M, ESPANET R, 'TOMPOUSSE, un modèle simplifié de prévision du rendement et du calibre de la tomate', In: A Baille (ed.), *Actes du Séminaire de l'AIP intersectorielle 'Serres'*, Avignon, INRA, 1997, 100–9.
52. DE HALLEUX D, GAUTHIER L, 'Adaptation d'un modèle de prévision de récolte de tomates en serre (TOMPOUSSE) à une culture commerciale québécoise', *Csae/Scgr (Canadian Society for Engineering in Agricultural, Food, and Biological Systems) Paper* No. 97–602, 1997, 1–8.
53. ANDRIOLO J L, DUARTE T DA S, LUDKE L, SKREBSKY E C, 'Avaliação da produtividade do tomateiro em cultivo protegido através de um modelo de simulação da produção', *Hortic Bras*, 1998 **16** 13–18.
54. BRISSON N, MARY B, RIPOCHE D, JEUFFROY M H, RUGET F, NICOULLAUD B, GATE P, DEVIENNE-BARRET F, ANTONIOLETTI R, DURR C, RICHARD G, BEAUDOIN N, RECOUS S, TAYOT X, PLENET D, CELLIER P, MACHET J M, MEYNARD J M, DELÉCOLLE R, 'STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterisation applied to wheat and corn', *Agronomie*, 1998 **18** 311–46.
55. CHALLA H, HEUVELINK E, VAN MEETEREN U, 'Crop growth and development', In: J C Bakker, G P A Bot, H Challa and N J Van de Braak (eds) *Greenhouse Climate Control, an Integrated Approach*, Wageningen, Wageningen Pers, 1995, 62–84.
56. FENG Y, LI X, BOERSMA L, 'The Arrhenius equation as a model for explaining plant responses to temperature and water stresses', *Ann Bot*, 1990 **66** 237–44.
57. DE KONING, Development and dry matter distribution in glasshouse tomato: a quantitative approach, Thesis, Wageningen Agricultural University, 1994.
58. JOHNSON I R, THORNLEY J H M, 'Temperature dependence of plant and crop processes', *Ann Bot*, 1985 **55** 1–24.
59. TIJSKENS L M N, VERDENIUS F, 'Summing up dynamics: modelling biological processes in variable temperature scenarios', *Agric Syst*, 2000 **66** 1–15.
60. BERTIN N, 'Competition for assimilates and fruit position affect fruit set in indeterminate greenhouse tomato', *Ann Bot*, 1995 **75** 55–65.
61. GARY C, BARCZI J, BERTIN N, TCHAMITCHIAN M, 'Simulation of individual organ growth and development on a tomato plant: a model and a user-friendly interface', *Acta Hortic*, 1995 **399** 199–205.
62. MARCELIS L F M, 'Sink strength as a determinant of dry matter partitioning in the whole plant', *J Exp Bot*, 1996 **47** 1281–91.
63. HEUVELINK M, 'Dry matter partitioning in a tomato plant: one common assimilate pool?' *J Exp Bot* 1995 **46** 1025–33.
64. HEUVELINK M, BERTIN N, 'Dry matter partitioning in a tomato crop: comparison of two simulation models', *J Hortic Sci*, 1994 **69** 885–903.

65. SCHOLBERG J, Adaptative use of crop growth models to simulate the growth of field-grown tomato, Thesis, University of Florida, Gainesville, 1996.
66. GARY C, CHARASSE L, TCHAMITCHIAN M, BERTIN N, REBILLARD A, BOULARD T, CARDI J P, BAILLE A, 'SIMULSERRE: an educational software simulating the greenhouse-crop system', *Acta Hortic*, 1998 **456** 451–8.
67. BERTIN N, GUICHARD S, LEONARDI C, LONGUENESSE J J, LANGLOIS D, NAVEZ B, 'Seasonal evolution of the quality of fresh glasshouse tomatoes under Mediterranean conditions, as affected by air vapour pressure deficit and plant fruit load', *Ann Bot*, 2000 **85** 741–50.
68. HO L C, BELDA R, BROWN M, ANDREWS J, ADAMS P, 'Uptake and transport of calcium and the possible causes of blossom-end rot in tomato', *J Exp Bot*, 1993 **44** 509–18.
69. SEGHI L, ORLANDINI S, GOZZINI B, *Survey of Simulation Models of Plant Disease*. Firenze, Italy, Centre for Computer Science Application in Agriculture (CeSIA), 1996.
70. NICOT P, BAILLE A, 'Integrated control of *Botryris cinerea* on greenhouse tomatoes', In: C E Morris, P C Nicot and C Nguyen-Thé (eds), *Aerial Plant Surface Microbiology*, New York, Plenum Press, 1996, 169–89.
71. VAN ROERMUND H J W, VAN LENTEREN J C, RABBINGE R, 'Biological control of greenhouse whitefly with the parasitoid Encarsia Formosa on tomato: an individual-based simulation approach', *Biol Control*, 1997 **9** 25–47.
72. WOLF S, RUDICH J, MARANI, REKAH Y, 'Predicting harvesting date of processing tomatoes by a simulation model', *J Am Soc Hortic Sci*, 1986 **111** 11–16.
73. MCNEAL B L, SCHOLBERG J M S, JONES J W, STANLEY C D, CSIZINSZKY A A, OBREZA T A, 'Application of a greenhouse tomato-growth model (TOMGRO) to field-grown tomato', *Soil Crop Sci Soc Fla Proc*, 1995 **54** 86–93.
74. GOUDRIAAN J, 'Box-car-train methods for modelling of ageing, development, delays and dispersion', In: J A J Metz and O Diekman (eds) *The Dynamics of Physiologically Structured Populations*, Berlin, Springer-Verlag, 1986, 453–73.
75. NEDERHOFF E, Effects of CO_2 concentration on photosynthesis, transpiration and production of greenhouse fruit vegetable crops, Thesis, Wageningen Agricultural University, 1994.
76. HEUVELINK E, 'Effect of fruit load on dry matter partitioning in tomato', *Scientia Hortic*, 1997 **69** 51–9.
77. GAUTHIER L, 'GX: a Smalltalk-based platform for greenhouse environment control. 1. Modeling and managing the physical system. 2. Supporting and implementing control strategies', *Trans ASAE* 1992 **35** 2003–20.
78. RELLIER J P, MARTIN-CLOUAIRE R, NAVARRETE M, JEANNEQUIN B, GARY C, MONTBROUSSOUS B, TCHAMITCHIAN M, BAILLE A, 'Modeling and simulating decision making for greenhouse tomato production: the Conserto project', *Acta Hortic*, 1998 **456** 485–92.

79. GAUTHIER L, GARY C, ZEKKI H, 'GPSF: a generic and object-oriented framework for crop simulation', *Ecol Model*, 1999 **116** 253–68.
80. MCKINION J M, BAKER D N, WHISLER F D, LAMBERT J R, 'Application of the GOSSYM/COMAX system to cotton crop management', *Agric Syst*, 1989 **31** 55–65.
81. KOZAI T, 'Ideas of greenhouse climate control based on knowledge engineering techniques', *Acta Hortic*, 1985 **174** 365–73.
82. GUAY R, GAUTHIER L, 'Knowledge representation in a tomato disorder diagnosis system', *Comput Electron Agric*, 1991 **6** 21–32.
83. HARAZONO Y, KAMIYA H, YABUKI K, 'A control method based on artificial intelligence technique and its application for controlling plant environment', *Acta Hortic*, 1988 **230** 209–14.
84. HARAZONO Y, 'Artificial intelligence technology to control plant environments', In: Y P S Bajaj (ed.) *High-tech and Micro-propagation*, vol. 1, Berlin, Springer-Verlag, 1991, 432–51.
85. SCHOTMAN P J, Improving support for greenhouse climate management. Thesis, Wageningen Agricultural University, 2000.
86. PONTRYAGIN L S, BOLTYANSKII V G, GAMKRELIDZE R V, MISHCHENKO E F, *The Mathematical Theory of Optimal Processes*, New York, John Wiley Interscience, 1962.
87. BELLMAN R E, DREYFUS S E, *Applied Dynamic Programming*, Princeton, Princeton University Press, 1962.
88. LEWIS F L, *Optimal Control*, New York, John Wiley Interscience, 1986.
89. GAL S, ANGEL A, SEGINER I, 'Optimal control of greenhouse climate: methodology', *Eur J Oper Res*, 1984 **17** 45–56.
90. SEGINER I, 'Optimising greenhouse operation for best aerial environment', *Acta Hortic*, 1980 **106** 169–78.
91. SEGINER I, 'Optimal greenhouse production under economic constraints', *Agric Syst* 1989 **29** 67–80.
92. SEGINER I, ANGEL A, GAL S, KANTZ D, 'Optimal CO_2 enrichment strategy for greenhouses: a simulation study', *J Agric Engng Res*, 1986 **34** 285–304.
93. CRITTEN D L, 'Optimisation of CO_2 concentration in greenhouse crops: a modelling analysis for the lettuce crop', *J Agric Engng Res*, 1991 **48** 261–71.
94. TCHAMITCHIAN M, VAN WILLIGENBURG L G, VAN STRATEN G, 'Optimal control applied to tomato crop production in a greenhouse', *ECC'93, European Control Conference*, Groningen, 1993.
95. TAP R F, VAN WILLIGENBURG L G, VAN STRATEN G, VAN HENTEN E, 'Optimal control of greenhouse climate: computation of the influence of fast and slow dynamics', *IFAC Conference*, Sydney, IFAC, 1993.
96. MARSH L S, ALBRIGHT L D, 'Economically optimum temperatures for greenhouse hydroponic lettuce production. Part I: a computer model', *Trans ASAE*, 1991 **34** 550–6.
97. MARSH L S, ALBRIGHT L D, 'Economically optimum temperatures for

greenhouse hydroponic lettuce production. Part II: results and simulations', *Trans ASAE*, 1991 **34** 557–62.

98. SEGINER I, MCCLENDON R W, 'Methods for optimal control of the greenhouse environment', *Trans ASAE*, 1992 **35** 1299–1307.
99. SEGINER I, SHER A, 'Optimal greenhouse temperature trajectories for a multi-state-variable tomato model', In: *The Computerized Greenhouse*, New York, Academic Press, 1993.
100. VAN HENTEN E J, Greenhouse climate control: an optimal control approach. Thesis, Wageningen Agricultural University, 1994.
101. KOKOTOVIC P V, KHALIL H K, O'REILLY J, *Singular Perturbation Methods in Control: Analysis and Design*, New York, Academic Press, 1986.
102. BATTILANI A, BUSSIÈRES P, COMBRES J P, ANCONCELLI S, DUMAS Y, 'Irrigere: a computer program to facilitate processing tomato irrigation management', *Acta Hortic*, 1999 **487** 537–46.
103. MARCELIS L F M, VAN DEN BOOGAARD R, MEINEN E, 'Control of crop growth and nutrient supply by the combined use of crop models and plant sensors', *Proceedings of IFAC Agricontrol 2000, International conference on Modelling and control in agriculture, horticulture and post-harvested processing*, 10–12 July 2000, Wageningen, 2000, 351–6.
104. MARTIN-CLOUAIRE R, KOVÁTS K, CROS M J, 'Determination of greenhouse climate setpoints by SERRISTE: the approach and its object-oriented implementation', *AI Applications*, 1993 **7** 1–15.
105. JEANNEQUIN B, MARTIN-CLOUAIRE R, TCHAMITCHIAN M, 'Planning daily climate set-points for a greenhouse tomato production: the SERRISTE system', In: H F M ten Berge and A Stein (eds) *Model-based Decision Support in Agriculture, Quantitative Approaches in Systems Analysis*, 1997 **15** 9–14.
106. TCHAMITCHIAN M, HENRY-MONTBROUSSOUS B, JEANNEQUIN B, LAGIER J, 'SERRISTE: climate set-point determination for greenhouse tomatoes', *Acta Hortic*, 1998 **456** 321–7.
107. GUERRIN F, 'Magma: a model to help animal manure management at the farm level', *Proceedings of IFAC Agricontrol 2000, International conference on Modelling and control in agriculture, horticulture and post-harvested processing*, 10–12 July 2000, Wageningen, 159–64.

11

Dairy production

C. F. E. Topp, Scottish Agricultural College, Ayr

11.1 Introduction

The political and economic environment in which dairy farmers operate have changed significantly over the past two decades. From a time where policy measures encouraged expansion in production and hence supported capital investment, circumstances have changed to a situation where policy seeks to control production by, for example, the introduction in 1984 of constraints on the volumes of milk produced. More recently, public concerns about the impact of agriculture on the environment has raised issues relating to water quality and other forms of environmental degradation. Increasingly, European agricultural support measures include cross-compliance elements. Accordingly, agricultural policy increasingly demands farmers to observe environmental criteria to benefit from support payments. Nevertheless, in the long-term farmers will only survive if they can guarantee sufficient financial surplus to allow for reinvestment, personal living expenses of those working on the farm and some financial returns for risks taken. Consequently, dairy farmers face the dual challenge of establishing a financially and environmentally sustainable production system. It is therefore becoming increasingly important from a farm management viewpoint to understand the inter-relationships between improving the efficiency of lowland livestock production systems and minimising the effects of this production on the environment.

One way of helping farmers and researchers to improve their understanding of the financial and environmental consequences of changes in their management is to make use of computer models of the dairy enterprise to evaluate the different management scenarios. To achieve this requires an holistic approach to the development of an integrated systems model of the dairy production

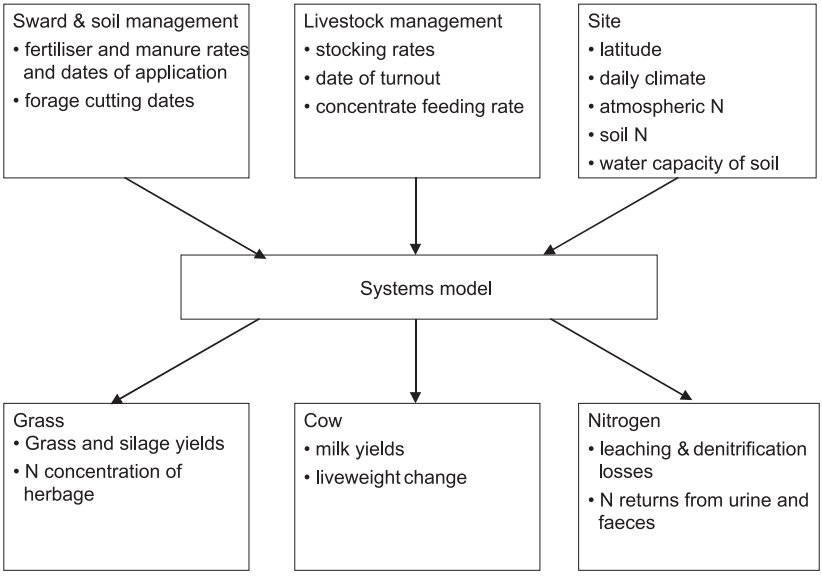


Fig. 11.1 The driving variables and outputs from the systems model.

enterprise, which combines these elements in order to understand the environmental and economic viability. Such a model requires the integration of the key components of the production system; namely forage production and utilisation, animal production and an economic evaluation component.

In order for the model to be relevant to the farmer, the inputs and outputs must either relate to the management practices or are describing the site. Hence, the driving variables of the integrated model are determined by farm management practices, and the outputs from the model must also be relevant to the farmer. The driving variables and the outputs from the model are illustrated in Fig. 11.1. Any model developed must be applicable to any site. As a consequence, the systems model will need to be process based, and hence a mechanistic model has been developed to describe the system. Accordingly, the behaviour of the system is based on the next lower level in the hierarchical system, and the description at the lower level may be mechanistic or empirical (France and Thornley 1984).

11.2 The model structure

11.2.1 Introduction

The systems model structure developed in this chapter is based on an integrated modelling approach (Fig. 11.2), and describes the following key factors:

- nitrogen use and its associated environmental effects
- grassland production

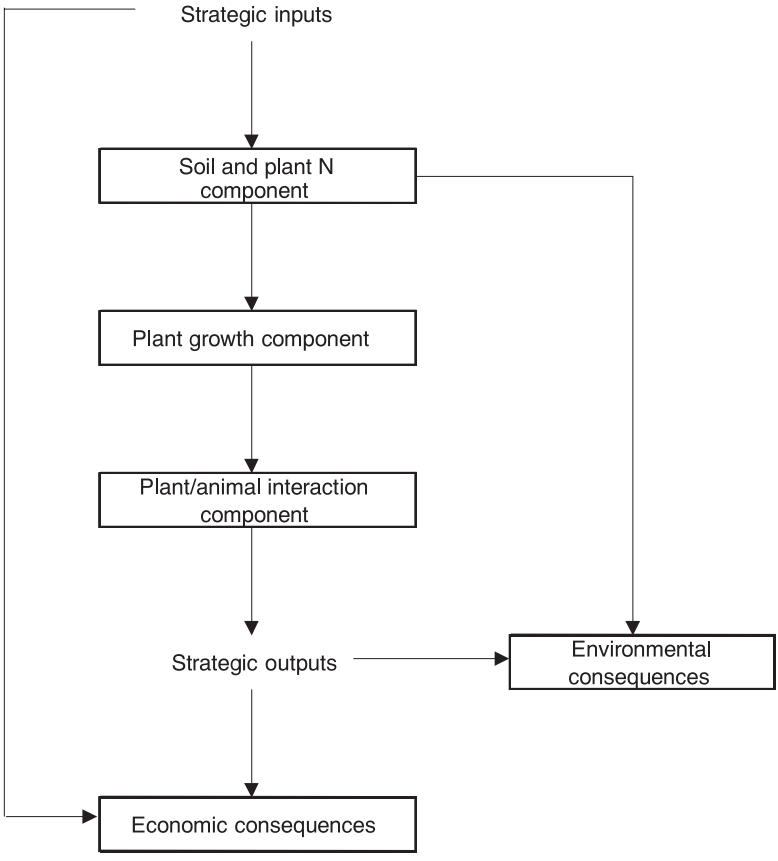


Fig. 11.2 A schematic diagram of the integrated model.

- grassland utilisation
- economic factors.

Within the model, time is measured in days from 1 January.

11.2.2 Nitrogen use and its associated environmental effects

The soil nitrogen sub-model describes the nitrogen transformations that are occurring in the soil, and hence the nitrogen that is available for crop growth and that which is lost through nitrogen leaching and denitrification (Fig. 11.3). Within the systems model, the major processes, transformations and outputs of nitrogen in agricultural soils are described in the SOILN model (Johnsson *et al.* 1987; Bergström *et al.* 1991), and the rates of transformations are affected by the soil temperature and soil water content, which are derived from running the SOIL model (Jansson 1996). The soil in the sub-model is divided into layers on the basis of physical and biological characteristics, and the organic and

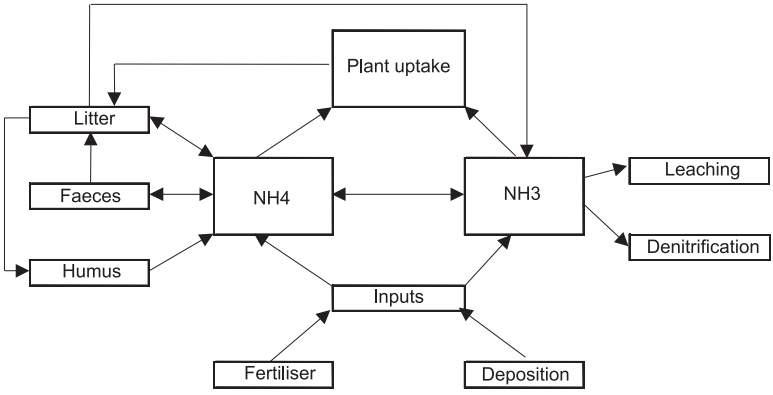


Fig. 11.3 A schematic diagram of the N sub-model.

inorganic pools are replicated for each layer. The soil nitrogen is described in terms of three organic pools, namely litter, humus and faeces, and two inorganic pools, which are ammonium and nitrate. The fertiliser nitrogen is added into the inorganic pool, while the manure and slurry nitrogen is added to the faecal and litter organic pools and the inorganic pools.

The linkage between the soil nitrogen model and the grass growth model is through the uptake of nitrogen. The demand for nitrogen by the plant is dependent on the maximum nitrogen concentration of the leaf, stem and root material (Eckersten and Jansson 1991), and it is therefore dependent on the quantity of new biomass produced during that time period. Following Angus and Moncur (1985), the maximum and minimum nitrogen concentration of the leaves is dependent on the stage of development of the crop, which is a function of the average daily temperature and the photoperiod (Gustavsson *et al.* 1995). The uptake of nitrogen, assuming the available nitrogen is less than the plant requires, is dependent on the quantity of nitrogen available within each layer of the soil. In the model, Eckersten and Jansson (1991) have assumed that the priority for allocating the nitrogen is roots, stem and then leaves. Consequently, the requirement of the roots for nitrogen is satisfied before nitrogen is allocated to the stem. The senescent sward material releases nitrogen into the litter pool within the soil.

The nitrogen excreted by the dairy cow impacts not only on pasture production, but also on the nitrogen losses from the system. This sub-model within the systems framework is based on the model developed by the Agricultural and Food Research Council (AFRC) (1993). The protein in the diet of the dairy cow is composed of quickly degradable protein (QDP, $g\,kg^{-1}$), slowly degradable protein (SDP, $g\,kg^{-1}$), effective rumen degradable protein (ERDP, $g\,kg^{-1}$) and undigestible protein (UDP, $g\,kg^{-1}$). A proportion of the UDP is digested in the lower intestines (DUP, $g\,kg^{-1}$). This is a function of the UDP and the acid detergent insoluble nitrogen within the feed. The ERDP available is a function of the fermentable metabolisable energy (FME, $MJ\,head^{-1}\,day^{-1}$), crude protein contents (CP, $g\,head^{-1}\,day^{-1}$) and the

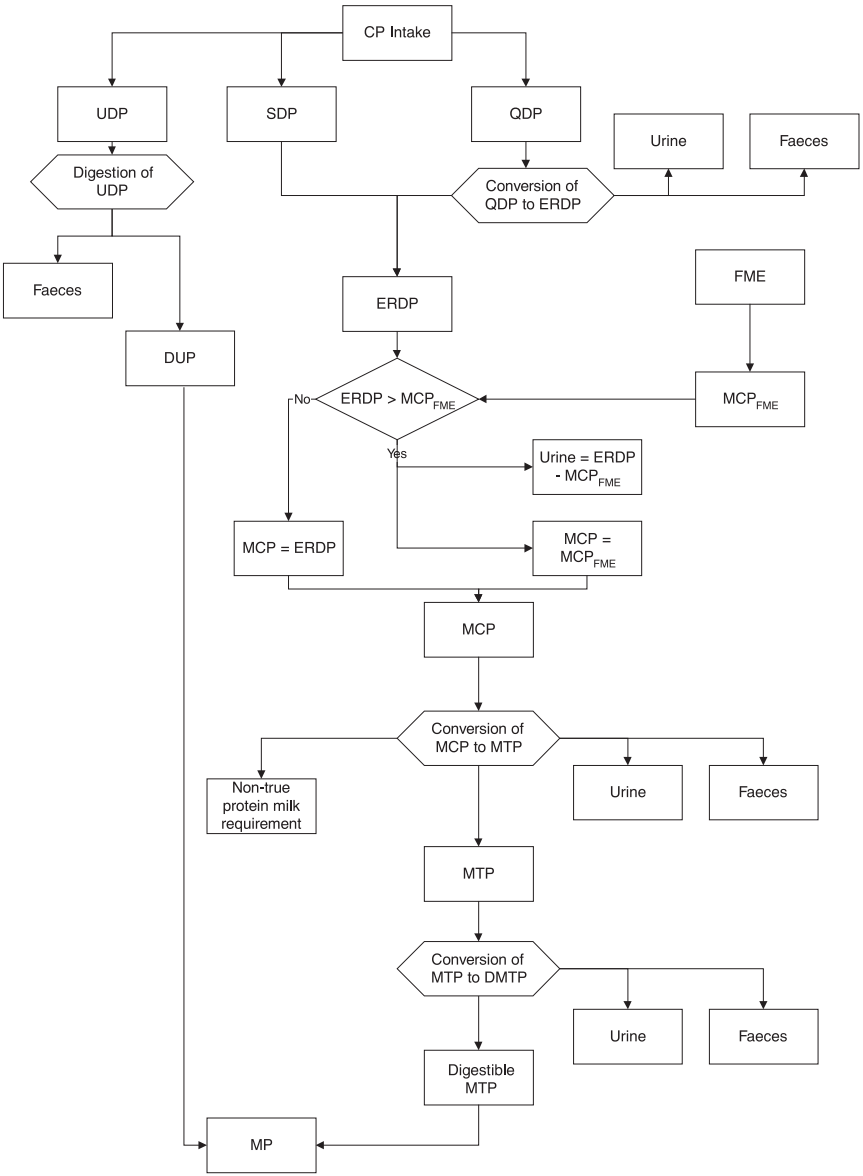


Fig. 11.4 A schematic diagram of the intake of crude protein and the production of microbial protein (MP).

degradability of the feed, which are predefined for a fresh grass and a dairy concentrate (AFRC 1993). The microbial crude protein (MCP, g head^{-1} day^{-1}) that is available is the minimum of the ERDP and the microbial crude protein yielded from the FME (MCP_{FME}, g head^{-1} day^{-1}). The microbial true protein content (MTP, g head^{-1} day^{-1}) is 75% of the MCP, as 25% of the MCP is in the

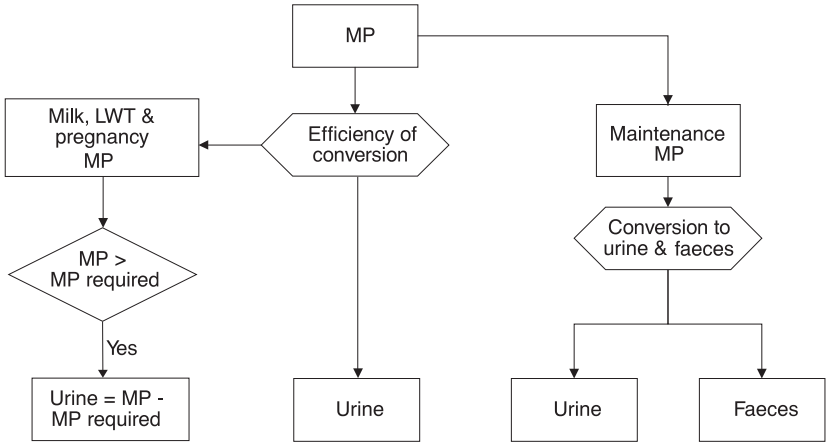


Fig. 11.5 A schematic diagram of the use of the microbial protein.

form of nucleic acids and non-protein nitrogen. Within the sub-model, the requirements of the grazing dairy cow for digestible true protein (DMTP, $g\,head^{-1}\,day^{-1}$) that is required for maintenance, pregnancy, milk production and live weight change are determined. The losses of nitrogen through faecal production are shown schematically in Figs. 11.4 and 11.5. It has been assumed that 15% of the urine nitrogen and 3% of the faecal nitrogen (Whitehead 1995) are volatilised in the field.

In the model, it is assumed that the urine and faecal nitrogen are distributed homogeneously across the sward, although it is recognised that the distribution of urine and faeces is non-uniform (Haynes and Williams 1993). In addition, the excreta can be concentrated where the livestock tend to congregate (Hilder 1966, MacDiarmid and Watkin 1972). Nevertheless, the urine and dung patches can cover between 25–35% of the sward (Haynes and Williams 1993). Each urination event is estimated to cover between 0.16 and 0.49 m^2 (Haynes and Williams 1993), with an estimated effective area of 0.5 to 0.7 m^2 (Richards and Wolton 1976, Lantinga *et al.* 1987). The surface area covered by a dung patch is between 0.05 and 0.09 m^2, but the effective area in wet areas including south-west Scotland can be up to six times greater than the area covered by the dung patch. Consequently, the area affected by the urine and faecal nitrogen will be greater than the base area covered.

11.2.3 Grassland production

The sub-model of the sward assumes that it is pure grass and a schematic diagram of the model is shown in Fig. 11.6. Forage production is calculated on a daily basis, and is presumed to be dependent on herbage mass, temperature, radiation, atmospheric carbon dioxide (CO_2) concentration, available water and nutrients. There are eight state variables, namely,

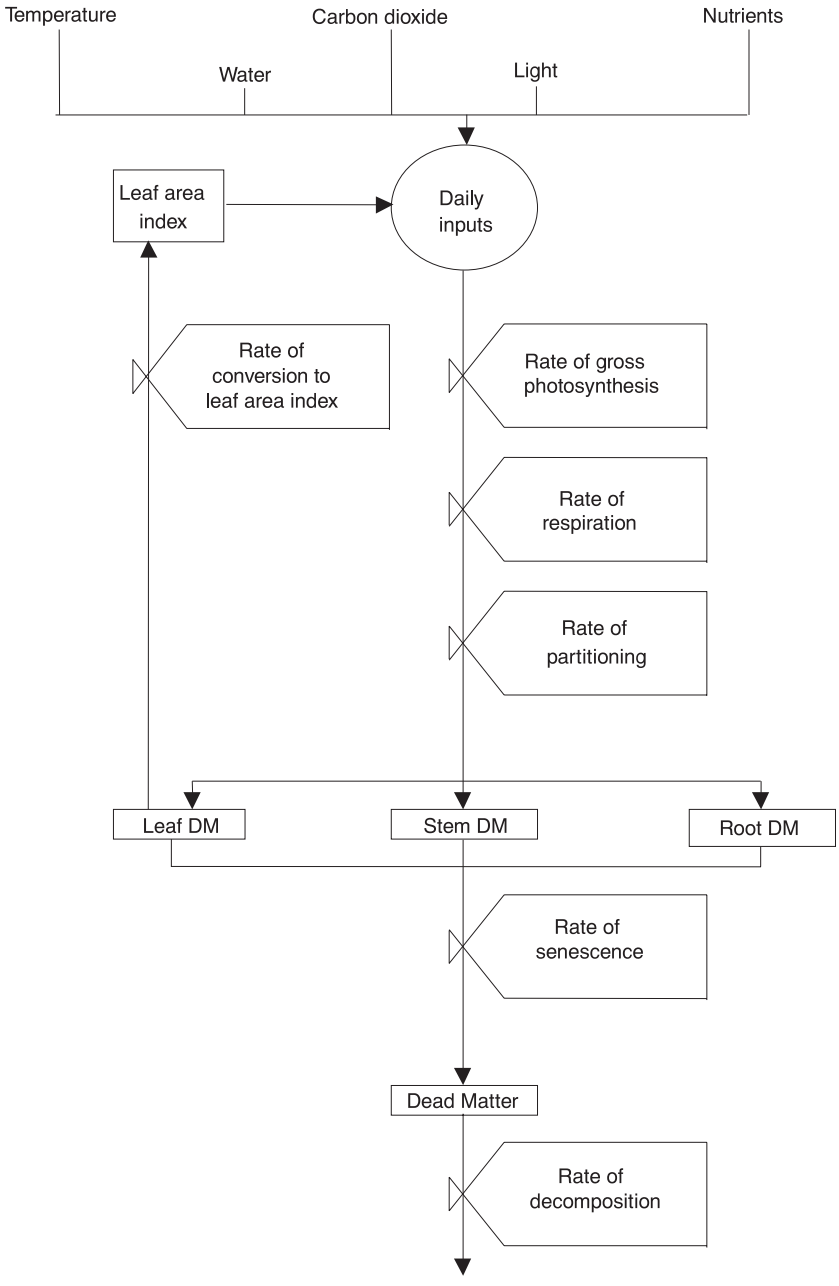


Fig. 11.6 A schematic diagram of the forage growth model.

1. leaf dry matter
2. leaf nitrogen pool
3. stem dry matter
4. stem nitrogen pool
5. root dry matter
6. root nitrogen pool
7. dead material
8. the leaf area index of the crop.

There are also five driving variables, namely the mean daily temperature, the level of photosynthetically active radiation, the atmospheric concentration of CO_2, the available moisture and the available nitrogen. Essentially temperature, photosynthetically active radiation and atmospheric CO_2 concentration are presumed to modify the rates of gross photosynthesis. Net photosynthesis is then derived by deducting respiration losses. The available moisture and nitrogen modify the net photosynthate, which is then partitioned between leaf, stem and root. The nitrogen taken up by the plant is then partitioned between the leaf, stem and root. The resultant leaf, stem and root materials are then either harvested or pass into the dead pool through decomposition, with the nitrogen being released into the litter pool within the soil. The 'stem' comprises tillers and latent developing leaves as well as true stem.

Given the structure of the model, it is convenient to divide its description into five sub-models concerned with (i) photosynthesis, (ii) respiration, (iii) water and nutrient stress, (iv) assimilate partitioning and senescence, and (v) herbage accumulation under cutting.

Photosynthesis

The canopy gross photosynthesis is described for a monoculture by the non-rectangular hyperbolae (Johnson and Thornley 1984). It is a function of the irradiance intercepted by the leaves, described by Beer's law, the photochemical efficiency and the leaf photosynthetic rate. The leaf photosynthetic rate is considered to be a function of the leaf area index (Johnson *et al.* 1989), and the mean daily temperature (Johnson and Thornley 1983). It is assumed that the photosynthetically active radiation and temperature do not vary throughout the day. Thus, the daily rate of photosynthesis can be calculated by multiplying the maximum hourly rate of leaf photosynthesis by the effective day length, where day length is based on nautical twilight. In addition to temperature affecting the rate of photosynthesis, the atmospheric concentration of CO_2 also modifies the rate as described by Thornley *et al.* (1991).

Respiration

The total respiration requirement of the sward can be divided into growth and maintenance components. The growth respiration is related to the gross photosynthate, and the maintenance respiration is related to the mass of the plant and the growth conversion efficiency (Thornley 1976). Following Johnson and

Thornley (1983), the maintenance respiration requirement increases linearly with temperature.

Water and nutrient stress

The effect of a reduction in the availability of water or plant nutrients will be to reduce the rate of net photosynthate of each component, either by reducing the efficiency of photosynthesis or by reducing the length of the growing period. The effect of water and nutrient stress on photosynthesis has been modelled by reducing the net photosynthesis in proportion to the stress experienced by the crop.

The principal limiting nutrient for pasture in Scotland is nitrogen. The degree of nitrogen stress experienced by the crop is the leaf nitrogen concentration expressed as a proportion of the maximum leaf nitrogen concentration. Similarly, the available soil water is expressed as a proportion of the soil water required for maximum growth. Empirically derived relationships expressing the effect of water and nutrient stress on the photosynthate for grass, have been estimated from part of the GM23 data (J. Gilbey, personal communication). The amount of nitrogen that is available to the sward is dependent on the available pool of nitrogen determined by the SOILN model.

Assimilate partitioning and senescence

The net photosynthesis expressed as kg $CO_2\,ha^{-1}$ (ground) day^{-1} is converted to dry matter by multiplying the net photosynthesis by the efficiency of converting CO_2 to dry matter. Following Doyle *et al.* (1989), pasture growth occurred when there is photosynthate surplus to requirements for tissue maintenance and growth respiration. A fixed proportion of the photosynthate is assumed to be partitioned to the root (Johnson *et al.* 1983), and the remaining photosynthate is partitioned between the leaves and the stem. Losses, through senescence, offset the production of new leaf and stem material. The senescent material passes into the pool of dead material, where it remains until it decomposes.

Sheehy *et al.* (1980) observed that, for grass, the physiological stage of development affects the proportion of photosynthate partitioned to the leaves and the rate of leaf senescence. In spring, during the reproductive phase, less assimilate is partitioned to the leaves. The apparent life of the leaf is increased, implying a lower rate of leaf loss. The commencement of the reproductive phase of each species varies with temperature and light (Cooper 1960). However, for simplicity, the changes in physiological states are assumed to occur on designated days.

Herbage accumulation under cutting

In the grass sward, the actual quantity of grass harvested under cutting is equated with the quantity of leaf and stem material in the sward less some predefined residual quantity of material that remains on the paddock.

11.2.4 Grassland utilisation

Within the model, it is assumed that half the area will be set aside for the first conservation cut and a third of the area for the second. However, if there is a shortage of pasture for grazing, the paddocks set aside for conservation are grazed. Any paddock that has not been grazed during the thirty days prior to the date of cutting is cut for conservation.

The interaction between the herbage growth and the grazing sub-models is through the removal of leaf area by the grazing animal. This removal of leaf affects the rate of photosynthesis and consequently the growth rate of the crop. This, in turn, affects the crop's morphology in terms of the leaf-to-stem ratio, which regulates the digestibility of the herbage on offer and therefore influences the herbage intake by the grazing animal.

At low herbage allowances, once the available herbage had been consumed, the animals abandon any attempt to graze closer to the ground (Le Du *et al.* 1979). Hence, in the model, it has been assumed that the dairy cows will not graze below a predefined herbage mass. The start of the grazing season is controlled by the availability of grass, and the quantity of herbage on the paddock must have increased. On the other hand, the grazing season is considered to end when one of the following criteria is met:

- the metabolisable energy from dry matter intake does not meet the metabolisable energy requirements of maintenance; or
- the predicted dry matter intake falls to less than 20% of the potential level; or
- the soil moisture availability would result in the animals poaching the paddocks.

However, grazing can still occur after the growing season has ended. Throughout the grazing season the rotation of the dairy cows around the paddocks is determined solely by the quantity of herbage on each paddock.

Basically a spring-calving dairy herd, rotationally grazed during the summer period on a pure grass sward, is simulated. The pasture is divided into a number of equal-sized paddocks. Herbage production is calculated for each paddock on a daily basis and is dependent on the existing herbage mass, the availability of nutrients, temperature, radiation and CO_2 concentration. The herd is represented in the model by the 'average dairy cow', which is assumed to comprise 25% first-year heifers, 25% second lactation cows and 50% cows in later lactations. Thus, each year 25% of the cows are presumed to be culled and replaced.

The intake of dry matter by the grazing dairy cow is assumed to be regulated by three factors (Loewer *et al.* 1983), namely

- the feed availability
- the physiological limit on intake
- the physical ability of the animal to consume feed.

Feed availability

When the quantity of herbage available for consumption is less than that required for 95% of maximum daily intake, the daily allowance of green herbage

regulates intake. The green herbage allowance is taken to be the green herbage mass above the minimum herbage mass of 900 kg DM ha^{-1} required for grazing. Zemmelink (1980) described the relationship for tropical grasses between herbage intake as a function of the daily allowance of green herbage and the maximum daily intake in kg DM per head per day. In the absence of any established relationships for temperate grasses, this relationship has been adopted. Consequently, the model has been calibrated using the assumption that the maximum daily intake of herbage is related to the metabolic live weight of the cow and is presumed to increase by 136 g DM for every kg of metabolic weight (McDonald *et al.* 1988).

Physiological limit to intake

The physiological limit to intake is considered to be regulated by the daily metabolisable energy (ME) requirements of the animal. Energy requirements by the dairy cow are divided into those for maintenance, pregnancy, milk production, and growth and fattening. The description of the maintenance requirements is based on Hulme *et al.* (1986). The net utilisation efficiency of ME for maintenance is related to the metabolisability of the feed, while the mean age of the 'average dairy cow' is assumed to be four. Daily energy requirements for pregnancy have been derived using relationships specified in Agricultural Research Council (ARC) (1980).

The potential energy requirement for lactation has been derived from estimates of the potential milk yield based on a Wood's lactation curve (Wood *et al.* 1980). The potential daily milk yield of the 'average dairy cow' is taken to be the weighted average of the potential daily milk yield of each age cohort. Finally, the estimates of the daily energy requirements for growth and fattening assume that the potential growth of an animal can be described by a Gompertz equation (Taylor 1968).

The physiological energy requirements of the 'average dairy cow' are then obtained in the model by summing the four elements, namely maintenance, pregnancy, lactation and growth energy requirements. As the energy retention of the cow is not linearly related to intake (Schiemann *et al.* 1971, van Es 1976), the physiological intake has been corrected for feeding level (ARC 1980).

Physical limit to intake

With feeds having a low digestibility, the actual intake may be lower than the physiological requirement. Feed intake is controlled by the rate of passage of undigested material through the digestive tract, and the rate is positively related to the digestibility of the feed (Conrad *et al.* 1964). It is also considered that the capacity of the digestive tract of the cow is influenced by the stage of lactation. Following Kahn and Spedding (1984), the ability of the digestive tract to process and void undigested residues was increased linearly up to a maximum value on day 150 of lactation, and then decreased linearly back to the base level at the end of lactation. At the same time, following the recommendations of the ARC (1980), the physical limit to herbage intake is corrected for the effects of

concentrate feeding. This is because as the level of concentrates increases, the intake of herbage decreases, so that the net effect of supplementing the diet only results in a small increase in the dry matter intake (Mayne 1990).

Components of intake

The actual daily intake is described by the most limiting factor of the physiological, the physical and the herbage limitations to intake (Fig. 11.7). However, this provides no information on the composition of the diet in terms of leaf, stem or dead material. Observations by Jamieson and Hodgson (1979) have shown that grazing lambs and calves preferentially select green material and hence the same has been assumed for dairy cows. The proportions of leaf, stem and dead material in the sward are also known to differ from the proportions in the diet (Rattray and Clark 1984). Accordingly, following Doyle *et al.* (1989) the mean daily intakes of leaf and stem are functions of the proportions of leaf and stem in the sward plus an intake of dead material. In the model, the proportionate digestibilities of the leaf and stem material have been assumed to decrease as the season progresses (Osbourn 1980).

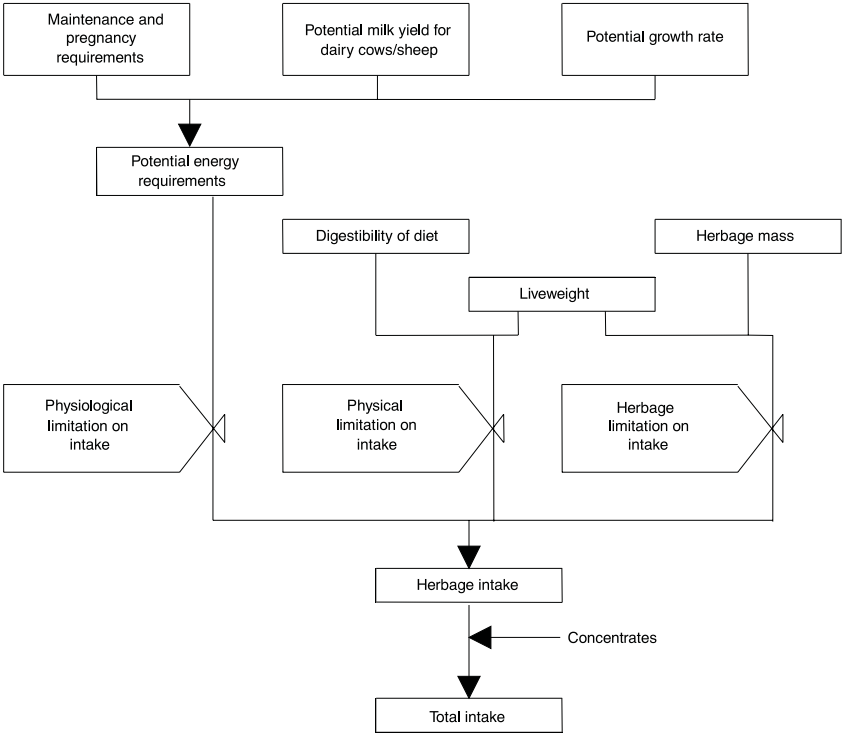


Fig. 11.7 A schematic representation of the factors limiting intake.

11.2.5 Economic considerations

It is crucial within the systems model to assess the impact of changing the management system on the profitability of the system, as well as on the environment. As a consequence, two common indicators, namely margin over concentrates and margin over forage, which allow comparisons to be made between different systems, are determined within the systems model. These two indicators require the determination of the expenditure on concentrates and fertilisers, and revenue from milk, and hence the milk yield.

The user enters the milk price, while the milk yield is determined within the modelling framework. The fertiliser usage is determined by the management strategy adopted and hence the user enters the quantity of fertiliser applied per hectare, as well as the price per unit of fertiliser. The user of the model can determine whether the concentrate-feeding regime adopted is flat rate or whether the cows are fed to yield. Accordingly, the user enters the concentrates fed either per litre of milk produced or per day and the cost per unit of concentrates. The level of milk production is a function of the herbage available and the concentrates fed, and the energy requirement of the dairy cow.

Dairy cow production

Within the model, the energy intake is partitioned between maintenance, pregnancy, live weight gain and milk production. The energy requirements for maintenance and pregnancy are considered to have priority. If there is insufficient energy available to meet the potential energy requirements of the animal, it is assumed that the potential energy requirements for milk and growth are reduced by an equal amount (Bruce *et al.* 1984), illustrated in Fig. 11.8. In the event of the maternal body being catabolised to meet maintenance and pregnancy requirements, the energy available for milk production may become less than zero. If this occurs, no milk is produced and the quantity of maternal body catabolised is restricted to the shortfall in energy requirements for maintenance and pregnancy.

11.2.6 Validation of the model

The herbage production model has been validated for several sites in the UK (Topp 1999a, Topp and Doyle 1996a, Topp and Hameleers 1998) and Northern Europe (Topp 1999b). In addition, milk production produced from the dairy cow model has been validated against data for an indoor feeding system (Topp 1999a) and a rotationally grazed pasture in Ireland (Topp 1999a, Topp and Doyle 1996b). Topp and McGechan (2000) have described the validation of the herbage production linked to the environmental consequences of the systems model. The model was validated against herbage yield, nitrogen concentration of the herbage and nitrogen leaching data, which was collected from lysimeters for the Crichton Royal dairy farm, Dumfries, Scotland for 1993 to 1996 (Hooda *et al.* 1998). This required the calibration of the SOIL (McGechan *et al.* 1997) and SOILN models (Wu *et al.* 1998) for the site.

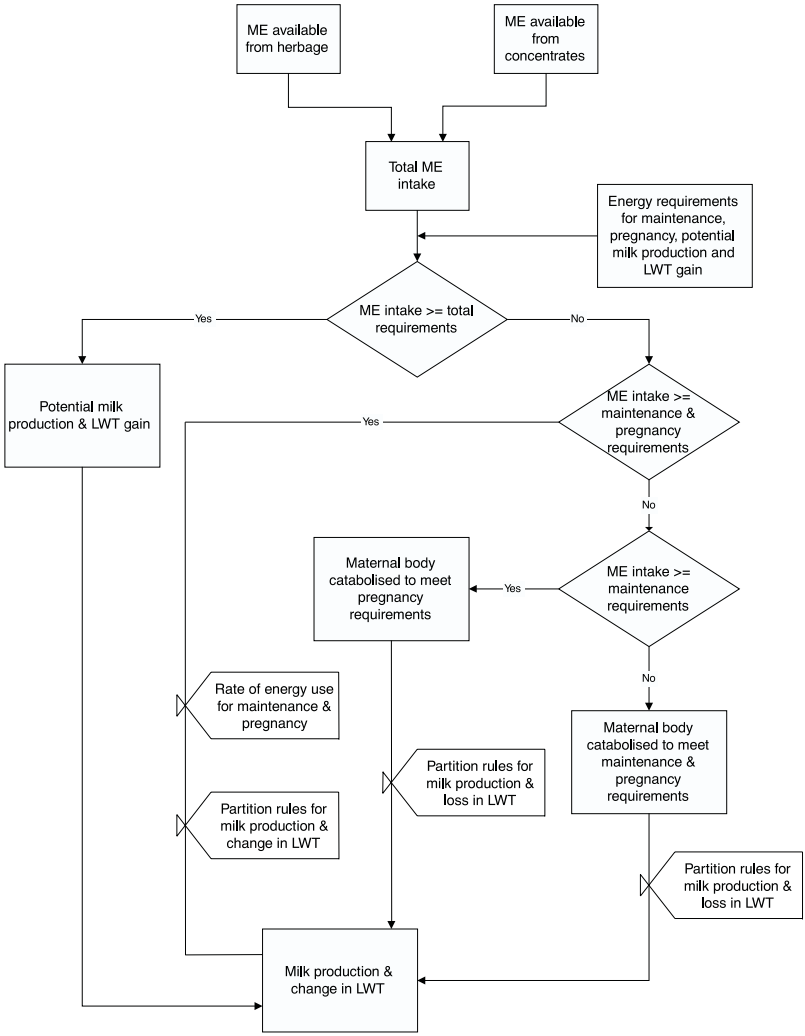


Fig. 11.8 A schematic diagram of the partitioning of metabolisable energy in the dairy cow.

The statistical technique used to assess the ability of the forage and livestock models to simulate production is Theil's inequality coefficient (u), which was defined by Theil (1970) as:

$$u = \frac{\sqrt{\frac{1}{n}\sum_{n=1}^{n}(\text{obs} - \text{pred})^2}}{\sqrt{\frac{1}{n}\sum_{n=1}^{n}(\text{obs})^2} + \sqrt{\frac{1}{n}\sum_{n=1}^{n}(\text{pred})^2}} \tag{11.1}$$

where pred and obs are the predicted and observed values respectively and *n* is the number of observations. The numerator of this equation is the root mean square error, while the denominator scales Theil's inequality statistics so that it always falls between zero and one, with zero indicating a perfect fit. As the difference between the actual and the simulated values are squared, large errors are penalised more than small errors. This technique also assesses the model's ability to duplicate turning points or rapid changes in the data. Theil's inequality coefficient can be decomposed into the bias, variance and covariance proportions. The bias proportion is an indication of the systematic error, while the variance proportion represents the ability of the model to replicate the degree of variability in the observed data. The covariance proportion represents the error remaining after accounting for the bias and variance proportions.

The grass production element of the model was validated using three sites at a range of fertiliser application rates in the UK (Morrison *et al.* 1980) and a large number of sites across Northern Europe during the period 1982–1986 (Corrall 1988, Peeters and Kopec 1996). The results of the validation, tested using Theil's inequality coefficient, suggest that the systems model can predict grass production over the growing season across a large number of sites in the UK (Topp and Doyle 1996a, Topp and Hameleers 1998, Topp 1999a) and Northern Europe for irrigated and non-irrigated situations (Topp 1999b), as shown in Tables 11.1–11.3. However, Table 11.4 reveals that the ability of the model to predict varies with the level of nitrogen fertiliser applications (Topp 1999a). The model tends to under-predict at low nitrogen application rates and over-predict

Table 11.1 Theil's inequality coefficient and bias, variance and covariance proportions of the coefficients for three sites in the UK

Site	Inequality coefficient	Bias proportion	Variance proportion	Covariance proportion
High Mowthorpe	0.23	0.225	0.205	0.570
Seale Hayne	0.18	0.000	0.015	0.985
Rosemaund	0.18	0.009	0.042	0.949

Table 11.2 Theil's inequality coefficient and bias, variance and covariance proportions for the FAO sites which were irrigated

Site	Inequality coefficient	Bias	Variance	Covariance
Crossnacreevy, UK	0.19	0.13	0.19	0.69
Hurley, UK	0.35	0.39	0.19	0.43
Uppsala, Sweden	0.42	0.31	0.27	0.42
South Savo, Finland	0.31	0.23	0.19	0.58
North Pohjannaa, Finland	0.29	0.13	0.16	0.71
North Wyke, UK	0.34	0.39	0.23	0.39

Table 11.3 Theil's inequality coefficient and bias, variance and covariance proportions for the FAO sites which were non-irrigated

Site	Inequality coefficient	Bias	Variance	Covariance
Braunschweig, Germany	0.31	0.19	0.07	0.74
Plas Gogerddan, UK	0.39	0.41	0.25	0.34
Crossnacreevy, UK	0.20	0.04	0.08	0.88
Johnstown Castle, UK	0.29	0.12	0.07	0.82
Grange, Eire	0.27	0.15	0.22	0.63
Moorpark, Eire	0.24	0.01	0.02	0.98
Hurley, UK	0.35	0.34	0.13	0.54
Uppsala, Sweden	0.51	0.32	0.28	0.41
South Savo, Finland	0.31	0.18	0.11	0.72
North Pohjannaa, Finland	0.27	0.03	0.02	0.96
MacRobert, UK	0.25	0.19	0.01	0.80
Auchincruive, UK	0.23	0.18	0.12	0.71
Kiel, Germany	0.24	0.00	0.00	1.01
North Wyke, UK	0.33	0.37	0.20	0.43
Bronydd Kawt, UK	0.37	0.32	0.22	0.48

Table 11.4 The ratio of the predicted yield: observed yield averaged nitrogen level at three sites in the UK

Nitrogen level ($N\,kg\,ha^{-1}\,yr^{-1}$)	High Mowthorpe	Rosemaund	Seal Hayne
0	0.79	0.38	0.30
150	0.90	0.76	0.96
300	1.17	1.07	1.03
450	1.78	1.28	1.31
600	2.24	1.68	1.35
Total	1.38	1.03	0.99

at high nitrogen application rates. Nevertheless, the error is mainly due to random variations, as the bias and variance proportions of the coefficient tend to be small.

The model predicted reasonably well milk production in the indoor production system, with a Theil inequality coefficient of 0.037 and bias and variance proportions accounting for 0.363 and 0.087 respectively (Topp 1999a). The grazing dairy cow model was validated for low (2.36-2.49 cows ha^{-1}) and high (3.1–3.23 cows ha^{-1}) stocking densities for the period 1985–1987 at Johnstown Castle, Ireland (Ryan 1988, 1989). Although Theil's inequality coefficient suggests reasonable predictions for the grazing herd (Table 11.5), the model tends to over-predict milk production towards the end of the grazing season as illustrated in Fig. 11.9 (Topp 1999a, Topp and Hameleers 1998).

When the herbage production model is validated for herbage yield and the impact on nitrogen leaching, the predictions for herbage production are

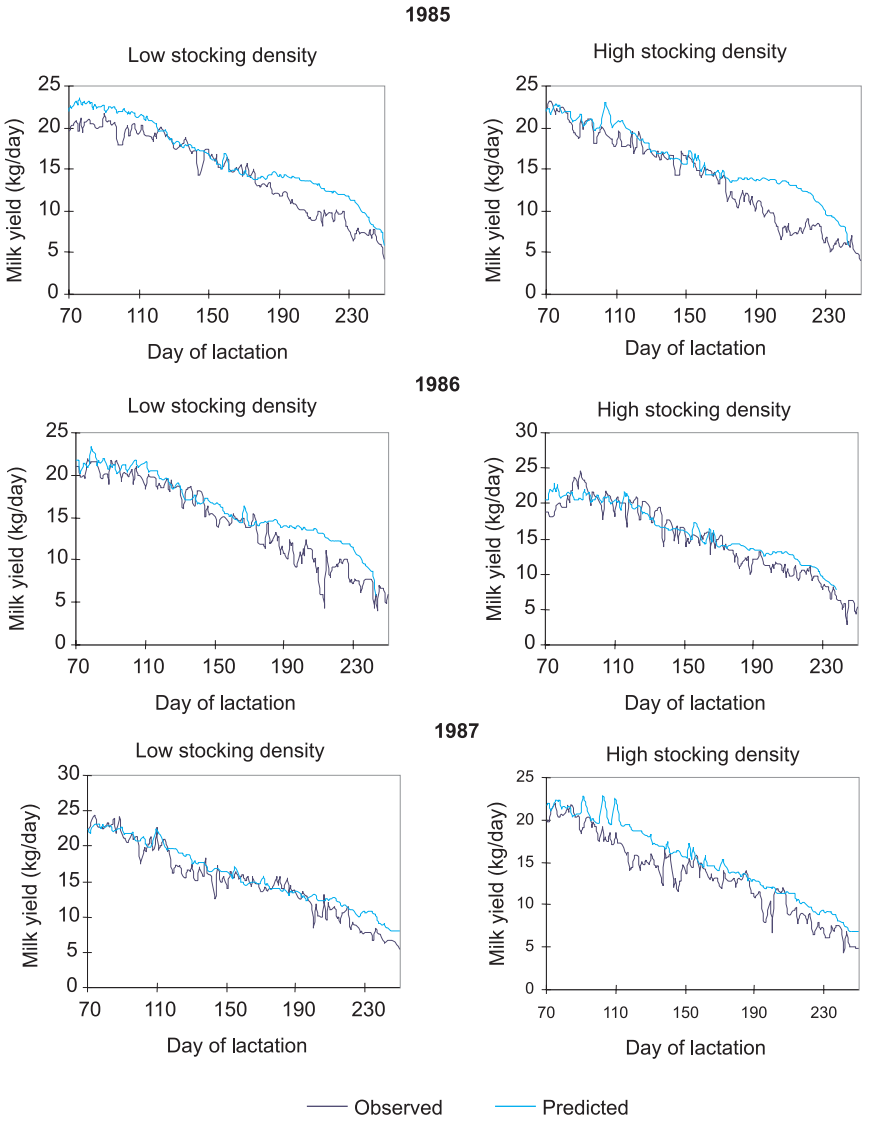


Fig. 11.9 Observed and predicted daily milk yields for the 'average cow' at An Foras Taluntais, Johnston Castle, 1985–1987.

reasonably good, with the bias and variance proportions of the Theil's inequality coefficient being relatively small. However, the Theil's inequality coefficient for nitrogen leaching suggests poor levels of fit. Nevertheless, the model does predict the general trends in nitrogen leaching production throughout the year, although there is a tendency for the predicted levels of leaching to be lower than measured, as illustrated in Fig. 11.10.

Table 11.5 Theil's inequality coefficient and bias, variance and covariance proportions of the coefficients

Stocking density	Year	Theil statistic	Bias	Variance	Covariance
Low	1985	0.069	0.575	0.020	0.405
	1986	0.070	0.470	0.184	0.346
	1987	0.051	0.291	0.131	0.578
	1985–1987	0.064	0.443	0.084	0.473
High	1985	0.082	0.494	0.168	0.338
	1986	0.053	0.154	0.141	0.705
	1987	0.075	0.622	0.000	0.378
	1985–1987	0.071	0.418	0.075	0.507

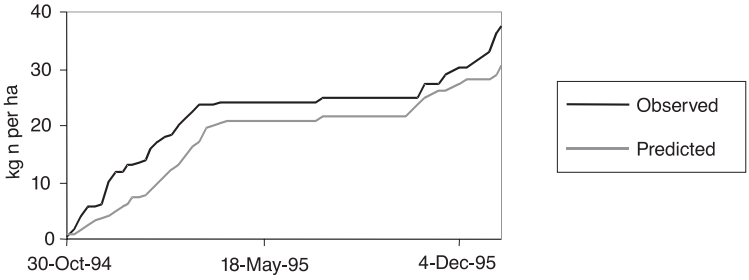


Fig. 11.10 The cumulative observed and predicted N leaching for the Crichton Royal Farm, Dumfries.

11.3 Conclusions

The holistic model of the dairy enterprise is dynamic and deterministic and hence is applicable across a wide range of sites, which has been demonstrated as part of the validation process. The model describes not only milk and herbage production, but also the nitrogen flows within the system, and hence the losses through leaching and denitrification. The model also considers the economic impacts of changing the management system. Accordingly, the dairy enterprise model can be used for scenario planning and asking 'what if' questions of the management system. However, the model does not determine the optimum management system, and hence it is unable to propose an optimum grazing regime or fertiliser application regime.

The model has been used to explore the consequences on herbage yield throughout the growing season and leaching losses of different fertiliser application rates and dates coupled with varying application rates and dates of slurry. Similarly, the impact of fertiliser and slurry on the availability of herbage for grazing and hence milk yield has also been explored. In addition, the impact of

grazing dairy cows returning nutrient to the pasture through urine and faecal deposits on the system has been assessed (Topp and Hameleers 1999). Furthermore, by running the model over a number of years, the interaction between the climate and the management factors on the production and viability of the dairy system can be explored. However, daily climate data, describing the average temperature, rainfall, radiation and evapotranspiration, are required. Thus to run the model for a series of years requires large quantities of climatic data. This data requirement and the limitations of current personal computer power currently limit the use of the model as an on-farm decision support system. However, this limitation will be reduced as computer power increases. Nevertheless, the model was developed to understand the system as opposed to being used as a decision support system. However, the model can be used to assess the impact of changing the management system on the efficiency of the system.

11.4 Acknowledgements

SAC receives financial support from the Scottish Executive Rural Affairs Department.

11.5 References

AGRICULTURAL AND FOOD RESEARCH COUNCIL, (1993) *Energy and Protein Requirements of Ruminants*. Wallingford, UK: CAB International.

AGRICULTURAL RESEARCH COUNCIL, (1980) *The Nutrient Requirements of Ruminant Livestock*. Commonwealth Agricultural Bureaux, Farnham Royal, UK.

ANGUS J F, MONCUR M W, (1985) Models of growth and development of wheat in relation to plant nitrogen. *Australian Journal of Agricultural Research*, **36** 537–44.

BERGSTRÖM L, JOHNSSON H, TORSTENSSON G, (1991) Simulation of soil nitrogen dynamics using the SOILN model. *Fertilizer Research*, **27** 181–8.

BRUCE J M, BROADBENT P J, TOPPS J H, (1984) A model of the energy system of lactating and pregnant cows. *Animal Production*, **38** 351–62.

CONRAD H R, PRATT A D, HIBBS J W, (1964) Regulation of feed intake in dairy cows. I. Change in the importance of physical and physiological factors with increasing digestibility. *Journal of Dairy Science*, **47** 54–62.

COOPER J P, (1960) Short-day and low-temperature induction in *Lolium*. *Annals of Botany*, **24** 232–46.

CORRALL A J, (1988) Prediction of production from grassland. *Herba* (FAO), **1** 25—8.

DOYLE C J, BARRS J A, BYWATER A C, (1989) A simulation of bull beef production under rotational grazing in the Waikato region of New Zealand. *Agricultural Systems*, **31** 247–78.

ECKERSTEN H, JANSSON P-E, (1991) Modeling water-flow, nitrogen uptake and production for wheat. *Fertilizer Research*, **27** 313–29.

FRANCE J, THORNLEY J H M, (1984) *Mathematical Models in Agriculture*, Butterworths, London.

GUSTAVSSON A-M, ANGUS J F, TORSSELL B W R, (1995) An integrated model for growth and nutritional value of Timothy. *Agricultural Systems*, **47** 73–92.

HAYNES R J, WILLIAMS P H, (1993) Nutrient cycling and soil fertility in the grazed pasture ecosystem. *Advances in Agronomy*, **49** 119–99.

HILDER E J, (1966) Distribution of excreta by sheep at pasture. *Proceedings of the 10th International Grassland Congress*, 977–81.

HOODA P S, MOYNAGH M, SVOBODA I F, ANDERSON H A, (1998) A comparative study of nitrate leaching from intensively managed monoculture grass and grass-clover pastures. *Journal of Agricultural Science*, **131** 267–75.

HULME D J, KELLAWAY R C, BOOTH P J, (1986) The CAMDAIRY model for formulating and analysing dairy cow rations. *Agricultural Systems*, **22** 81–108.

JAMIESON W S, HODGSON J, (1979) The effects of variation in the sward characteristics upon the ingestive behaviour and herbage intake of calves and lambs under a continuous stocking management. *Grass and Forage Science*, **34** 273–82.

JANSSON P-E, (1996) Simulation model for soil water and heat conditions – a description of the SOIL model. Department of Soil Sciences. The Swedish University of Agricultural Sciences, Uppsala, Sweden.

JOHNSON I R, THORNLEY J H M, (1983) Vegetative crop growth model incorporating leaf area expansion and senescence, and applied to grass. *Plant, Cell and Environment*, **6** 721–9.

JOHNSON I R, THORNLEY J H M, (1984) A model of instantaneous and daily canopy photosynthesis. *Journal of Theoretical Biology*, **107** 531–45.

JOHNSON I R, AMEZIANE T E, THORNLEY J H M, (1983) A model of grass growth. *Annals of Botany*, **51** 599–609.

JOHNSON I R, PARSONS A J, LUDLOW M M, (1989) Modelling photosynthesis in monocultures and mixtures. *Australian Journal of Plant Physiology*, **16** 501–16.

JOHNSSON H, BERGSTRÖM L, JANSSON P-E, PAUSTIAN K, (1987) Simulated nitrogen dynamics and losses in a layered agricultural soil. *Agriculture, Ecosystems and Environment*, **18** 333–56.

KAHN H E, SPEDDING C R W, (1984) A dynamic model for the simulation of cattle herd production systems: 2. An investigation of various factors influencing the voluntary intake of dry matter and the use of the model in their validation. *Agricultural Systems*, **13** 63–82.

LANTINGA E A, KEUNING J A, GROENWOLD J, DEENAN P J A G, (1987) Distribution of excreted nitrogen by grazing cattle and its effects on sward quality, herbage production and utilization. In: van der Meer H G, Unwin R J, van Dijk T A and Ennik G C (eds) *Animal Manure on Grassland and Fodder Crops. Fertilizer or Waste*, pp. 103–7, Dordrecht: Martinus Nijhoff Publishers.

LE DU Y L P, COMBELLAS J, HODGSON J, BAKER R D, (1979) Herbage intake and milk production by grazing cows. 2. The effects of level of winter feeding and daily herbage allowance. *Grass and Forage Science*, **34** 249–60.

LOEWER O J, SMITH E M, GAY N, FEHR R, (1983) Incorporation of environment and feed quality into a net energy model of beef cattle. *Agricultural Systems*, **11** 67–94.

MACDIARMID B N, WATKIN B R, (1972) The cattle dung patch. 3. Distribution and rate of decay of dung patches and their influence on grazing behaviour. *Journal of the British Grassland Society*, **27** 48–54.

MAYNE C S, (1990) Effects of supplementation on the performance of both growing and lactating cattle at pasture. In: Mayne, C S (ed.) *Management Issues for the Grassland Farmer in the 1990's*, British Grassland Occasional Symposium No. 25, 55–71.

MCDONALD P, EDWARDS R A, GREENHALGH J F D, (1988) *Animal Nutrition* (4th Edn), John Wiley and Sons Inc., New York.

MCGECHAN M B, GRAHAM R, VINTEN A J A, DOUGLAS J T, HOODA P S, (1997) Parameter selection and testing the soil water model SOIL. *Journal of Hydrology*, **195**, 312–34.

MORRISON J, JACKSON M V, SPARROW P E, (1980) The Response of Perennial Ryegrass to Fertiliser Nitrogen in Relation to Climate and Soil: Report of the Joint ADAS/GRI Grassland Manuring Trial – GM20. *Grassland Research Institute Technical Report No. 27*. The Grassland Research Institute, Hurley, Maidenhead, Berkshire, UK.

OSBOURN D F, (1980) Grazing management. In: Holmes W (ed.) *Grass and Its Production and Utilization*, Blackwell Scientific Publications, London, pp. 70–124.

PEETERS A, KOPEC S, (1996) Production and productivity of cutting grasslands in temperate climates of Europe. In: Parente G, Frame J and Orsi S (Eds) *Grassland and Landuse Systems*, 16th EGF Meeting 1996, pp. 59–73.

RATTRAY P V, CLARK D A, (1984) Factors affecting the intake of pasture. *New Zealand Agricultural Science*, **18** 141–6.

RICHARDS I R, WOLTON K M, (1976) The spatial distribution of excreta under intensive cattle grazing. *Journal of the British Grassland Society*, **31** 89–92.

RYAN M, (1988) Irish studies on milk production from grass/white clover swards. In: *The Grassland Debate: White Clover versus Applied Nitrogen*, Proceedings of a Conference held at the National Agriculture Centre, Stoneleigh, 19 October. Kenilworth: Royal Agricultural Society of England, Agricultural Development and Advisory Service.

RYAN M, (1989) Development of a legume-based dairy system. In: Plancquaert P and Hagger P (Eds) *Legumes in Farming Systems*, Kluwer Academic Publishers, Dordrecht, pp. 159–67.

SCHIEMANN R, JENTSCH W, WITTENBURG H, (1971) Zur Abhängigkeit der Verdaulichkeit der Energie und der Nährstoffe von der Höhe der Fetleraufnahme under der Rationszusammensetzung bei Milchkühen. *Archiv für Tierrernährung*, **21** 223–40.

SHEEHY J E, COBBY J M, RYLE G J A, (1980) The use of a model to investigate the influence of some environmental factors on the growth of perennial ryegrass. *Annals of Botany*, **46** 343–65.

TAYLOR ST C S, (1968) Time taken to mature in relation to mature weight for sexes, strains and species of domesticated mammals and birds. *Animal Production*, **10** 157–69.

THEIL H, (1970) *Economic Forecasts and Policy* (2nd edn), North Holland Publishing Company, Amsterdam.

THORNLEY J H M, (1976) *Mathematical Models in Plant Physiology*. Academic Press.

THORNLEY J H M, FOWLER D, CANNEL M G R, (1991) Terrestrial carbon storage resulting from CO_2 and nitrogen fertilization in temperate grasslands. *Plant, Cell and Environment*, **14** 1007–11.

TOPP C F E, (1999a) The implications of climate change on forage-based livestock systems in Scotland. Ph.D. thesis, University of Glasgow.

TOPP C F E, (1999b) *Low-input animal production based on forage legumes for silage, Project Task 5, Economic Modelling Working Paper 4, Validation of Forage Model*, EU Contract No. FAIR CT 96–1832.

TOPP C F E, DOYLE C J, (1996a) Simulating the impact of global warming on milk and forage production in Scotland: 1. The effects on dry-matter yield of grass and grass–white clover swards. *Agricultural Systems*, **52** 231–42.

TOPP C F E, DOYLE C J, (1996b) Simulating the impact of global warming on milk and forage production in Scotland: 2. The effects on milk yields and grazing management of dairy herds. *Agricultural Systems*, **52** 243–70.

TOPP C F E, HAMELEERS A L J, (1999) Modelling nutrient flows within dairy production systems. In: Corrall A J (ed.) *Accounting for Nutrients: A challenge for grassland farmers in the 21st Century*, Occasional Symposium of the British Grassland Society, No. 33. 29–34.

TOPP C F E, MCGECHAN M, (2000) Modelling nutrient flows within a grassland system. In: Søegaard K, Ohlsson C, Sehested J, Hutchings N J and Kristensen T (eds) *European Grassland Federation 2000 Grassland Farming. Balancing Environmental and Economic Demands*, **5**, 465–7. Rounborgs grafiske hus, Holstebro, Denmark.

TOPP K, HAMELEERS A, (1998) Mechanistic modelling of dairy production Systems. In: Tijskens L M M and Hertog M L A T M (eds) *Proceedings of the International Symposium on Applications of Modelling as an Innovative Technology in the Agri-Food-Chain Model-It, Acta Horticulturae*, **476**, 157–63.

VAN ES A J H, (1976) Factors influencing the efficiency of energy utilisation by beef and dairy cattle. In: Swan H and Broster W H (eds) *Principles of Cattle Production*. London: Butterworths (Proceedings Easter School Agricultural Science University of Nottingham, 1975), pp. 237–53.

WHITEHEAD D C, (1995) *Grassland Nitrogen*, Guildford: CAB International.

WOOD P D P, KING J O L, YOUDAN P G, (1980) Relationships between size, live-weight change and milk production characters in early lactation in dairy

cattle. *Animal Production*, **31** 143–51.

WU L, MCGECHAN M B, LEWIS D R, HOODA P S, VINTEN A J A, (1998) Parameter selection and testing the soil nitrogen dynamics model SOILN. *Soil Use and Management*, **14** 170–81.

ZEMMELINK G, (1980) Effect of selective consumption on voluntary intake and digestibility of tropical forages. *Agricultural Research Reports 896*, Pudoc, Wageningen.

12

Beef cattle production

K. G. Rickert, University of Queensland, Gatton

12.1 Introduction

This chapter refers to computer models that simulate beef cattle production. Such models consist of mathematical equations and instructions which mimic the roles, interactions and influences of the various inputs to beef cattle production. The chapter recognises that modelling is a term which refers to both building and using models, and that beef cattle production includes the complex interactions between the physical environment, financial environment, management, feed supply, and animal reproduction and growth. The chapter considers the challenge faced by model builders in dealing with such complexity, overviews possible applications, and gives an example of a simple beef production model.

Pasture and animal scientists started to model beef cattle production after computers first became available for research in the 1960s. A rapid expansion in the range, scope and role of models followed in response to the even more rapid expansions in the power and accessibility of computers. Insight into the progress and philosophy of modelling pasture and animal production are obtained from recent reviews.[1, 2] Models have been a valuable aid to research, extension, and management at the farm, industry or government levels because of the following three attributes.

1. If each equation in a model is regarded as a hypothesis pertaining to a specific process or component, then a model can be regarded as a collection of hypotheses, derived from past research that can be further modified and developed through new research. In this way, a model becomes a repository for past research and a precursor for future research.[3, 4] Model construction is now a common activity that gives research direction and focus.

2. Models provide a quantitative description of the many interacting components which may have conflicting responses in a beef production system.[4] This is a powerful and unique attribute that greatly exceeds the analytical capacity of the human mind. For example with beef cattle, as stocking rate increases (the number of animals per unit area of land), the liveweight and value per animal decrease, variable costs increase and production per hectare at first increases and then decreases.[5] A manager must balance the trade-offs between profit, risk, pasture degradation and premium prices.[6] Similar trade-offs between productivity, stability and sustainability are common in farming systems[7] and a model allows users to experience 'virtual' reality in managing grazing systems.
3. Models can give a quantitative extrapolation in space and time of information derived from past research and experiences. For example, by processing historical records of daily weather data, estimates of variability in output can be expressed as probability distributions.[8] Similarly, by processing the historical weather for different land units in a region, and thereby estimating spatial and temporal variations in forage production, estimates of safe stocking rates can be compared against trends in actual regional stocking rates to indicate periods of overgrazing.[9] Further, if the spatial model uses current weather data as input, the output is a near real-time display of pasture and/or animal production[10] that can influence government or industry policies. All of these applications rely on a model's ability to extrapolate information in temporal and spatial dimensions, and this attribute is fundamental to the role of models in information transfer.[11]

Today a wide range of models on different aspects of plant and animal production are being used as aids to research, farm management, and to determine government or industry policies.[1]

12.2 Elements of beef cattle production

Beef cattle production deals with the conversion of climatic and edaphic inputs into plant products, which are consumed by various classes of animals in a beef cattle herd to give meat for human consumption. This beef production system consists of four interacting biophysical and bioeconomic subsystems, which are manipulated through the management subsystem in response to the climate subsystem (Fig. 12.1). The structure and significance of the various subsystems are described in more detail below.

The climate subsystem is largely outside the management subsystem but it directly affects the four subsystems influenced by a manager. For example, rainfall supplies soil water for plant growth, may cause soil erosion, and influences the rate of waste decomposition in soil. Further, prevailing temperature, humidity and radiation influence plant growth, and the incidence of plant and animal pests and diseases. Climatic inputs also display seasonal and

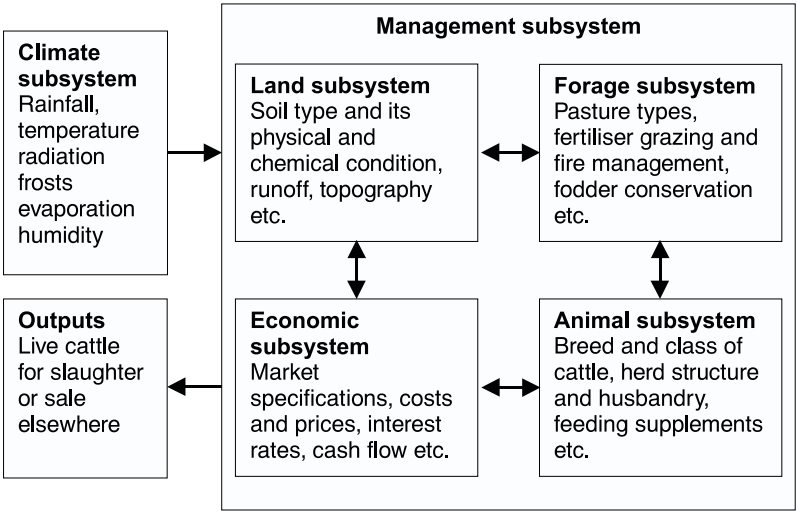


Fig. 12.1 Interrelationships between biophysical and bioeconomic subsystems (rectangles) with the management subsystem of the farmer. The biophysical and bioeconomic subsystems contain processes that determine their status. The interface between two subsystems (arrows) represents a conversion of materials into a new form. The manager is constantly responding to the climate subsystem, which impacts to varying degrees on the soil, pasture, animal and economic subsystems.

year-by-year variations and a manager must devise strategies to cope with these variations. Indeed, matching the farming system to the level and variability of climate inputs is a big challenge for a farm manager.[12] Seasonal variations in climate give rise to seasonal variations in quality and type of forage which may trigger fodder conservation (e.g. hay) to offset periods of forage deficiency. Wide year-by-year variations in climate inputs, often expressed as droughts or floods which lead to major perturbations in forage supply and market prices, need to be handled through skillful and resourceful management.[13] However, long-term weather forecasts now give managers prior warning of likely climatic extremes. For example, in northern Australia seasonal forecasts indicate the probability of rainfall in the forthcoming three to six months exceeding the historical median value, thereby permitting managers to make an early response to a likely distribution of rainfall.[14] Also extremely hot or cold temperatures can cause deaths in plants and animals, and computer models such as GRAZPLAN,[15] coupled to weekly weather forecasts, give early warning of likely mortalities in susceptible classes of animals. In both cases, recent improvements in the reliability and skill of weather forecasting are helping farmers to cope with wide variations in climate.

The land subsystem supplies water and nutrients for plant growth. Since it includes many of the ecological processes that sustain the whole system, both the manager and interest groups in the wider community are keen to keep the

land subsystem in good condition. Land degradation through soil erosion, desertification, salinisation, acidification and nutrient decline is a major concern in many of the world's grazing lands and has led to the notion of landscape management. With this approach, managers in a region with a common attribute, such as a river catchment, are encouraged to adopt strategies that enhance sustainable development rather than exploitation of the land subsystem. Landscape management also recognises that grazing lands produce food as well as ecosystem services, such as water and biodiversity that are needed to sustain the cities where most people live. Preferred management strategies for a landscape may arise through different management options being assessed by government agencies or local communities, and computer models are often useful tools in this process.[16]

Plants within the forage subsystem supply digestible nutrients when grazed by cattle. Forage accumulates through plant growth and forage not eaten, together with faeces and urine from cattle, return to the soil subsystem through the detritus food chain. The quality of forage on offer varies with the growing conditions and type of plant species in the system. New growth is the most digestible and there is a steady decline in quality as plant parts age, die and senesce. Since temperate grasses have a higher digestibility than tropical grasses, grazing systems in temperate zones tend to display higher animal performance than tropical zones, Leguminous species tend to have higher digestibility than gramineous species.[17] If a grazing system is based on sown pastures the manager may select to grow a mixed-pasture which usually consists of a few species that are well suited to a particular situation. This contrasts with native rangelands where the system consists of many different species, often including trees. Here a manager aims to keep the pasture in good condition by maintaining adequate plant cover to reduce soil erosion and a predominance of desirable rather than undesirable plant species.[18] In both sown pasture production systems and native rangelands, forage condition and animal performance can be manipulated by management options such as the choice of stocking rate, type and amount of fertiliser application, periods of grazing and conservation, level of supplementary feeding, and fire in the case of rangelands.[19, 20]

The cattle subsystem produces animals for sale through the processes of reproduction and growth within a herd consisting of different animal classes. The number of different animal classes on a farm largely depends on the quality of the pasture subsystem and on the objectives of a manager. In essence, breeding cows produce calves and after weaning these move into different classes as they grow and age (Table 12.1). Usually young female cattle (heifers) are selected to replace aged or culled cows and are mated for the first time when they reach maturity and a specific weight that depends on the breed and prevailing nutrition. Under good nutrition, heifers may be mated first at 15–18 months of age, but with the poorer nutrition in extensive rangelands, mating usually takes place at 24–30 months. Heifers that are not required for replacing cows might be sold for slaughter or for breeding purposes elsewhere. Male cattle are commonly castrated before weaning although a small number of high-

Table 12.1 Classes of cattle commonly found in beef cattle herds in extensive grazing systems. Adult equivalent, being the ratio of the energy requirement of a class to the energy requirement of an adult animal, is a coefficient for equating animal numbers in each class to a common base. Intensive grazing systems with a higher level of nutrition will have fewer classes since cattle are sold at a younger age.

Animal class	Adult equivalent	Age years	Comments
Cows and calves	1.3	2–12	Managers aim to have breeding cows calve annually. Calves are usually weaned at about 6 months of age.
Yearling heifers	0.55	0.5–1.5	Heifers are females that have not had one calf. When mature at 1.5 to 2.5 years, depending on breed and growing conditions, some are mated to replace culled cows. Surplus heifers may be sold for slaughter or as breeding stock.
2-year-old heifers	0.75	1.5–2.5	
Yearling steers	0.55	0.5–1.5	Steers, or castrated males, are sold for finishing elsewhere, or for slaughter. Age and weight at sale depends on the level of nutrition they experience, the specifications of available markets, and on the price advantage of different markets. Within limits set by prevailing climatic and economic conditions, a manager can target a specific market by manipulating feed supplies in the pasture subsytem.
2-year-old steers	0.8	1.5–2.5	
3-year-old steers	1.0	2.5–3.5	
4-year-old steers	1.1	3.5–4.5	
Culled cows	1.0	3–12	Cows no longer suitable for breeding due to age or infertility. Usually conditioned and sold for slaughter.
Bulls	1.1	3–7	Male animals for mating with cows. One bull is required for every 20 to 25 cows.

performing males may be retained to replace aged bulls. Depending on the prevailing nutrition and markets, male cattle may be retained for one to three years after weaning, to be sold for slaughter or for finishing elsewhere on another farm or in a feedlot. Thus, which market to target, and how the cattle should be fed to meet the market, are key strategic decisions for a manager. Deciding when to sell specific groups of cattle is a key tactical decision for a manager.

The different classes of cattle in a beef herd have different nutritional requirements because they differ in weight and age. The term adult equivalent (*AE*) relates the energy requirement of different classes to a common base, the energy requirement for maintenance of an adult animal, such as a non-lactating cow. The *AE*s of Table 12.1 can be determined from feeding tables but a first approximation for growing cattle is given by:

$$AE = LW^{0.75}/105.7 \qquad (12.1)$$

where LW and $LW^{0.75}$ are the liveweight and metabolic weight of animals in a specific class and 105.7 is the metabolic weight of a non-lactating bovine with a liveweight of 500 kg/head.[21]

The market subsystem refers to the different markets for beef cattle available to a manager along with the prices and profit margins associated with each market. Specifications for markets vary with location. In an extreme case there is no specification, and all cattle are sold as beef with no separation of cuts at retail outlets. At the other extreme, individual animals are prepared for a specific market and traced through the supply chain, with carcasses being graded for quality and various cuts of meat separated and sold at prices that reflect consumer preferences and the grade. Farmers in countries that export beef, such as USA, Australia, Canada and New Zealand, commonly have a range of market options that are specified in terms of age, gender, weight and fat thickness of a carcass. However, the classification scheme is not standardised internationally, although there is an international trend to reduce the allowable limits for residues of pesticide and growth promotants in export beef. Penalties for farmers in not meeting specifications for chemical residues are usually severe, including condemnation of all meat in the case of excess chemical residues.

12.3 Challenges for modellers

The above description of beef production is deceptively simple. In practice a model builder is faced with the challenge of expressing the complex interactions between components of the system (Fig. 12.1). Specific challenges include

- how to match the primary purpose of the model to the most appropriate structure
- how to handle natural variability in the biophysical components and the interface between the subsystems, and
- how to validate the completed model.

Answers to these questions are interrelated and reflect back to the history and philosophy of model building.

12.3.1 Matching purpose and structure

Models of beef production systems are commonly built as aids to research, farm management or policy evaluation and their structure may be mechanistic, empirical or a combination of both.[1] Empirical models estimate outputs by empirical equations developed from experimental observation of output in relation to one or more influencing variables, while mechanistic models reflect a theoretical understanding of the factors that control outputs. The relative merits of mechanistic and empirical structures have been hotly debated and the choice of structure is a critical and often difficult decision for a model builder.[2, 4, 22, 23] Mechanistic models, because of their stronger theoretical base, tend to be more

versatile and are more likely to explain responses than empirical models, but they may not be more accurate and often contain parameters that are difficult to determine in practical situations. Conversely, the robustness of an empirical model depends on the range of experimental data used in its derivation, and spurious results might occur if it is applied outside this range. Thus model builders should specify the derivation and application of an empirical model, and users should adhere to these specifications. As a variation on the above distinction, some models combine both empirical and mechanistic elements, such as an empirical model being used to process and interpret the results previously stored from many simulation experiments with a mechanistic model.

Research models are built by researchers to analyse the complex interactions in beef production systems. They can be regarded as a repository for past research since they collate and integrate information from past research. They are also a precursor for future research since gaps in knowledge and understanding are highlighted. Because research models focus on processes and their interactions, they are often mechanistic in structure and have a limited distribution. However, GRAZE is an exception to this statement, being a comprehensive mechanistic model of forage and animal growth that is widely distributed and well documented.[24] Sometimes a research model evolves into a management or policy model, thereby reducing development costs.

Models for farm management are usually designed to evaluate management options pertaining to one or more components of the system. They aid management by evaluating different scenarios thereby allowing preferred strategies to be identified, but importantly, a manager is free to accept or reject the output. Developing this type of model requires considerable time and effort, since to be accepted by potential users, the package needs to operate in a convenient and reliable manner, have a high degree of validity or skill, and have a commercial arrangement for distribution and after-sales service.[1] FEEDMAN[25] is an example of many commercial decision support systems that focus on farm management. However, history suggests that experienced farmers do not readily use such software for common routine decisions unless its use is clearly beneficial and it is promoted by a trusted product champion.[26–28] On the other hand, professional farm advisors who are paid to recommend preferred management options are likely to use the software to justify a recommendation. Because a farm advisor may have many clients, decision support software that is regularly used by a few farm advisors may still have a big impact on farm management. Both mechanistic and empirical sub-models are widely used in management software.

Policy models serve government or industry leaders by estimating outcomes to possible scenarios and initiatives in policy. Both mechanistic and empirical sub-models are used in policy models dealing with pasture and animal production. Policy models range from those that provide a one-off analysis of a specific problem to those that provide a regular ongoing service. An example of a one-off analysis that influenced policy was the rejection of a plan, based on results from field research over ten years, to construct farm dams and use the stored water to irrigate crops to improve the forage supply in north western Queensland.

Simulation studies based on long-term records of climate showed that the plan was not viable because rainfall was too variable.[29] Apparently the field study that supported the plan coincided with a run of high-rainfall years. An example of a regular ongoing service is the monthly maps of relative pasture yield, adjusted for prevailing stocking rates, which are derived from a pasture production model operating on a 5×5 km grid for the State of Queensland.[30] The maps provide an objective assessment of drought status for government and industry. Constructing and maintaining a policy model of this scale requires an integrated team of scientists, programmers and support staff. As with management models, a policy model's credibility depends on its scientific base and validity.

12.3.2 Coping with linkages between components

With regard to Fig. 12.1, the status of each subsystem is expressed by several different terms, which reflect the purpose of the overall model and the structure of the sub-models that simulate each subsystem. Since the subsystems are interdependent, they need to be linked in an appropriate manner, an issue in model building that is often called the interface problem. As an illustration, simple expressions of the status of each subsystem might be:

1. climate subsystem – inputs of solar radiation and/or temperature on plant growth and rainfall on soil water supply;
2. land subsystem – amount of soil water (mm) available for plant growth in response to daily rainfall runoff, drainage and evapotranspiration;
3. pasture subsystems – yield (kg/ha) of leaf and stem, potentially for each plant species in the pasture, in response to daily plant growth less consumption and senescence;
4. animal subsystem – liveweight (kg/head) of each animal class, in response to an initial liveweight and accumulated daily liveweight gain; and
5. economic subsystem – farm profit ($ or $/ha) in response to value of animals sold less variable costs.

Interface between climate, land and pasture subsystems

Mechanistic models often estimate plant growth as the product of intercepted solar radiation and radiation use efficiency. Intercepted radiation depends on leaf area of the forage, and radiation use efficiency links the soil and climate subsystems, being dependent on prevailing climate, soil nutrient status and soil water supply.[31] In practice, radiation interception and radiation use efficiency are difficult to simulate in pastures in rangelands that are a mixture of C3 and C4 species growing as spaced plants under trees in a semi-arid environment, and are grazed selectively by cattle. Under these complex circumstances an empirical model based on field observations can be a useful tool. For example, pasture growth (PG kg/ha) can be estimated as:

$$PG = WUE * WU \tag{12.2}$$

where *WUE* is water use efficiency, another term that links the two subsystems for a specified site (kg/mm), and *WU* is water use over a specified time step (e.g. mm/day).

Equation (12.2) avoids the difficulties associated with radiation interception by recognising the strong direct relationship between water use via transpiration and forage growth via photosynthesis, two gaseous transfer processes that are controlled by leaf stomata. It can be applied at different temporal and spatial scales.[32] On a daily time step, *WUE* becomes transpiration efficiency and *WU* is daily transpiration estimated by a sub-model of soil water balance, but on monthly or seasonal time step, *WUE* becomes rainfall use efficiency and effective rainfall (actual rainfall less runoff) is an approximation of *WU*. Although *WUE* varies with fertility status of the soil, seasonal conditions and the number of trees present, it is a parameter that can be determined simply for a site from measurements of plant growth in relation to *WU*. The FEEDMAN decision support system estimates monthly plant growth through this approach and the default values of *WUE* for many different soil-forage combinations were either obtained from field experiments or by integrating output from a daily plant growth model. In either case, the default values can be customised to reflect local conditions.

Interface between pasture and animal subsystems

This interface must account for nutritional demands of different classes of animals, all of which have the ability to move and select a preferred diet from a pasture that exhibits wide spatial and temporal variation in yield and quality.

In mechanistic terms, animal production is dependent on intake of digestible nutrients, and once the amount and quality of diet is known, models for estimating different forms of production (e.g. liveweight change, milk production, wool growth) in different animal classes already exist.[33] Thus the interface problem becomes how to estimate, either directly or indirectly, two interdependent terms, the amount (intake) and quality (digestibility) of diet. Actual intake is usually less than a potential intake, which depends on the breed and liveweight of animals, due to constraints arising from the amount and quality of forage on offer. Forage digestibility declines with age, is greater in leaf than stem, and varies across species. Mechanistic models commonly simulate diet selection by partitioning the forage on offer into digestibility or age categories with animals then selecting progressively from high to low categories until their appetite is satisfied.[34] Whilst this approach tends to mimic diet selection in temperate pastures reasonably well, the descriptive functions are essentially empirical relationships derived from field experiments. The approach has been less successful in rangelands with a more heterogeneous botanical composition and sward structure.[35] However, a more realistic algorithm for diet selection in heterogeneous forages places plant species into broad preference categories (e.g. preferred, desirable, undesirable, toxic, emergency and non-consumed) and then computes the proportion of each preference class in the

diet.[36] The algorithm assumes that an animal has experience with the vegetation, and has learned to avoid toxic species and non-consumed species. The 'emergency' category accounts for species that are only eaten after the preferred, desirable and undesirable species are depleted.

The above 'mechanistic' models are essentially based on 'empirical' expressions derived from diet selection studies with parameters that are rather abstract and site specific. To avoid these difficulties, the FEEDMAN package used the notion of potential liveweight gain to characterise the seasonal variation quality of different forages. Potential liveweight gain is the monthly liveweight gain of a standard animal (a 200 kg cross-bred steer, *Bos taurus* by *Bos indicus*) grazing the forage at a low stocking rate in a good season. It is a bioassay for forage quality that can be measured, but more importantly, it is meaningful to farmers and can be adjusted to reflect local experience and knowledge. With potential liveweight gain for a standard animal given, the energy concentration of the forage can be estimated and applied to different animal classes, after taking account of the impact of high stocking rate on reducing intake and dry conditions reducing forage quality.[25] Because this approach uses a bioassay to characterise forage quality, and a mechanistic model to estimate animal performance, it can be readily adapted to herds of different species, breeds and classes of livestock.

Interface between animal and economic subsystems

Operating profit of a beef cattle enterprise on a farm is given by:

$$Gross\_profit = Number\_sold \times (Animal\_value - Variable\_costs)(\$) \quad (12.3)$$

where *Number_sold* is the number of animals sold, *Animal_value* is the average value of sale animals, *Variable_costs* are average variable or operating costs per animal associated with different management options. Comparison of the gross profit for different management options indicates the relative profitability of the options.

Estimation of *Variable_costs* is a simple arithmetic exercise, but since there is wide variation in local costs, a model must allow a user to modify and recall this information, and a user must update the information as required. On the other hand, estimation of *Animal_value* is a two-step process where animals are first allocated to a market category (if more than one exists), each with a corresponding sale price that usually exhibits spatial and temporal variation. Thus, tables of market prices for use in the calculation of *Animal_value* need to be updated regularly. The determination of market categories is location specific since there is wide national and international variation in the title and specifications for each category. In countries with well developed beef markets, categories may be specified by age, sex and breed of cattle, by weight expressed as liveweight or carcass weight, and by an indication of the degree of 'finish' expressed as a condition score in live cattle or fat thickness for carcasses. However, markets are not necessarily mutually exclusive in that while a

premium market may have narrow specifications, cattle suited to a premium market may also be suited to a lower-priced market with wider specifications. Mechanistic models attempt to estimate animal growth and development, and the associated fat deposition.[37, 38] Condition score has been derived empirically from the history and status of animal performance,[39] but neither approach has been applied to a full range of market specifications. FEEDMAN uses a simple approach to estimate *Animal_value* in that the characteristics of each herd are compared against entries in a table of markets, specified in terms of monthly sale price, and breed, age, class and liveweight of cattle. The highest price match is then selected and used to calculate *Animal_value*.

12.3.3 Coping with natural variability

On-farm complexity

Creating a 'user friendly' presentation of software that mimics pasture and animal production on a farm is a challenge because a multi-dimensional scenario must be described through a keyboard and monitor. The multi-dimensional scenario might consist of descriptions of fields in the farm, pastures in the fields, number and class of animals in herds, grazing management of herds, and period, type, and amount of supplementary feeding (Fig. 12.1). In addition, potential users commonly prefer the software to have keystrokes and a screen layout similar to other familiar software. Also, outputs must be clear, easily understood, and suitable for further analysis or storage. One approach used by model builders to meet these requirements is to consult with a panel of potential users on a regular basis and progressively modify the software in response to suggestions from the panel.[11] Such 'interactive prototyping' is a time-consuming task that can lead to major changes in the layout of screens for entering data and displaying results, but experience has shown that model builders, who know a package intimately, are not experts in 'user friendly' presentations. In practice, there are tradeoffs between the capacity of a decision support package to handle wide variations in farm production systems and the need for the package to be 'user friendly'. Extensive help notes, default values for input parameters, and training exercises and examples all assist a novice user in mastering a package. In addition to complexity due to on-farm variations mentioned above, climate and prices are off-farm inputs that display wide spatial and temporal variations.

Climate

In the case of climate a user may wish to evaluate management options over a range of seasonal conditions contained in historical records of climate. One approach is to use all historical data as an input and then express key outputs, such as farm profit, as a probability distribution. Another approach is to use a probability distribution of historical annual rainfall to establish categories of 'seasons' that reflect natural variations, such as:

very dry, rainfall likely to be less than this category in 10% of years;
dry, rainfall likely to be less than this category in 30% of years;
median, rainfall likely to be less than this category in 50% of years;
wet, rainfall likely to be less than this category in 70% of years; and
very wet, rainfall likely to be less than this category in 90% of years.

The former approach demands access to a large database of historical records of climate, particularly if a model is to apply to a wide range of locations, each with a different climate history. The second approach, to select from the same comprehensive database a relatively small number of typical climate categories for each location, thereby eliminates the need for regular access to a large database of historical records. Both approaches are an attempt to assess management options simulated by the model in terms of the risk or likelihood of certain outcomes. This is a key attribute of models of beef production in variable climates, which is not obtained by using average or median climate data. Indeed, if only median climate data is used, animal production at high stocking rates is overestimated because year-by-year variations and interactions are ignored.[8]

In addition to analysing historical records of climate, model users are frequently interested in evaluating management options in relation to the current status of cattle and forage on a farm and future climate scenarios that are based on long-term weather forecasts.[13] Currently long-term weather forecasts indicate the probability of rainfall in the next three or six months being above or below median rainfall, and the skill of the forecasts is improving.[40] To cater for this requirement, models must allow users to enter potential future rainfall.

12.3.4 Verification and validation

Model verification ensures that the computer programs on which a model is based are free of 'bugs' and perform properly within specific limits. Usually a model builder uses special input data and parameters to test components of a model and their interactions under a wide range of operating conditions. The program needs to be corrected if values of the various variables and processes exceed an acceptable range. Problems may arise from a flaw in the algorithm describing a process, particularly as upper or lower limits are approached, or from a typing error in the program code. A sensitivity analysis is another component of verification that indicates the relative importance of accuracy in model inputs. Here a simulation experiment is designed to test the relative sensitivity of inputs and parameters that influence a system. Obviously accuracy is more important with sensitive than with insensitive inputs. The relative sensitivity of different inputs is indicated by comparing the change in output caused by a specific change in the different inputs (e.g. percent change in output after a 5, 10 or 20% change in an input parameter). Whilst verification is primarily the responsibility of model builders, simple exercises on these lines give model users a good appreciation of the operation and limitations of a model.

Model validation refers to how well a model mimics the system it is meant to represent. Validation is commonly demonstrated by first instructing a model to mimic a wide range of scenarios that have been actually observed, and then by comparing predictions from a model against the observations. The validation data should be independent of the data used in developing a model. Linear regressions of observations against predictions are commonly used to make the comparisons. The closer the slope and coefficient of determination for a regression are to unity, and the intercept to zero, the better the validity of a model. However, there are theoretical and practical problems with validation based on regression analysis,[41] and the confidence of the model builders should be recognised as a model undergoes development and modification.[42,43] Of course, serious users also develop confidence in a model through less formal validations as they compare predictions against their own observations and experiences. In practice, validation is an ongoing activity that warrants considerable effort by the model builders and independent experts, particularly when the model attempts to mimic large variation in production systems and is used as an aid to politically or financially sensitive decisions.[44] In essence a model is 'valid' when it sufficiently mimics the real world to fulfill its objectives, and when decisions based on the model are superior to those made without the model.[45]

12.4 Simple model of herd structure

It is obvious from Fig. 12.1 and Table 12.1 that for a given farm, the number and class of cattle in the animal subsystem depends on the amount and quality of growth in the forage subsystem. These interactions are captured in the following simple empirical model of herd structure in relation to broad management options. It also illustrates how a model that incorporates a few basic parameters can be a powerful analytical tool.

The notion of farm carrying capacity (*CC*) is a good starting point. This is the long-term safe stocking rate for a farm, one that does not cause ecological deterioration of the production system. It is a vital concept for managed grazing systems that incorporate the biological, commercial and social elements pertaining to good land care. It is commonly used to quantify a farm for sale or leasing in Australia and the USA, and because different classes of cattle have different nutritional requirements, it is commonly expressed as adult equivalents (see equation (12.1)).

In rangelands where forage growth is dependent on rainfall, carrying capacity is largely dependent on the amount of forage growth and on the proportion of growth that can be eaten (utilisation, *U*) without causing degradation of the pasture. Thus, based on the report by Johnston *et al.*[32]

$$CC = R * WUE * A * U/I \quad (AE) \tag{12.4}$$

where *CC* is farm carrying capacity, *R* is effective rainfall (mm/year, in subtropical climates this is annual rainfall less runoff), *WUE* is water use

efficiency (e.g. 5 kg/ha/mm), A is area of the farm (ha), U is safe utilisation (e.g. 0.25) and I is annual intake for an adult animal (e.g. 4000 kg/year). Whilst WUE varies with the inherent fertility of the soil, fertiliser applications and presence of trees, it is simple to measure. On the other hand U is not simply measured but studies have shown it ranges from about 0.1 in arid infertile environments to about 0.5 in moist fertile environments. Although equation (12.4) demonstrates the derivation of CC from first principles, in practice farm CC is usually determined from local knowledge and experience.[32] The next task is to determine herd structure or the distribution of carrying capacity across the various animal classes.

When all cattle on a farm originate from the breeding cows (i.e. no off-farm purchases) the system is characterised by three performance indicators, which underpin a simple but versatile mathematical model of herd structure.

(1) Weaning rates refer to the number of calves weaned per hundred cows mated. This key indicator depends on the nutritional health status of cows and on the number and fertility of bulls. It commonly ranges from 95% in high-performing herds to less than 50% in herds of poor performance, a value that will not sustain the herd in the long term.
(2) Survival rates refer to the proportion of each class of cattle that survive a year. Mortality from poor health, accident or predators is common, particularly in extensively-managed beef production systems. The animal classes most prone to mortality are breeding cows and calves soon after weaning. Clearly high survival rates are desirable and susceptible classes of cattle commonly receive special feeding to avoid mortality from poor nutrition.
(3) Culling rates refer to the proportion of breeding cows culled annually for age, infertility, or other imperfections. Hence, if the effective breeding life of a beef cow is about ten years, culling helps to maintain high weaning rates. The rate of culling, plus the mortality of breeding cows defines the number of replacement heifers required to maintain a constant number of breeding cows.

The following model, which is suitable for a spreadsheet, provides a 'steady state' estimate of number in the various classes of cattle in a herd (herd structure, Table 12.1), in response to a few key assumptions and parameters, and local knowledge of performance criteria. The model depends on four assumptions.[46] First, all animal classes on a farm with breeding and growing cattle can be specified by a manager, and are related numerically to the number of cows mated, provided extra animals are not purchased. Second, the overall carrying capacity (CC) of a farm, in terms of number of adult equivalents, is either known or can be estimated by equation (12.4). Third, for simplicity, cows and calves are regarded as a single animal class until the calves are weaned. Fourth, the number of cows mated (CM) is fixed for each situation because if one dies or is culled from the breeding herd it is replaced with a heifer. Thus the 'n' classes of cattle on a farm can be represented as

$$CC = A1 \times CM + A2 \times CM + A3 \times CM + \ldots\ldots\ldots An \times CM \qquad (12.5)$$

and after collection of terms and simplification

$$CM = CC / \sum Ai \qquad (12.6)$$

where *Ai* is a coefficient that relates the number of animals in the '*i*' th class of cattle to *CM*, the numbers of cows mated. *Ai* is the product of four factors:

$$Ai = PFi \times CFi \times SRi \times BRi \qquad (12.7)$$

where *PFi* is a flag to indicate if the *i* th class of animal is present (1, present; 0, absent); *CFi* is a factor to convert the *i* th class of animal to adult equivalents (Table 12.1); *SRi* is the proportion of the original number surviving in the i th class; and *BRi* is the ratio of the number of animals in the *i* th class to the number of breeders when survival in the class is 100%.

WR is weaning rate, expressed as a percentage of the number of calves weaned to number of cows mated. If half the weaners are assumed to be female, it follows that $BRi = WR/2$ for each class of steers in the herd, and for heifer cattle *BRi* is similar to steers until replacement heifers enter the breeding herd.

Replacement heifers enter the breeding herd when two or three years of age by adjusting *PFi* accordingly. First dead cows are replaced (*DEATHS* = percentage of *CM* dying each year), then culled cows are replaced according to a specified culling policy (*CULL* = preferred percentage of *CM* replaced each year). If there are too few heifers for the culling policy, all available heifers are used as replacements and the shortfall is noted by the lack of surplus heifers for subsequent sale and a reduced ratio for culling. If there are too few heifers to replace the dead cows the herd cannot be sustained. Thus for culled cows:

$$BR_{cull\ cows} = MAX(0, MIN(CULL, WR/2 - DEATHS)/100) \qquad (12.8)$$

and for any surplus females

$$BR_{surplus\ females} = MAX(0, (WR/2 - CULL - DEATHS)/100) \qquad (12.9)$$

Once the number of cows mated have been calculated using equation (12.6), the number of cattle in the remaining animal classes is given by

$$Ni = CM * PFi * SFi * Bri \qquad (12.10)$$

where $i > 1$ since for cows, being class 1, $Ni = CM$.

Table 12.2 illustrates the application of this model to four scenarios pertaining to breeding and growing beef cattle on extensive rangelands. Case 1 represents a herd where disease and/or poor nutrition severely restricts performance of the breeding herd and this limitation is removed in Case 2. Case 3 is similar to Case 2 except for a 50% increase in farm carrying capacity, which might occur through farm development options such as buying more land, controlling woody weeds or sowing improved pasture. Case 4 illustrates the effects on herd structure of a further improvement in performance of breeding cows along with a reduction in age of selling steers and mating heifers, as might

Table 12.2 Herd structures generated by the simple model given above in response to changes in key parameters that might occur as heath, nutrition and management improves in a 'closed' herd consisting of breeding and growing cattle on extensive rangeland.

Key parameters	Case 1	Case 2	Case 3	Case 4
Farm carrying capacity (*CC*) adult equivalents	1000	1000	1500	1500
Weaning ratio (*WR*) (% of cows mated)	50	80	80	90
Cow mortality rate (*DEATHS*) (%)	15	5	5	3
Ideal culling ratio for cows (*CULL*) (%)	20	20	20	20
Age of steers at sale: years	4	4	4	3
Age of surplus heifers at sale: years	3	3	3	2
Simulated results				
Total number of cattle in herd	1088	1121	1682	1700
Number of breeding cows	421	303	455	532
Proportion of herd as breeding cows (%)	39	27	27	31
Number of culled cows	42	61	91	106
Proportion of breeding cows culled (%)	10	20	20	20
Number of surplus heifers sold	0	44	66	112
Number of steers sold	99	114	171	227
Total number of cattle sold	141	219	328	446
Proportion of sale cattle in herd (%)	13	20	20	26

occur from a further improvement in herd nutrition and management. Whilst Table 12.2 is a static representation that ignores the transitional states that would occur when changing from Case 1 to Case 4, it shows the broad implications of management options on herd structure and number of cattle for sale. It also illustrates that simple 'spreadsheet' models can be a useful first step in selecting broad management options that warrant a more detailed evaluation.

12.5 Future developments

Modelling pasture and animal production has come a long way in three decades. Its future as an aid to research is assured since it provides direction and context to research programs.

While farmers have been slow to adopt decision support packages that aid routine decisions, professional advisors who need to give good advice to many clients are more receptive to new tools that assist in evaluating management options within complex systems across a wide range of environments. Future developers of farm management models will probably regard farm advisors or service agencies rather than farmers as the primary customers. Also, the models will be more user-friendly through the use of improved graphics and visualisation techniques, and the provision of support and upgrades via the World Wide Web.

The scope and range of policy models are expanding rapidly because they provide policy makers with an objective assessment of complex problems. This

trend will continue, but policy models are likely to expand from the traditional biophysical base to include socioeconomic components and estimates of the impact of policies on the 'triple bottom line' – ecological sustainability, profitability and social acceptability.[47–49] Indeed, a future challenge will be how to better integrate the technologies pertaining to hard and soft systems, such as pasture and animal production models being part of participatory action research, and thereby involving stakeholders in defining and evaluating policies.[16, 50]

A global network of information for model development and proven software modules is expanding through the World Wide Web. Model developers will have increasing access to libraries of algorithms, and computer operating environments which will encourage more rapid development of new models and a rich set of shared applications and experiences. However, since models are repositories for information and results from past research, there remains a global need for scientists and government agencies to organise creditable databases of information, which are critical to the future development of decision support systems and integrated policy models.[51]

12.6 References

1. RICKERT K G, STUTH J W, MCKEON G M, 'Modelling pasture and animal production'. In *Field and Laboratory Methods for Grassland and Animal Research*, L T Mannetje and R M Jones (Eds), pp. 29–65, Wallingford, CAB Publishing, 2000.
2. HERRERO M, DENT J B, FAWCETT R H, 'The plant/animal interface in models of grazing systems'. In *Agricultural Systems Modeling and Simulation*, R M Peart and R B Curry (Eds), pp. 495–542, Marcel Dekker Inc, New York, 1998.
3. EBERSOHN J P, 'A commentary on systems studies in agriculture'. *Agric Syst*, 1976 **1**(3) 173–84.
4. BLACK J L, DAVIES G T, FLEMING J F, 'Role of computer simulation in the application of knowledge to animal industries'. *Aust J Agric Res*, 1993 **44**(3) 541–55.
5. JONES R J, SANDLAND R L, 'The relation between animal gain and stocking rate. Derivation of the relation from the results of grazing trials'. *J Agric Sci UK*, 1974 **83**(2) 335–42.
6. RICKERT K G, 'Stocking rate and sustainable grazing systems'. In *Grassland Science in Perspective*, A Elgersma, P C Stuik and L J G van de Maesen (Eds), Wageningen Agricultural University Papers 96.4, pp. 29–63. Wageningen Agricultural University, Wageningen, 1996.
7. CONWAY G R, 'Sustainability in agricultural development: trade-offs between productivity, stability, and equitability'. *J Farm Syst Res Ext*, 1994 **4**(2) 1–14.
8. RICKERT K G, MCKEON G M, 'Models for native pasture management and development in south east Queensland'. In *Ecology and Management of*

the World's Savannas., J.C. Tothill and J.C. Mott (Eds), pp. 299–302, 1985.

9. MCKEON G M, DAY K A, HOWDEN S M, MOTT J J, ORR D M, SCATTINI W J, WESTON E J, 'Northern Australian savannas: management for pastoral production'. *J Biogeogr*, 1990 **17**(4–5) 355–72.
10. HALL W H, Near-real Time Finacial Assessment of the Queensland Wool Industry on a Regional Basis. PhD Thesis, University of Queensland, 1997.
11. STUTH J W, HAMILTON W T, CONNER J C, SHEEHY D P, BAKER M J, 'Decision support systems in the transfer of grassland technology'. In *Grasslands for our World*, M J Baker (Ed.), pp. 234–42, SIR Publishing, Wellington, 1993.
12. LANDSBERG R G, ASH A J, SHEPHERD R K, MCKEON G M, 'Learning from history to survive in the future: management evolution on Trafalgar Station, North-East Queensland'. *Rangel J*, 1998 **20**(1) 104–18.
13. JOHNSTON P W, MCKEON G M, BUXTON R, COBON D H, DAY K A, HALL W B, QUIRK M F, SCANLAN J C, 'Managing climate variability in Queenslands's grazing lands – new approaches'. In *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems – the Australian Experience*, G Hammer, N Nicholls and C Mitchell (Eds), Kluwer Academic Press, Amsterdam, 2000.
14. MCKEON G M, ASH A J, HALL W B, STAFFORD SMITH D M, 'Simulation of grazing strategies for beef production in north-east Queensland'. In *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems – the Australian Experience*, G Hammer, N Nicholls and C Mitchell (Eds), Kluwer Academic Press, Amsterdam, 2000.
15. DONNELLY J R, MOORE A D, FREER M, 'GRAZPLAN: decision support systems for Australian grazing enterprises–I. Overview of the GRAZPLAN project, and a description of the MetAccess and LambAlive DSS.' *Agric Syst*, 1997 **54**(1) 57–76.
16. ABEL N, 'Resilient rangeland regions'. In *People and Rangelands: Building the Future*, D Eldridge and D Freudenberger, (Eds), *Proceedings of the VI International Rangeland Congress, Townsville, Queensland, Australia, 19–23 July, 1999* **1**, 21–30.
17. NORTON B W, 'Differences between species in forage quality. Temperate and tropical legumes and grasses, digestibility'. In *Nutritional Limits to Animal Production from Pastures*, J B Hacker (Ed.), Farnham Royal, Commonwealth Agricultural Bureaux, 1982.
18. BROWN J R, 'State and transition models for rangelands. 2. Ecology as a basis for rangeland management: performance criteria for testing models'. *Trop Grassl*, 1994 **28**(4) 206–13.
19. HODGSON J G, *Grazing Management. Science into Practice*, Harlow, Longman, 1990.
20. HEITSCHMIDT R K, WALKER J W, 'Grazing management: technology for sustaining rangeland ecosystems?'. *Rangel J*, 1996 **18**(2) 194–215.

21. MINSON DJ, WHITEMAN PC, 'A standard livestock unit (SLU) for defining stocking rate in grazing studies'. *Proceedings of the XVI International Grassland Congress, 4–11 October 1989, Nice, France, 1989, 1117–18*, 1989.
22. SELIGMAN NG, BAKER MJ, 'Modelling as a tool for grassland science progress.' In *Grasslands for our World*, M J Baker, (Ed.), pp. 228–33, SIR Publishing, Wellington, 1993.
23. MONTEITH JL, 'The quest for balance in crop modeling'. *Agron J*, 1996 **88**(5) 695–7.
24. LOEWER OJ, 'GRAZE: a beef-forage model of selective grazing'. In *Agricultural Systems Modeling and Simulation*, R M Peart and R B Curry (Eds), pp. 301–417, New York, Marcel Dekker Inc., 1998.
25. SINCLAIR SE, RICKERT KG, PRITCHARD JR, FEEDMAN – A feed-to-dollars beef and deer management package, QZ00004, Brisbane, Queensland Department of Primary Industries, 2000.
26. COX PG, 'Some issues in the design of agricultural decision support systems'. *Agric Syst*, 1996 **52**(2) 355–81.
27. BUXTON R, STAFFORD-SMITH MD, 'Managing drought in Australia's rangelands: four weddings and a funeral'. *Rangel J*, 1996 **18**(2) 292–308.
28. RICKERT KG, 'Experiences with FEEDMAN, a decision support package for beef cattle producers in south eastern Queensland'. *Acta Hort*, 1998 **476** 227–34.
29. CLEWETT JF, *Shallow Storage Irrigation for Sorghum Production in North-west Queensland*. Bulletin QB85002, Brisbane, Queensland Department of Primary Industries, 1985.
30. HASSETT RC, WOOD HL, CARTER JO, DANAHER TJ, 'Statewide ground-truthing of a spatial model'. In *People and Rangelands: Building the Future*, D Eldridge and D Freudenberger (Eds), *Proceedings of the VI International Rangeland Congress, Townsville, Queensland, Australia, 19–23 July, 1999*, **2**, 763–4.
31. HAMMER GL, WRIGHT GC, 'A theoretical analysis of nitrogen and radiation effects on radiation use efficiency in peanut'. *Aust J Agric Res*, 1994 **45**(3) 575–89.
32. JOHNSTON PW, MCKEON GM, DAY KA, 'Objective ''safe'' grazing capacities for south-west Queensland, Australia: development of a model for individual properties'. *Rangel J*, 1996 **18**(2) 244–58.
33. FREER M, MOORE AD, DONNELLY JR, 'GRAZPLAN: decision support systems for Australian grazing enterprises – II. The animal biology model for feed intake, production and reproduction and the GrazFeed DSS'. *Agric Syst*, 1997 **54**(1) 77–126.
34. DOVE H, 'Constraints to the modelling of diet selection and intake in the grazing ruminant'. *Aust J Agric Res*, 1996 **47**(2) 257–75.
35. HALL WB, RICKERT KG, MCKEON GM, CARTER JO, 'Simulation studies of nitrogen concentration in the diet of sheep grazing Mitchell and mulga grasslands in western Queensland'. *Aust J Agric Res*, 2000 **51** 163–72.

36. QUIRK M F, STUTH J W, WEST N E, 'Verification of POPMIX preference algorithms for estimating diet composition of livestock'. In *Rangelands in a Sustainable Biosphere*, Denver, Society for Range Management, 1996.
37. SCA, *Feeding Standards for Australian Livestock: Ruminants. Standing Committee on Agriculture*, East Melbourne, CSIRO Publications, 1990.
38. WILLIAMS C B, JENKINS G M, 'Predicting empty body composition and composition of empty body weight changes in mature cattle'. *Agric Syst*, 1997 **53**(1) 1–25.
39. RICKERT K G, MCKEON G M, 'A computer model of the integration of forage options for beef production'. *Proc Aust Soc An Prod*, 1984 **15** 15–19.
40. DAY K A, AHERNS D G, PEACOCK A, RICKERT K G, MCKEON G M, 'Climate tools for northern grassy landscapes.' *Proceedings of Northern Grassy Landscapes Conference Katherine, 29–31 August 2000*, Darwin, Northern Territory University, 2000.
41. MITCHELL P L, 'Misuse of regression for empirical validation of models'. *Agric Syst*, 1997 **54**(3) 313–26.
42. HARRISON S R, 'Regression of a model of real-system output: an invalid test of model validity'. *Agric Syst*, 1990 **34**(3) 183–90.
43. HARRISON S R, 'Validation of agricultural expert systems'. *Agric Syst*, 1991 **35**(3) 265–85.
44. SCANLAN J C, MCKEON G M, DAY K A, MOTT J J, HINTON A W, 'Estimating safe carrying capacities of extensive cattle-grazing properties within tropical, semi-arid woodlands of north-eastern Australia'. *Rangel J*, 1994 **16**(1) 64–76.
45. DENT J B, EDWARD JONES G, MCGREGOR M J, 'Simulation of ecological, social and economic factors in agricultural systems'. *Agric Syst*, 1995 **49**(4) 337–51.
46. RICKERT K G, ESPIE N J, STOCKUP: a program to assess the structure of beef cattle herds, Brisbane, Department of Primary Industries, 1990.
47. PANDEY S, HARDAKER J B, 'The role of modelling in the quest for sustainable farming systems'. *Agric Syst*, 1995 **47**(4) 439–50.
48. BELLAMY J A, LOWES D, ASH A J, MCIVOR J G, MACLEOD N D, 'A decision support approach to sustainable grazing management for spatially heterogeneous rangeland paddocks'. *Rangel J*, 1996 **18**(2) 370–91.
49. ABEL N, ROSS H, WALKER P, 'Mental models in rangeland research, communication and management'. *Rangel J*, 1998 **20**(1) 77–91.
50. PARK J, SEATON R A F, 'Integrative research and sustainable agriculture'. *Agric Syst*, 1996 **50**(1) 81–100.
51. BESWICK A, JEFFREY S J, MOODIE K B, 'Climate databases: fundamental to the effective management of Australia's rangelands'. In *People and Rangelands: Building the Future*, D Eldridge and D Freudenberger (Eds), *Proceedings of the VI International Rangeland Congress, Townsville, Queensland, Australia, 19-23 July, 1999*, **2**, 850–1.

Part IV

Applications: processing technologies

Introduction

Mathematical models have been used for a long time in food process design to ensure microbiologically safe and tasty food. In fact, sterilization processes are one of the few production processes in industry which rely on a mathematical model to ensure the safety of the process. Although initially most models were empirical, gradually this has changed and many generic models are now used in practice to improve food processes. The advent of cheap high speed computers has been substantial in furthering progress in this area given the often high complexity of the models. In this part it will be shown how models can be used successfully in a wide range of processing applications including fermentation technology, modified atmosphere package design, and thermal processing including heating, cooling and freezing of foods.

In Chapter 13, model-based design of fermentation processes will be discussed. This involves the combination of a wide range of models describing growth and inactivation of micro-organisms, transport phenomena, down-stream unit operations and, last but not least, economic models to calculate the profit of the operation.

Modified atmosphere packaging is now widely used to extend the shelf life of various foods. Whereas still often a purely experimental approach is taken to select the packaging foil, the advantages of using mathematical models are now well recognized. In Chapter 14 an overview will be given of how conceptual models can be used to improve modified atmosphere packaging processes.

Thermal processes are certainly amongst the most important processes in the food processing industry, and the state of the art in modelling of these processes is reviewed in Chapter 15. The key issue in most thermal processes is to achieve

sufficient inactivation of micro-organisms with as little damage to other quality attributes of the food as possible. For some applications, such as sterilization of conduction heated foods, an appropriate generic modelling methodology is available. However, in most cases the heat transfer processes are quite complicated, and although the governing models are known well, their numerical solution remains a challenging task.

Cooling and freezing operations are widely used in the food industry, and the prediction of cooling or freezing time, product temperature, heat load and moisture loss is very important for a proper process design. Numerous empirical and semi-empirical formulas have been suggested in the literature to predict these process variables, but more complicated generic models are now available. An extensive overview of these models will be given in Chapter 16.

Bart Nicolaï

13

The use of models in process development: the case of fermentation processes

J. J. M. Hofmeester, DSM Food Specialties, Moorebank

13.1 Introduction

In the food industry, models are used for exactly the same purpose as in the scientific world. The most important reason to use models is to structure ideas and concepts to be able to prove or disprove those ideas. In this way knowledge is developed about the relevant processes.

The practical application of models in industrial processes is very important for the development of new processes, and for communication between people with different backgrounds. For this reason, models of very different levels of complexity are used. Some models are used to describe qualitatively the situation in the process, without too much accuracy. Other models describe in great detail the situation, and are used for in-depth study of the process.

This chapter describes the application of models using the development of processes in the fermentation industry as an example. Using this example the possible types of models are demonstrated. This example is chosen because fermentation is the most important production technology at DSM Food Specialties, and it shows how models can be used across a multi-disciplinary team. The description of a fermentation process will always be a rough simplification of reality, since detailed knowledge about what exactly happens in a cell under fermentation conditions is not available, as fermentation processes are generally too small to justify the research required to develop this knowledge. This means that complete knowledge of the process will probably never become available.

The development of models is probably the most important part of the work people do in research and development, because models help them to think about processes that are too complicated to understand in every detail.

Simplifying reality in models will help to focus on the relevant issues of the process, and to communicate among co-workers. Apart from that it is very important to develop knowledge that is more generally applicable than to just one single fermentation process. Although computers make it possible nowadays to develop very complex models, in the end people must be able to translate the results of those models into practical changes in the process.

13.2 What is a fermentation process?

For those people that are not familiar with fermentation processes, typical properties of the process are given in this brief explanation.

13.2.1 Types of products produced

A wide range of products is made using fermentation processes. Traditionally beer, wine and cheese are important products in which fermentation plays a major role. This chapter will focus on single strain processes as used for the production of antibiotics and enzymes.

13.2.2 The microorganism

The core of the process is the microorganism that will produce the desired product. The production capacity of that microorganism will ultimately determine whether a certain process can be performed economically. For that reason the selection of the strain is very important, and is normally one of the most important parts of a development process. Traditionally, only fungi, yeasts and bacteria were used in these processes. Nowadays also cells from other organisms are used for production.

13.2.3 The medium

Since the microorganism must produce the product, it is important that all the organism's needs for growth and production are met. This means that not just all the desired nutrients must be available, also the temperature, pH, etc., must be controlled. Since some organisms may have very special needs, the development of media can be very laborious. Media normally contain a carbon and nitrogen source, trace elements and vitamins. To avoid the growth of undesired organisms, the medium and the equipment are sterilized before the process starts.

13.2.4 Control strategies

To be able to produce large amounts of products, the concentration of organisms and product must be as high as possible. To achieve that, the environment of the organism must be controlled. This means that corrective agents are added to

control pH, the temperature is controlled by cooling, and depleted medium components are added separately.

13.2.5 Scale of the process

Products with high added value (e.g. vaccines) are sometimes made in small fermentors several litres in volume. Large, low added value products must be made in much bigger volumes to be able to use the economy of scale to get the cost price down. These products are normally made in fermentors up to several hundred cubic metres in volume (see Fig. 13.1).

13.2.6 Organization

Development of a fermentation process requires the cooperation of different disciplines, e.g. biochemistry (enzymology), molecular biology (genetics),

Fig. 13.1 Size and geometry of a production fermentor.

microbiology, reaction technology (fermentation), downstream processing (recovery) and, last but not least, integration with economic evaluation from the very beginning. For optimal results, these people have to work closely together during the development of a new process. The use of models that simplify reality is an essential tool that allows people to contribute in fields that are outside their main expertise.

13.3 Models used during process development

13.3.1 Biological models

Genetic models

Genetic models are used to describe processes applied during genetic modification of genes. These models are not mathematical models, but rather verbal descriptions of the processes. The use of these verbal models is necessary to enable discussion of the processes between people familiar with the work as well as between people from different backgrounds. The most obvious model is the use of the words 'cut' and 'paste' for changing the positions of genes. This description of the process is a very rough simplification of a complex enzymatic treatment of the DNA.

The disadvantage of such a simplification is, of course, that it does not say much about the actual process. The (very large) advantage of such a model is that it can be used to explain the process to outsiders in terms that everybody can understand. Today many different ways are used to visualize this transformation of genes. These visualizations are models that are used to describe the process qualitatively. These models cannot be used for quantitative descriptions.

Another 'model' that should be mentioned here is also not a mathematical model, but has been proven to be very useful in the past. This is the model for the spatial structure of DNA, represented by the double helix. This image made it possible for everybody to understand that structure, without the necessity of long biochemical training.

Metabolic pathway engineering

A fairly recent development is the use of metabolic pathway engineering for the optimization of processes. This technique requires detailed knowledge of the metabolic pathway that leads to the final product. Not only qualitative knowledge is required, but also quantitative knowledge is being put into the model. This leads to estimates of the fluxes of all the steps involved, and finally it will give information on the rate-limiting step of the production process. Normally a lot of reactions are involved in the production of a certain product. Also, the rates of the reactions involved are determined by concentrations of substrates, precursors and enzymes. Modelling such a process takes much more than just the mathematical development of a model. The model will only be useful if:

- Proper parameter estimation can be done. This means that knowledge has to be acquired on all the reactions involved. Estimating the reaction rates can do this, but also a lot of high-level biochemical analyses will be necessary to determine concentrations, enzyme activities, etc.
- The tools are available to make controlled changes in the cell to remove bottlenecks for production.

Because of the last reason this technique was only developed when genetic tools became available to change the genetics of the organism. The model will give information on how to change the genes of an organism for optimal results.

This new tool has already proven to be very successful. It is used for the optimization of existing processes, and also for the development of processes that could not be done without this technique. A major disadvantage of this tool is that it cannot be used easily. High-level knowledge and experience in modelling, genetics and biochemistry are essential to do this successfully. Because of the complexity of this tool, one of the pitfalls is that a lot of time can be lost acquiring basic knowledge on the fluxes of the pathway, before any actual results are achieved.

Growth models

Models to determine growth of microorganisms are very important in the fermentation industry, since the growth ultimately determines the number of organisms, and therefore productivity of a process. A model that is very often used is the Monod model describing the influence of the concentration of a limiting substrate on the growth rate of an organism. The limiting substrate can be sugar or a nitrogen source.

$$\mu = \mu_{\max} * \frac{C_s}{C_s + K_s} \tag{13.1}$$

where μ is the specific growth rate of the organism (1/h); $\mu_{\max}$ is the maximum specific growth rate of the organism (1/h); C_s is the actual substrate concentration (mmol/l); and K_s is the affinity constant for the substrate (mmol/l).

Figure 13.2 gives the growth rate for various values of K_s. Models like this have several advantages. First of all, both extremes are definitely predicted correctly. When the substrate is not limiting the growth is $U_{\max}$, and when no substrate is available the growth rate is zero. The second advantage of the model is that only one parameter is required to calculate the substrate concentration as function of the growth rate. Finally the model is mathematically very simple.

On the other hand a disadvantage of this model is that the value of the affinity constant can be very low (in the order of several mmol/l). To measure the value of the affinity constant, normally a chemostat is used. This allows organisms to grow at a controlled specific growth rate. Measuring the substrate concentrations at various growth rates will give the affinity constant. When the affinity constant is very low, however, it cannot be determined accurately, since the organism will

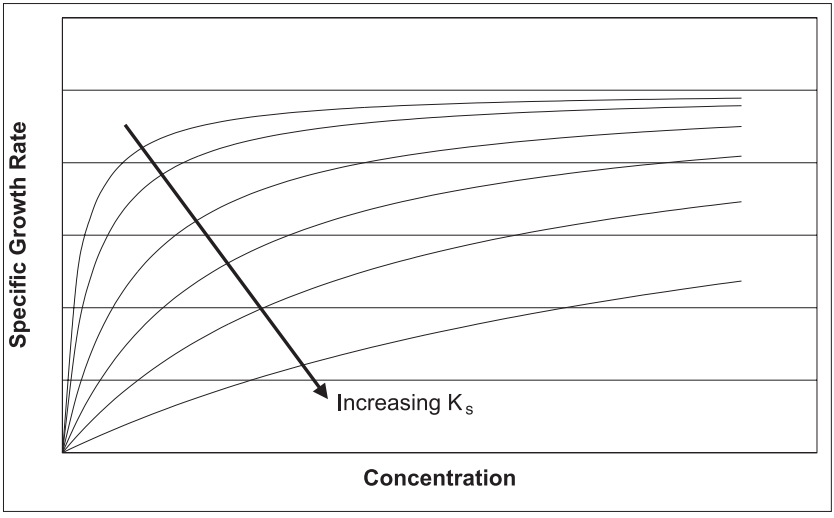


Fig. 13.2 Specific growth rate v. substrate concentration for various values of K_s.

consume the little bit of substrate before a proper analysis is done. Even though this means that in that case the growth control is almost on/off, this kind of equation is still used with low values of K_s, because calculation is simpler.

Killing-off models

A very important aspect of fermentation is sterility. Sterility is required to maintain a single strain culture of the desired production organism. Sterility can be achieved in several ways. Heat treatment, filtration and radiation are commonly used.

Models are used for killing-off by using heat treatment, because heat treatment should be applied carefully. In most media the quality of the medium decreases when it is at high temperature for a long time. Therefore, the ideal heat treatment is just sufficient to kill all the organisms, and not more than that. This requires a proper model for the death rate of the organisms. Much research has been carried out to describe the killing-off, and this is normally described as the time required for a decimal reduction of a certain type of organism at a certain temperature. When the infectious organisms are counted in the material that has to be sterilized, a minimal sterilization time can be calculated. This model is a very powerful tool to determine the sterilization time and temperature, and the results are normally very close to reality. For more detailed information see the chapter on inactivation in this book.

On the other hand, the financial consequences of contamination are too large to take any risks in this field. So even when a proper estimate of the sterilization can be made, normally a large excess of sterilization time is used to avoid any risk. This is an example where proper models are available, but they are not used properly in practice to avoid any risk.

Production models

When the product is not the microorganism itself, but, for instance, some protein or secondary metabolite, growth models are not sufficient to describe the production process. In that case a model is required to describe the amount of biomass in the process, and a second model should describe the amount of product the biomass produces over time. Two different models are given here, including the consequences for optimizing the process based on those models.

Constant production rate

This is the simplest model imaginable for description of the production. The model simply states that the biomass starts production at a certain rate as soon as it is formed, and keeps producing at that rate until the process is stopped. In other words, the specific production rate is constant, and independent of the physiological state of the organism. If experiments show this model to be correct for a certain process, optimization is very simple. Just grow as much biomass in as short a period as you can, and then keep that biomass at a low growth, and keep it producing.

Not many processes can be described by this model, but serious efforts are being made to design processes according to this model. When the growth and production can be uncoupled by adding a component essential for growth, but not for production, during the early stages of the process, this model can be realized in practical processes. If this can be done successfully, high yielding processes can be developed. This shows the strength of the use of models. Use of this simple model actually led to the idea of uncoupled growth, a more practical model.

Production coupled to growth

A model that is more applicable can be derived when the specific production rate of the organism is coupled to the growth rate of the cells. This model actually states that the production rate is coupled to growth because the limiting step for both processes is identical. This can be understood when, for instance, the limiting step is the amount of sugar that is used by the microorganism. The sugar is used by the microorganism for growth and production at a certain ratio.

If such a model describes a certain process (and only experimental work can prove whether this is true or not), optimization of the process leads to completely different results than the process in the last paragraph. When growth and production are coupled, it is important to keep the organisms growing, because otherwise the production rate will decrease. This means that the biomass concentration in the process will increase continuously during the process. Since some of the essential component for growth can only be added in limiting amounts (like oxygen), the amount of biomass that can be produced in a certain amount of time is limited. In that case a higher biomass concentration means a lower specific growth rate. Optimizing such a process requires delicate balancing of growth and production. Therefore it is very important to determine the actual relation between growth and production accurately.

13.3.2 Technological models

The genetic properties and the physiological state of the production organism determine how much product an organism produces. After that the productivity of the process is determined by the amount of organisms that can contribute to the production. Optimizing the biomass in the process gives some technological challenges. This can be illustrated by comparing the concentration of biomass and the biological activity of the process to situations in normal life.

In a production fermentor high concentrations of biomass are present. This can only be done when the environment is kept constant, and all necessary media components are available at all times. Normally in a production process the biomass concentration can be up to 100 g of dry biomass per litre, growing at a rate of 0.1 per hour (this is a 10% increase per hour). In doing this, the biomass consumes 20 g of substrate per litre per hour. The problems in maintaining that activity can be shown by comparing these data with human biological activity, i.e. to obtain the same concentration of people you need about four persons in one cubic metre! Apart from that these four people should be fed with around 20 kg of sugar every hour, i.e. 5 kg of sugar per person per hour!

Maintaining the correct temperature and making sure the nutrients are available is the subject of the technological work required to run a fermentation process. The major technological problems in fermentation are related to:

- heat transfer
- oxygen transfer
- mixing.

Heat transfer

Heat transfer is essential for fermentation processes, since heat is generated by the biological activity of the cells. If that heat is not removed, the temperature will increase, which will slow down production, or even kill the organism. The heat generated is removed by using cooling water in coils inside the fermentor (see Fig. 13.1). This is one of the classical examples of a scale-up problem, as the volume (and, therefore, the total heat produced) increases by the third power of the scale, whereas the surface area through which the heat is removed increases by only the second power of the scale. This means that a limit will be reached at a certain scale. Practical solutions are available to increase the surface area by adding cooling to the outside of the fermentor, the baffles, and even to the shaft.

Instead of modelling the cooling process, a standard model for heat exchangers is often used for estimating the heat transfer capacity of a fermentor. When this is not accurate enough, the process is adjusted slightly when it appears that heat exchange becomes limiting.

Oxygen transfer models

A lot of models are developed for the description of oxygen transfer in large-scale fermentors. These models describe the transfer of oxygen through a thin boundary layer around the bubble: the oxygen transfer increases when the surface area

increases (or when the bubble size decreases), and when thickness of the boundary layer decreases. Since each of these properties as such cannot be determined easily, the oxygen transfer is often described by the combined parameter $k_l a$, representing both the specific surface area a and the resistance for oxygen k_l.

The main advantage of this model is that the whole process is described by only one parameter that can be determined experimentally. This means that the influence of changes in the process on the oxygen transfer can be studied easily. In practice, empirical relations are available for the determination of the oxygen transfer as a function of, e.g., fermentor geometry, stirrer speed, airflow, and viscosity with prediction accuracy of ±10–20%. The reason is that these models do not describe the very complex reality inside a large-scale fermentor. In reality there are positions inside the fermentor that have a very high oxygen transfer, especially in the stirrer zone, whereas near the wall and near the surface have only very low oxygen transfer. In the case of a large-scale fermentor the oxygen concentration in the gas phase is not constant either. This means that the organism that is transferred through the fermentor experiences moments of high and low oxygen transfer.

As there is no generally accepted oxygen transfer model available for large-scale fermentations, the oxygen transfer rate has to be determined experimentally for a specific process in specific equipment. This situation is not very likely to change in the near future, since experimental work is very expensive.

Mixing models

Mixing models are probably the best developed technological models used in the fermentation industry. These models are of paramount importance, since they describe the variation the organism experiences in the fermentor.

Models using mixing time

The first models were developed based on concentration measurements after small pulses of salt or acid additions in water. These models described the mixing, as the time t_m required to realize 95% of the final concentration in the liquid (see Fig. 13.3). In the same way as for oxygen transfer experimental correlations are available to determine this time.

The advantage of such a model is that the effect of changing the process properties on the mixing can be studied by determining the effect of the change on this mixing time. The disadvantage is that it says nothing about the maximum and minimum concentration the organism may experience when chemicals are added into the fermentor. This is very relevant, since during fermentation very aggressive chemicals like concentrated sulfuric acid are added for pH control. If an organism experiences too high local concentrations of these components for too long a time, it will die.

Compartment models

When describing mixing in terms of mixing times is not sufficient, the use of compartment models is a better way of describing mixing. The idea behind these

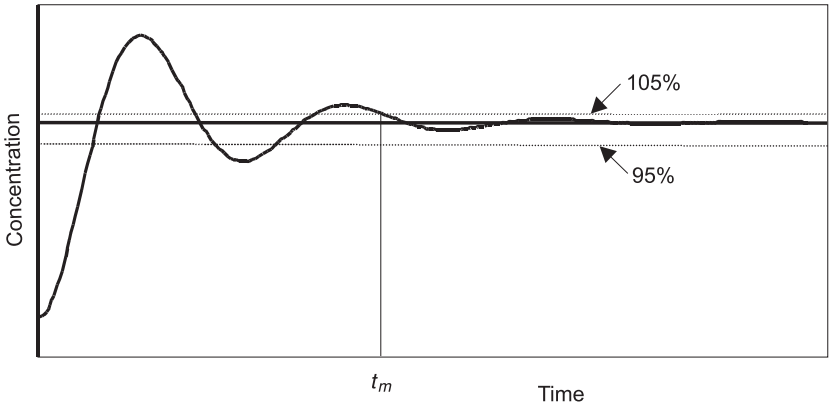


Fig. 13.3 Concentration measurement to determine mixing time.

compartment models is to divide the fermentor into compartments that are supposed to be ideally mixed. For every compartment a description is made for production and consumption of various components, and for the interaction between the different compartments. This description may be based on very thorough experimental work, or may be based on *a priori* knowledge. With such a model better results can be obtained than by just using mixing times. In particular, when the compartments are chosen cleverly, proper estimates of concentrations at certain places in the fermentor can be made. Also the effect of small changes of geometry can be investigated.

In the past much research was carried out to validate these models. This was accomplished by dividing the fermentor into two or more compartments with different properties. For instance all the feed was added to one compartment, and all the oxygen in the other. The production organism was pumped from one compartment into the other, mimicking the fluctuations present in the actual fermentor. The effect of poor mixing in the production fermentor on the production can be determined by comparing these results to a process where mixing is very fast. In this way the mixing effect of scaling-up is separated from other scale-up effects. These kinds of system are still in use for estimating the effect of poor mixing on the actual outcome of the fermentation.

Computational fluid dynamics

One step further from the compartment models is the use of computational fluid dynamics. When modelling the flow in a vessel like a fermentor, an obvious approach is to use the knowledge developed in fluid dynamics (see Chapter 4). In fact the differential equation describing fluid flow has been known for a long time as the Navier-Stokes equation. In principle the flow of liquids can be calculated with this equation in very fine detail. This is done by dividing the content of the vessel into small volume elements, and by solving the Navier-Stokes equation numerically for every volume element. The amount of calculations required increases considerably when the size of the volume

elements decreases. On the other hand, small elements are necessary to be able to calculate the effect of small eddies. There will always be a trade-off between the calculation power available and the level of detail required.

This technique is currently developing very fast. Standard software is available to do this work. Also many scientific groups are working on further development of this software. The most important advantage of this technique is that basically all the information required of the liquid flow is available after the calculations are done. This not only includes average flows, but also the fluctuation of the flow, the local energy dissipation, local concentrations, etc. These models are very useful in relatively simple systems. In one-phase systems these calculations give results that match very well to the actual flow, and can be used for determining mixing times. For this reason this technique is widely used in industry when one-phase flow is applied.

Inside a fermentor the situation is much more complex. Often a lot of air is sparged into the liquid, and the properties of the liquid are not well defined and continuously changing. Apart from that, the response of the microorganism to certain environmental conditions is not known, and can be changed by the physiological state of the organism. Also, the actual size of industrial fermentors makes the number of volume elements very large, and therefore the amount of calculations required to obtain useful results is also very large.

Since the development of very fast computers will continue in the future, this modelling will give better results and will be used in more complex systems in the future. It will be very hard, however, to describe the interaction between the rheology and physics of the system and the organism.

13.3.3 Downstream processing models

The recovery of the product out of the fermentation broth normally requires several processing steps. A typical downstream process contains process steps for cell removal, concentration and purification of the product. The fermentation broth should have a low concentration of the product, and contain cells and undesired components in the final product of the proper quality. What kinds of steps are used is dependent on the type of product, and the purity in which the product can be sold. For some products cell removal and concentration may be sufficient to get the desired quality, for others several purification steps are used to get the proper product. To be able to design an optimal downstream processing system, modelling is performed on two different levels.

Modelling of separate unit operations

It is known how to handle the different unit operations with standard materials. Most of these unit operations are well described in the literature for the chemical engineer. However, the product from fermentation can be rather complex, and may vary during the process. For this reason the unit operation must be tested with the proper medium, to make sure this particular process step can run without any problems, and determine the deviations when the unit operation is

run with real broth compared with when it is run with model liquids. Modelling this knowledge will not lead to results. To do this, normal equipment is tested with the actual fermentation broth. To reduce costs of this experimental work, the equipment required is rented when possible.

Modelling of the entire recovery factory
When all unit operations are tested and designed, a model for the entire factory must be developed. This model will describe how the scheduling in the factory can be done, and how the interaction between the various processing steps is organized. It will also be used for the sizing of the equipment. This work will include estimates of workload of the operators in the factory, and what to do when disturbances occur. This part of the design of the factory is of paramount importance for proper functioning of the factory.

Both applications of the models are essential for the design of a proper downstream processing factory. If one of the unit operations is badly designed, the process may not work, resulting in a product of undesired quality or in no product at all. If the modelling of the entire factory is done poorly, the factory will not operate optimally, leading to long processing times and disturbances. Errors in the design of the process will lead to large costs and delays in process development if they have to be corrected afterwards. Proper modelling of the process can avoid these problems.

13.3.4 Economic models

Although not of scientific or technological nature, economics and, therefore, reliable financial models, are very important in the development of industrial fermentation processes. This aspect is very often forgotten by scientists and technicians during the development of a process, while most of the time money was the reason for the development or optimization of a process in the first place!

The similarity of economic models to the other models is that the actual result of the development of the process will improve when the models are improved. Close co-operation of scientists and technicians with financial experts during the execution of the development work is crucial, and mutual understanding of the incorporated technical and financial models is an important aspect. An integrated business and R&D approach offers the right condition for such a way of working and modelling.

13.3.5 Mixed models

Of course the result of any model improves the more information is used. Current computer systems are capable of running very complex models in real time. Because of that, models are being developed that use all information available for controlling and optimizing processes. This information does not only imply the models mentioned above, but also other obvious things that

create useful information. The use of elemental balances for all the elements in the process gives very useful information on what part of the input is used for the production of essential products.

These combined models can also use so-called fuzzy logic on the features of the process on which no other information is available, and for parameter estimation. In this way the model may learn from historical results to make even better predictions in the future. This structured way of developing models has already been shown to be very effective in various industrial processes.

13.4 Future trends

More and better models will be built for further development of processes and knowledge. There is very little doubt that these models will become more complex, because of increasing computing power. Building models will be the main task of people working in this field of development. However, while making these complex models, the development of very simple models should not be neglected. Simple models will give the opportunity to explain what happens in the process to outsiders to the discipline. These outsiders can be very important since very often the people who pay for the work that has been done are a part of that outsider group. A proper understanding of the process results by managers and customers is probably the most important thing that can be achieved by using models. If that understanding can be created, a developed process can be transferred from a technological and scientific success into a business success. This will be essential to develop future processes and models.

14

Improving modified atmosphere packaging through conceptual models

M. L. A. T. M. Hertog, Massey University, Palmerston North, and N. H. Banks, Zespri Innovation, Mt Maunganui South

14.1 Introduction

Conceptual models are descriptions of our understanding of a system that are used to shape the implementation of solutions to problems.[1] The quality and quantum of innovation that will occur in development of modified atmosphere packaging (MAP) strongly depends upon the insights gained from robust conceptual models of components of MAP. In this chapter, we outline a number of simple principles about modified atmosphere (MA) systems that we believe will assist industries that apply MA technology to move beyond the rather empirical '*pack-and-pray*' approach that still predominates in commercial practice.

MA is generally used as a technique to prolong the keeping quality[2] of fresh and minimally processed fruits and vegetables.[3] In the widest sense of the term, MA technology includes controlled atmosphere storage, ultra low oxygen storage, gas packaging, vacuum packaging, passive modified atmosphere packaging and active packaging.[4–7] Each of these techniques is based on the principle that manipulating or controlling the composition of the surrounding atmospheres affects the metabolism of the packaged product, such that the ability to retain quality of the product can be optimised. The different techniques come with different levels of control to realise and/or maintain the composition of the atmosphere around the product. While controlled atmosphere storage can rely on a whole arsenal of machinery for this purpose, active packages rely on simple scavengers and/or emitters of gases such as oxygen, carbon dioxide, water or ethylene either integrated in the packing material or added in separate sachets. Passive MA packaging, as an extreme, relies solely on the metabolic activity of the packaged product to modify and subsequently maintain the gas composition surrounding the product.

Although much research has been done to define optimum MA conditions for a wide range of fresh food products,[8] the underlying mechanisms for the action of MA are still only superficially understood. The application of MA generally involves reducing oxygen levels (O_2) and elevating levels of carbon dioxide (CO_2) to reduce the respiratory metabolism.[6] Parallel to the effect on the respiratory metabolism, the energy produced to support other metabolic processes, and consequently these processes themselves, will be affected accordingly.[9] This still covers only part of the story of how MA can affect the metabolism of the packaged produce. The physiological effects of MA can be diverse and complex.[10] In MAP, the success of the package strongly depends on the interactions between the physiology of the packaged product and the physical aspects of the package; MAP is a conceptually demanding technology. Much of the work in the area of MAP has been, and still is, driven by the practical needs of industry.[11] This has enabled commercial development based upon pragmatic solutions but has not always contributed substantially to advancing the conceptual basis upon which future innovation in MA technologies depends. As a result, there is a substantial potential for models to contribute to the field of MAP by making the complex and vast amount of, sometimes fragmental, expert knowledge available to packaging industries.

In this chapter, we bring together existing concepts, models and sub-models of MAP to build an overall conceptual model of the complex system of MAP. Starting from this overall model, dedicated models can be extracted for specific tasks or situations. The benefits and drawbacks of the modelling approach are discussed, together with an identification of the future developments needed to create advantage to MAP commercial operations.

14.2 Key principles and methods

14.2.1 A general conceptual approach

The ideal model integrating all critical aspects of MAP would inevitably have a multidisciplinary nature and a complexity that, at least in its mathematical form, is far beyond the scope of this chapter. Here we attempt to provide a sound conceptual model to assist understanding of the underlying mechanisms.

Going in aggregation level from the macro (palletised packs) via the meso (individual packs) to the micro level (packaged product) the emphasis shifts from physics and engineering to include more and more biology, physiology and microbiology. In parallel to this shift, the level of complexity and uncertainty increases.

Macro level

The macro level is schematically presented in Fig. 14.1. Much research has been undertaken on heat and mass transfer, the effects of boundary layers and different flow patterns given different geometries, types of cooling and ventilation.[12, 13] The same techniques have been applied to the storage of living

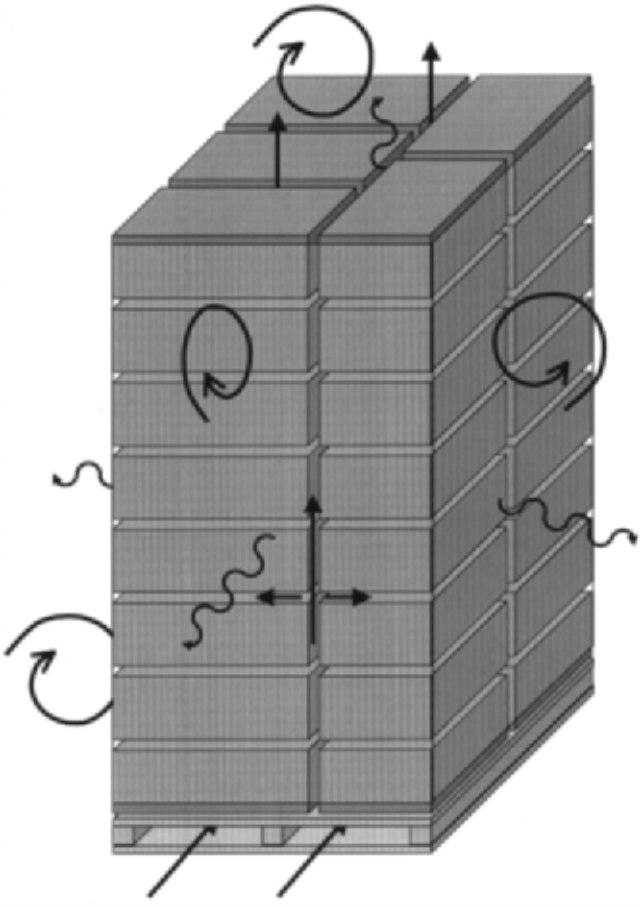


Fig. 14.1 A schematic outline at the macro level of MAP where forced airflow and turbulent convection are responsible for heat and mass transfer to and from the individual MA packs.

and non-living food and non-food products all over the world. These techniques enable, in general, a good understanding of the storage environment of palletised or stacked packs, whether or not MA packs. Cooling is needed to remove heat from the packages and to counteract continuously the heat produced by the living product. Both forced airflow and turbulent convection are at this level major contributors to the transport of heat, water, gases and volatiles, to and from the packs.

Meso level

At the level of individual packs (Fig. 14.2) the emphasis moves towards natural convection and diffusion processes driven by concentration and thermal gradients. Heat produced by the product is conducted directly, or through the

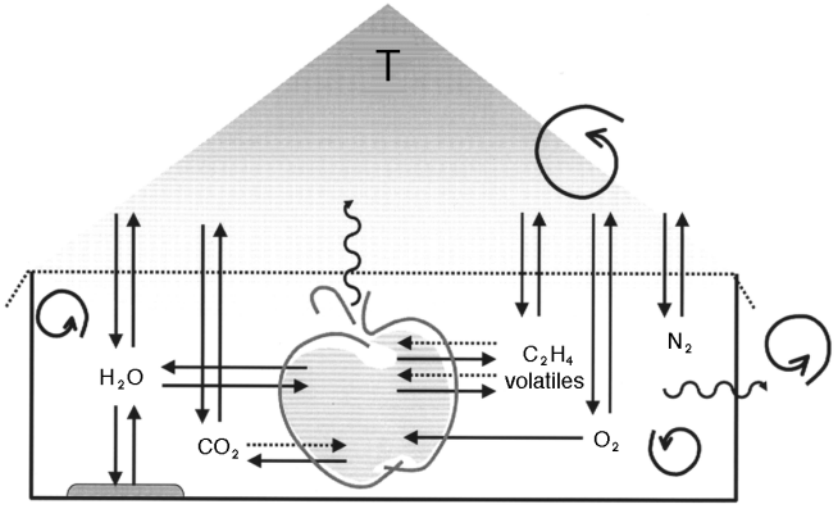


Fig. 14.2 A schematic outline at the meso level of MAP where heat and mass transfer from and to the packaged product are ruled by natural convection and diffusion processes. The packaging film acts like a selective semi-permeable barrier between the package and the surrounding atmosphere. Temperature has a marked effect on all processes going on at the meso level.

atmosphere in the package, to the packaging material and, eventually, is released to the air surrounding the pack. Water vapour, respiratory gases, ethylene and other volatiles are exchanged between the package atmosphere and the surrounding atmosphere by diffusion through (semi-)permeable packaging materials. Those packaging films can be either selective semi-permeable films or perforated films. Especially in the case of perforated films, the diffusion rate of a gas can be influenced by a concurrent diffusion of a second gas.[14] A countercurrent generally hinders the diffusion while a current in the same direction promotes the diffusion of the first gas.

Inside the package, the metabolic gases are either consumed (O_2) or produced (H_2O, CO_2, C_2H_4 and other volatiles) by the product. Each of these gases may promote or inhibit certain parts of the product's metabolism. In the end, the overall metabolism of the packaged product is responsible for maintaining the product's properties. As long as the product properties relevant for the quality as perceived by the consumer stay above satisfactory levels the product remains acceptable.

The steady state gas conditions realised inside an MA pack are the result of both the influx and the efflux through diffusion and the consumption and production by the product which are themselves strongly dependent on the composition of the package atmosphere.[15] For instance, water loss by the

product is the main source for water accumulating in the pack atmosphere. The product elevates humidity levels within the pack to an extent that depends upon relative water vapour permeances of film and product. This elevated humidity inhibits further water loss to a progressively greater extent as relative humidity approaches saturation. This substantial benefit carries a risk of condensation that is exacerbated by temperature fluctuations. Condensation creates favourable conditions for microbial growth that will eventually spoil the product and also depress the permeance of the packaging film.

The time needed for a package to reach steady state is important as from that moment on, the maximum benefit from MA being realised. In the extreme situation, the time to reach steady state could outlast the shelf-life of the packaged product. A typical example of how the atmospheric composition in an MA pack and gas exchange of the packaged product can change during time is illustrated in Fig. 14.3. The dynamics of reaching steady state depends upon the

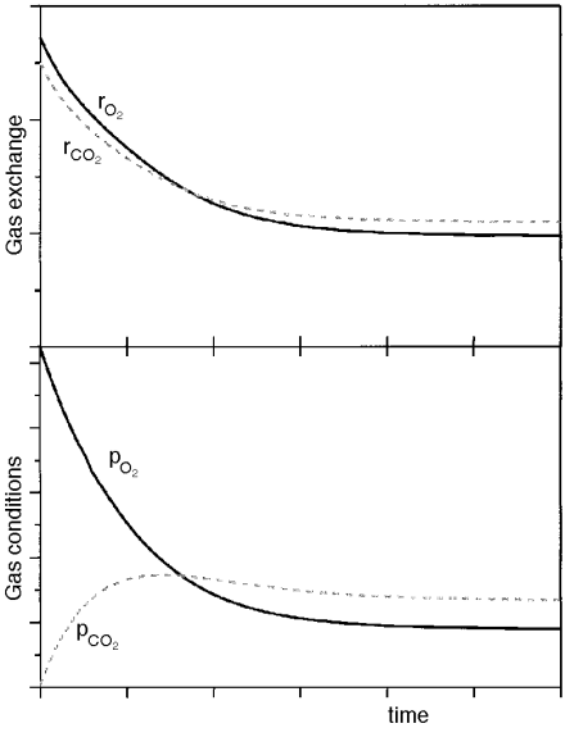


Fig. 14.3 A typical example of the dynamics of MA. Due to the gas exchange, CO_2 starts to accumulate while the O_2 level starts to decrease (bottom). In response to the changing gas conditions, gas exchange rates are inhibited (top). Driven by the increasing concentration gradients between package and surrounding atmosphere, O_2 and CO_2 start to diffuse through the packaging film. In combination, this slows down the change in gas conditions. Eventually, gas exchange by the product and diffusion through the film reach steady state levels at which the consumption and production of O_2 and CO_2 equals the influx and efflux by diffusion.

rates of gas exchange and diffusion and upon the dimensions of the package in relation to the amount of product packaged. Packages with large void volumes take longer to reach steady state levels. Temperature has a major effect on the rates of all processes involved in establishing these steady state levels[16] and hence on the levels of the steady state themselves.

Micro level

Gas exchange

The complexity of the biological system inherent in each fruit (Fig. 14.4) contributes significantly to the uncertainties in current knowledge on issues critical to the outcome of MA treatments. One of the central issues is the impact of MA upon the product's gas exchange, its consumption of O_2 and production

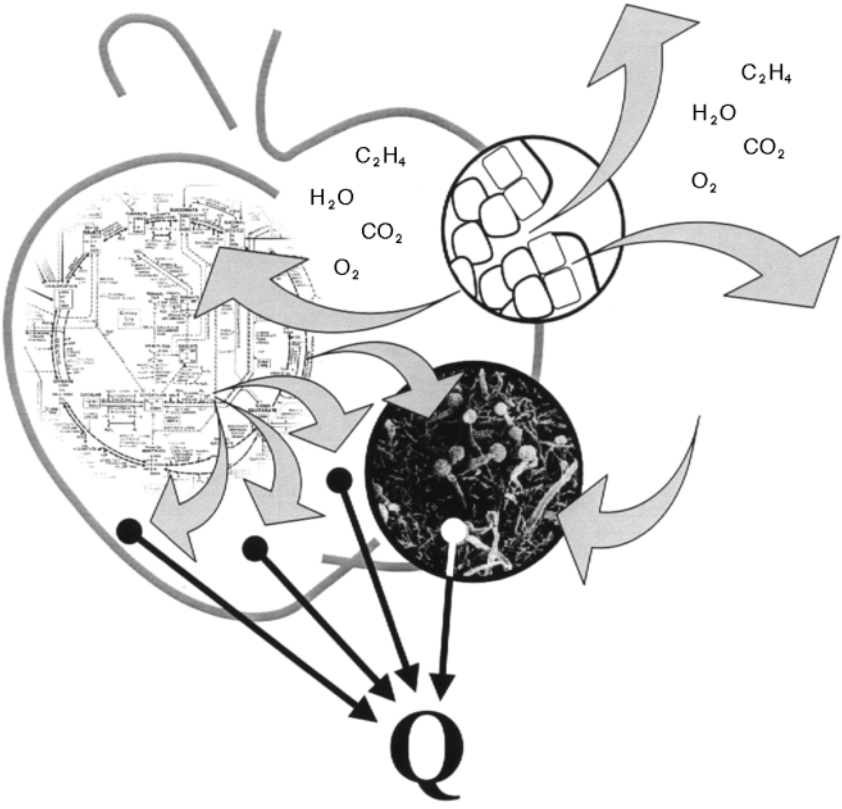


Fig. 14.4 A schematic outline at the micro level of MAP where the product is considered to generate its own MA conditions due to the resistance of the skin. The internal gas conditions are responsible for affecting large parts of the metabolism either directly or via the gas exchange. This will influence quality related product properties determining the quality (Q) as perceived by the consumer. Depending on the MA conditions, microbes can interact with the product's physiology influencing its final quality.

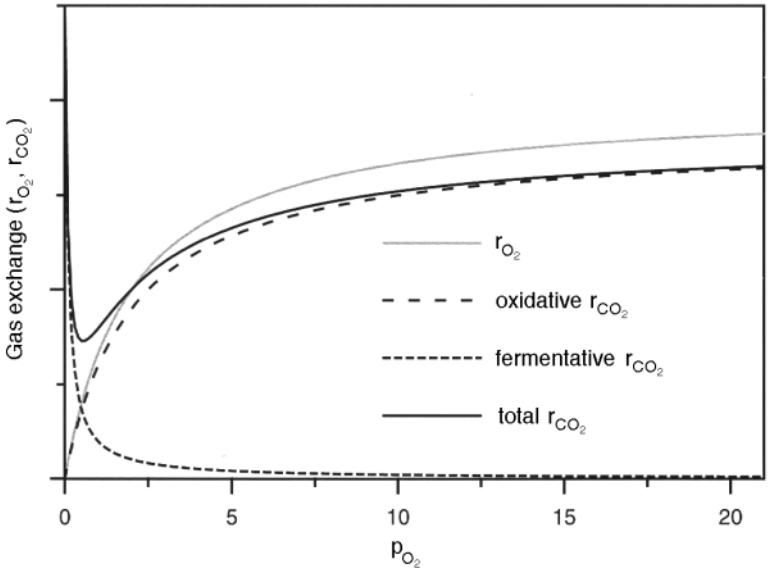


Fig. 14.5 A typical example of gas exchange as a function of O_2 partial pressures (P_{O2} in kPa). (r_{O_2}) is related to the oxidative part of CO_2 production (r_{CO_2}) via the respiration quotient. Additionally, at low O_2 levels fermentative CO_2 production can take place resulting in an increased CO_2 production as compared to the decreasing O_2 consumption.

of CO_2 (Fig. 14.5). Total CO_2 production consists of two parts, one part coming from the oxidative respiration in parallel to the O_2 consumption and the other part originating from the fermentative metabolism.[17] At high O_2-levels, aerobic respiration prevails. In this situation, the respiration quotient (RQ; ratio of CO_2 production to O_2 consumption), influenced by the type of substrate being consumed, remains close to unity. At lower oxygen levels, fermentation can develop, generally causing a substantial increase in RQ. This is due to an increased fermentative CO_2 production relative to an O_2 consumption declining towards zero. Besides the effect of O_2 on respiration and fermentation, CO_2 is known to inhibit gas exchange in some produce as well.

Although it would be convenient to consider gas exchange to be constant with time, there can be considerable ontogenetic drift in rates of gas exchange.[18] Especially in so-called climacteric fruits, a respiration burst can be observed when the fruit starts to ripen. In addition, freshly harvested, mildly processed or handled fruit generally shows a temporarily increased gas exchange rate.[9] Microbial infections can also stimulate gas exchange.[19]

Gas diffusion

When one considers gas exchange as a function of O_2 and CO_2 levels, one is generally inclined to look at the atmospheric composition surrounding the product as the driving force. However, the actual location of gas exchange is inside the cells, in the mitochondria. Depending on the type of product, this

means that an O_2 molecule has to diffuse through the boundary layer surrounding the product, through a wax layer, cracks, pores or stomata, through intercellular spaces, has to dissolve in water, and has to pass the cell membrane to get into the cell.[10] The CO_2 molecule produced by the gas exchange has to travel the same way in the opposite direction. The driving force for the diffusion comes from the partial pressure difference for O_2 and CO_2 between the fruit's internal and external atmospheres generated by the gas exchange. The intracellular, *in situ*, O_2 and CO_2 concentrations are much more relevant for the gas exchange than the fruit external gas conditions. Generally, it is assumed, however, that the largest resistance in the diffusion pathway from the surroundings into the fruit exists at the skin of the fruit.[20,21] Therefore, the largest gradient in concentration occurs at the skin while the concentration differences within a fruit are small.

Even at identical external atmospheres, different species of fruit will have completely different internal gas compositions due to their different skin permeances. Fruits with a wax layer, like apples, have a much lower permeance than leafy vegetables like cabbages, which generally have a large amount of stomata present.[22] The skin permeance of different apple varieties will be strongly affected by thickness of their natural wax layers. Due to such a wax layer, the skin of tomato and bell pepper is relatively impermeable, forcing all the gas exchange through the stem end of the fruit.[23] Consequently, some fruits go internally anaerobic at conditions where others are still aerobic.

Water diffusion and water loss

The diffusion of water vapour is limited by skin permeance in the same way as the diffusion of O_2 and CO_2. The slight difference is that the diffusion of O_2 and CO_2 is mainly going through pores connected to intercellular spaces while water vapour is more easily released through the whole skin surface.[24,25] Water loss is driven by the partial pressure difference for water vapour between the fruit's internal (close to saturation) and external atmospheres. Water loss is an important issue in relation to the overall mass loss, firmness loss and shrivelling or wilting of the product. Inside an MA pack, water loss can also be responsible for generating conditions favourable for microbial growth (high RH).

Ethylene effects

Being a plant hormone, ethylene takes a special place among the gases and volatiles produced by the product because of its potential impact on the product's own metabolism. The pathways of biosynthesis and bio-action of ethylene are still the subject of extensive study.[26] Most of the climacteric fruits show a peak of ethylene production at the onset of ripening. In most of these fruits, ripening can be triggered by exogenous supplied ethylene. This creates the situation that one ripening fruit in an MA pack will trigger the other fruit to ripen simultaneously, due to the ethylene accumulating in the pack. MA can inhibit the normal development and ripening of products postponing the climacteric ethylene production thus extending the keeping quality of the

product. With kiwifruit, however, advanced softening of the fruit occurs before ethylene is being produced.[27] Although the fruit is not producing any ethylene yet, the softening process is extremely susceptible for exogenous applied ethylene.

Product quality

The quality of the packaged product is based on some subjective consumer evaluation of a complex of quality attributes (like taste, texture, colour, appearance) which are based on specific product properties (like sugar content, volatile production, cell wall structure) (see also Chapter 17).[28] These product properties are generally changing as part of the normal metabolism of the product. Those developmental changes that are directly influenced by O_2 or CO_2 or driven by the energy supplied by respiration or fermentation will all be affected by applying MA conditions, potentially extending the keeping quality of the product. Some processes are more affected than others due to the way they depend on atmospheric conditions. To understand the mode of action of MAP for a specific product, a good understanding of how the relevant product properties depend on gas conditions and temperature is required.

Spoilage and pathogenics

MA conditions can also provide conditions favourable for the growth of microbes potentially limiting the keeping quality of the packaged product due to rot (see also Chapter 18).This is especially the case for soft fruits or minimally processed fruit and vegetable salads when high humidity levels are combined with a tasty substrate.[29] Some microbes are known to be opportunistic, waiting for their chance to invade the tissue when ripe, damaged or cut. In this case, MA conditions inhibiting the ripening of fruit in combination with a proper handling and disinfection can prevent some of the problems. Other microbes are more actively invading the tissue, causing soft patches on the fruit. More insight is needed into how MA can inhibit not only the metabolism of the product but also that of the microbes present on the products. High CO_2 levels are generally believed to suppress the growth of microbes, although sometimes the CO_2 levels needed to suppress microbial growth exceed the tolerance levels of the vegetable produce packaged.[29, 30]

Variation

Although the general concept of MAP is now almost complete, there remains one thing that affects all the other issues outlined so far, and that is variation (see also Chapter 8). Variation can occur on different levels, like time and spatial variation in temperature control in storage, irregularities in the stacking of cartons influencing ideal flow patterns, irregularities in the thickness or perforation of films or differences between batches of film used. However, the most important non-verifiable one is biological variation. Besides the more obvious differences between cultivars, distinct differences exist between produce from different harvests, years, soils or locations.[31] Even within one

batch, considerable variation between individual items can occur.[32] The amount of biological variation that can be expected generally depends on the organisation level examined. Within packages, product is generally coming from one grower resulting in a relatively homogeneous batch with limited fruit-to-fruit variation. Comparing different pallets involves product potentially originating from different growers and harvest dates resulting in a much larger variation.

When developing small consumer MA packages, variation in the rate of gas exchange is almost impossible to take into account. The larger the package, the more these differences tend to average out. However, in case of fruit interactions, individual outliers can affect the other fruit in a pack, as with the spreading of rots, the onset of ripening through C_2H_4 production or with off-flavour development.

14.2.2 Sub-models

Over the years, different elements of what has been discussed above have been the subject of mathematical modelling. Other subjects are still to be explored. Models describing the physics of MAP are usually more fundamental than the ones describing the physiology of MAP. This is due to the increased complexity and the lack of knowledge on the underlying mechanisms. For this reason, empirical 'models' (arbitrary mathematical equations fitted to experimental data) are still prevailing in post harvest physiology. This section gives an overview of the type of MAP-related models available in the literature with the emphasis on the physiological aspects of MAP.

Macro level

With the strong development of computers, rapidly increasing computational power becomes available to food and packaging engineers. Associated with this, engineers can add new numerical tools to their standard toolkit such as Computational Fluid Dynamics, infinite elements and finite differences. In general, when modelling heat and mass transfer, conservation laws are applied to formulate energy and mass balances (see Chapter 4).[12] The space under study is subdivided in a number of defined elements. Each of them is represented by one point within the three-dimensional space and is assumed to exchange mass and heat with its neighbouring elements according to the heat and mass balances defined. The accuracy of such a model strongly depends on the number and size of elements defined and the knowledge of system input parameters. To improve both accuracy and computational time, smaller elements can be defined in areas with steep gradients and larger elements in the more homogeneous areas.

Theoretically, this approach is applicable at both the macro level to describe airflow in a cold room, at the meso level to describe diffusion within a pack, and at the micro level to describe gradients within the product. The main application is, however, at the macro level and to a lesser extent at the meso level when large bulk packages are involved.[33] For small consumer size packages the

simplification of treating the pack atmosphere as one homogeneous unit is generally acceptable. At the micro level, there are too many system inputs still undefined to enable formulation of such a model, not to mention to parameterise and validate it.

Meso level

At the level of small consumer size packages the physics simplifies to relatively easy diffusion equations based on Fick's law describing gradient driven fluxes from point A to point B through a medium with a certain resistance. Gas permeates into (or out of) the package faster with increased film area, with thinner films and with larger concentration differences.[34] The permeance of a film typically depends on the material used. With the current range of polymers available, a wide range in permeability can be realised. Most films are selective barriers with different permeances to the different gases.[35]

The standard industry test for determining permeance of a specific film is done at the single temperature of 23ºC using dry air conditions. The conditions at which a film is exposed in MAP of fresh produce, ranges however from zero to 25ºC and high humidity levels ($> 90\%$). Depending on the actual temperature, the permeance of the film changes accordingly. This temperature dependence is generally described using an Arrhenius equation (see Chapter 3). This is an exponential relationship originating from chemistry where it is used to describe the rate of chemical reactions as a function of temperature. The central parameter quantifying the temperature dependence is the activation energy. The higher the activation energy the faster permeance increases with increasing temperatures. An activation energy of zero means that the permeance does not change with temperature. The activation energy is characteristic for the film material used and is different for the different gases.

The effect of humidity and condensation on the permeance of films is widely recognised and still the subject of study. At high humidities, water can be absorbed by the film changing the permeance for other gases as well. Furthermore, due to temperature changes, water can condensate on the film forming an extra barrier for diffusion. Both aspects are still to be modelled.

When perforated films are considered, the diffusion through the film can be decomposed into two processes: diffusion through the film polymer and diffusion through the pores. Perforations are generally much less selective as this involves just diffusion through air. In addition, the effect of temperature on diffusion through pores (air) is much less compared to its effect on diffusion through the polymer. As diffusion through the pores accounts for most of the total diffusion through a perforated film, the activation energies for perforated films are close to zero. Due to the effect of boundary layers, diffusion through pores is not linearly related to pore area and film thickness and some corrections have to be made depending on pore size and pore density. Models, originally developed to describe stomatal resistance in leaves, have been applied for this.[36]

The effect of concurrent diffusions has been modelled using Stefan-Maxwell equations.[14] These equations take into account the effect of collisions between

countercurrents of different species of molecules on their final diffusion rates and can explain some of the observed diversions from Fick's law of diffusion.

The effect of pack volume on the dynamics of MAP is something that does not need to be modelled explicitly. As both diffusion and respiration are defined as a function of partial gas pressures, and as these partial pressures depend by definition on the amount of molecules present per unit of volume, the volume is already incorporated implicitly. For instance, doubling the void volume of an MA pack means that twice the amount of oxygen molecules are available. To reduce the oxygen concentration in the void volume to a certain level, twice the number of molecules have to be removed, which takes about twice as long.

Micro level

Several attempts have been made to model the gas exchange by either empirical models, or strongly simplified fundamental or kinetic models using, for instance, a single Arrhenius equation.[37,38] A more fundamental approach was used by Chevillotte[39] who introduced Michaelis Menten kinetics to describe respiration on the cell level. Lee[40] introduced and extended this approach in the postharvest field to describe the respiration of whole fruit. After him, several other authors successfully applied this Michaelis Menten approach to a wide range of products[41,42] and extended the original Michaelis Menten equation to include different types of CO_2 inhibitions[17] and to account for the effect of temperature.[43–45] Traditionally, the effect of temperature was described using the Q_{10} system. More recently, the use of the Arrhenius equation is being favoured. The general applicability of the Michaelis Menten approach is probably due to the fact that it is simplified enough to enable parameterisation, and that it is detailed enough to account for the different phenomena observed.

Driven by dissatisfaction with the Michaelis Menten approach, as it may not describe the respiration of fresh produce because actual respiration is composed of many steps of metabolic reactions, Makino *et al.*[46,47] felt the need to develop an even more simplified model. Based on Langmuir's absorption theory, an O_2 consumption model was developed which, in the end, appears to be an exact copy of the Michaelis Menten approach, with parameters meaning the same, only labelled differently. Instead of developing an alternative for the Michaelis Menten approach, Makino unintentionally reinvented it and validated its assumptions via an analogous mechanistic approach.

Although proven extremely applicable for practical use and indispensable for enhancing the understanding and interpretation of gas exchange data, the Michaelis Menten type of formulation is a considerable simplification of the biochemical reality. This stimulates the ongoing research to generate models that are more detailed.[48] The developmental effect on gas exchange has not been modelled so far, except for some empirical corrections for an assumed drift of respiration during time.[9]

Burton[10] added a whole new dimension to MA research by stimulating research on internal atmosphere compositions of products as a key concept in the responses of fruits and vegetables to MA. He emphasised the concept of the skin

being a barrier between external and internal atmospheres. The same way film permeance alters the gas conditions inside the package, the skin alters the internal gas atmospheres.[49] Basically, the fruit can be considered as the smallest possible MA package. The mathematics behind modelling internal atmospheres is the same as that applied in modelling pack atmospheres. Assuming the largest resistance in the diffusion pathway exists at the skin of the fruit, diffusion from the pack atmosphere to the fruit internal atmosphere can be described with a simple diffusion equation using the permeance and area of the skin. The relation between fruit internal and external atmosphere conditions can be understood completely from the combined effect of skin permeance and gas exchange characteristics. However, the gas exchange model now has to be parameterised as a function of fruit internal gas conditions instead of pack atmosphere conditions.

With regard to gas exchange, the combination of diffusion equations and Michaelis Menten type kinetics resulted in generally accepted and applicable models. As far as product specific issues are concerned, models are completely lacking or only available in a rudimentary form. However, to complete the overall MAP model we do need sub-models on how MAP is influencing the physiology of the packaged product beyond their gas exchange. How do the quality determining product properties depend on the gas conditions, either direct or via the changed gas exchange? The development of such models is severely hampered by the lack of physiological knowledge and complete sets of data for validation. Though empirical or statistical models can be useful to describe simple relationships found in a specific experiment, robust mechanistic models are needed to develop predictive models that can be applied under a wide range of conditions.

A relatively simple problem like shrivelling of apples due to water loss can be understood easily from the diffusion of water from the fruit internal into their external atmosphere.[50] The analysis of the results is hampered, though, by the large biological variation in skin permeance.[51] However, due to its generic mechanistic approach, the model can be integrated easily within the larger MAP model for a wide range of products.

The colour change of some products (tomatoes,[52] cucumber[53]) has been modelled successfully. What remains to be investigated, is how these colour changes are affected by the gas conditions. With the colour change of broccoli buds, MA conditions were shown to have an effect on the rate of colour change.[54] Whether this was directly related to the reduction in gas exchange was not tested. In the case of rot development in strawberries, Hertog *et al.*[55] assumed that the metabolic rate was the direct driving force for the progress of ripening enabling microbes to develop rot. Reduction of spoilage under MA could be explained from this reduction of gas exchange.

An extremely complex, and relevant issue of how ripening of (climacteric) fruits is affected by gas conditions has not been unravelled, let alone been modelled. However, based on some general concepts, Tijskens *et al.*[56] developed a simplified mechanistic model describing the softening of apples

under MA including some of those climacteric developmental changes. Although this model is a strong simplification of the physiological reality, it shows the generic potential of well formulated mechanistic models.

Understanding the mode of action of MAP for a specific product requires knowledge of how the relevant product properties depend on the gas conditions (composition and temperature). This is what makes MAP a laborious exercise as each product can have different quality determining product properties responding in slightly different ways to the MA conditions applied. One way to get around this is by developing generic models describing phenomena like shrivelling, softening, sweetening, mealiness, flesh browning or skin colour change that can be validated independently for a wide range of products.

Another option that has already proved itself successful is to stick to a more general level, describing keeping quality independent of the underlying product properties. This generic approach was originally developed by Tijskens and Polderdijk[2] to describe the effect of temperature on keeping quality for a wide range of commodities. This approach was extended to include the effect of MA, assuming all quality decay is driven by the metabolic rate.[54,57] Some further refinement would be needed to discriminate, for instance, between respiration and fermentation driven quality decay processes. This approach can give insight into how much MAP is able to extend keeping quality without the need to unravel the exact mechanism of how, for instance, firmness of apple is influenced by MA conditions at the biochemical level.

Modelling in microbiology has always been important. Usually growth curves are described as a function of temperature, pH, water activity and in response to the presence of competing microbes at well-defined growth media (see also Chapter 18).[58] The relevance of microbes for MAP increased with the increasing demand for convenience foods stimulating the markets for MA packaged cut and slightly processed fruit and vegetable mixes. Low numbers of microbes in foods may already result in hazardous situations. However, the currently available models in predictive microbiology are not set up to deal with these low numbers.[58] Instead of modelling actual numbers, probability of presence should be taken into account. In addition, the composition of the natural growth medium in MA packages (being the fruit and vegetables) is not well defined and highly variable. To predict growth of microbes in MA packages, both the biological interaction between produce and microbes and the direct effect of changed atmospheric conditions on the growth rates of microbes should be taken into account. Research in this field is still developing[29] and given its complex nature, mechanistic models integrating the outlined microbial aspects will not be readily available.

Although (biological) variation is hard to model, the effect of variation in a system can be easily demonstrated once a model of that system is available. By simply running the model multiple times, taking randomly distributed values for one or more of the model parameters, the effect of variation becomes clear. Such a so-called Monte-Carlo approach can be applied to a MAP model by drawing randomly distributed values for, for instance, film thickness or the product's

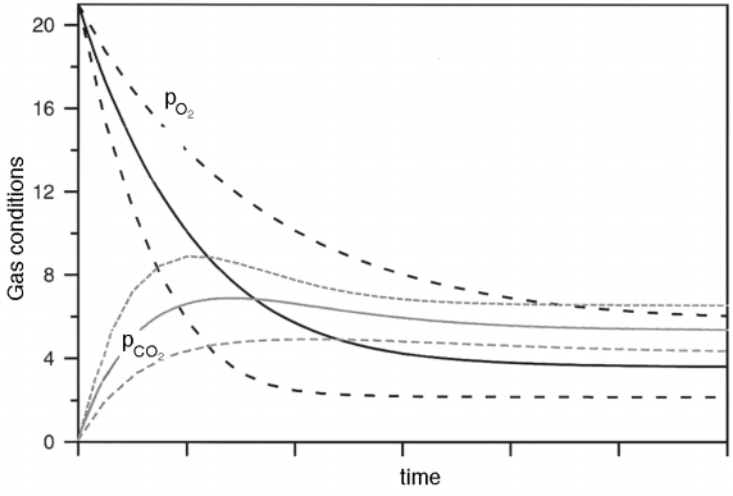


Fig. 14.6 For a given MA package the effect of ±25% biological variation on gas exchange and ±10% variation in packaged biomass were calculated. The average MA conditions developing during time (solid lines) and their 95% confidence intervals (dotted lines) were plotted based on 200 simulations.

respiration rate. Based on the results, a 95% confidence interval for the gas conditions inside an MA pack can be formulated (Fig. 14.6). Variation can induce certain risks, especially when package atmospheres are targeted close to what is feasible. When desired O_2 levels are close to the fermentation threshold,[59] the risk is that some of the packages, depending on the variation in gas exchange rate, result in O_2 levels dropping below the fermentation threshold.[60] This results in packages with unacceptable fermented produce. Biological variation is generally much larger (±25% is not exceptional) than the physical sources of variation (generally less than ±10%) as the physical factors are generally easier to control. Biological variation also comes back in the initial quality of the packaged product resulting in different length of keeping quality or shelf-life. Some quality change models try to account for these sources of variation.[53, 55]

14.2.3 Dedicated models

Though it is possible to develop a model covering all facets of MAP at all levels, such a model would be impossible to operate. Before one is able to use such a model it needs to be fully parameterised. Generally, this information is not available in all situations. Moreover, it is not always relevant to go to such a level of completeness. Depending on the specific issues involved in a particular application, dedicated MAP models can be extracted from the overall conceptual model. Some elements need to be worked out in more detail while others can be simplified or assumed constant, depending on the dedicated application.

A retailer trying to deliver the best for the end users is mainly interested in consumer size packs and how keeping quality develops during display in retail and after purchase at the consumer's place (shelf-life). In this case, the emphasis would be on a product-specific keeping quality model linked to the change in pack atmosphere conditions. The surrounding conditions are taken as they are. A large exporting company sending off wrapped pallets with product would be interested in whether the MA conditions stay within some given target limits. The emphasis is now on how to control the conditions inside a container to maintain constant MA conditions and how to optimise package design and pallet stacking to promote homogeneous flows and heat exchange throughout the bulk load. When developing packages for minimally processed salads, the emphasis is on incorporating predictive models on microbial growth together with specific models on the product's physiology.

14.3 Areas of application

The previous section already mentioned some potential applications for MAP models. In this section, we explore some of the potential of MAP models to enhance the practical implementation of MAP and to lift it beyond the phase of 'pack-and-pray'.

14.3.1 Dimensioning MAP

There is more than one 'right' solution to the search for a suitable MA package for a specific product. Assuming the product is known, including its gas exchange characteristics and some optimum target MA conditions, and the external storage conditions are set but beyond control, there is still a number of degrees of freedom through which the MA package can be manipulated for better or for worse.

To realise the target MA conditions the total permeance of the package has to be dimensioned in relation to the amount of product packaged. Besides choosing a different film material with a higher or lower permeability, film thickness and film area can be chosen. A film that is too permeable to be used as a wrapping can give good results when used to seal the top of an impermeable tray because of the reduced diffusion area. A film that is suitable for a small consumer pack can be too impermeable to be used as a liner in a carton because of the increased amount of biomass per unit of available diffusion area. Trying to influence this ratio, by packing less produce in a package, results in an increased void volume. This increases the time needed for the product to bring the package gas levels to the target MA conditions. This is not favourable, as it takes longer before the product gets the maximum benefit of the optimum MA conditions. However, the buffering capacity of a relatively large void volume can have its positive effects when the MA package has to survive short periods of suboptimal conditions. During a short warm period, a package with a small void volume could rapidly

generate anaerobic conditions while a package with a large void volume is already transferred back to cooler conditions before becoming anaerobic.

Even optimal dimensioning an MA package appears to be complex due to the different interactions. A MAP model can considerably enhance this search for a package with a fast enough dynamic phase, resulting in steady state values close to the target gas conditions and enough buffering capacity to be applicable in practice.

14.3.2 Developing new films

In case a company wants to bring a new MA pack on the market for a specific product with the package dimensions already set by other market requirements, one needs to search for the right film to complete the MA pack.

Normally, film permeance is used as an input in the MAP model. However, the model formulation can be turned around and film permeance needed can be calculated based on the product's gas exchange characteristics, assuming some known optimum target MA conditions, and given the external storage conditions. As the storage and transport conditions throughout a chain will not be constant, this exercise should be repeated over a range of temperatures or a number of different temperature scenarios. MA conditions that are optimal at one temperature do not need to be optimal at another temperature. For instance, the tolerance to low oxygen levels decreases with increasing temperature.[59, 61]

Once models are available to describe the effect of humidity and condensation on film permeance, they can be used to predict humidity levels inside the package and to predict how film permeability is affected by this. This will help to set detailed specifications to films with regard to this aspect. This will be especially usefully in the MAP of minimally processed produce, soft fruits and leafy vegetables because of the high humidity levels occurring in these packages.

14.3.3 Optimising logistic chains

Given the ultimate MA package for a certain product, its eventual success mainly depends on the temperature control between the moment of packing and the moment of opening the package at the consumer's place. In a logistic chain where temperature is not controlled throughout, application of MAP is a waste of time, money and produce. Using a MAP model to simulate a package going through a logistic chain will give insight in the strong and weak parts of that chain.[15] It will make clear which parts of the chain are responsible for the largest quality losses of the packaged product and need improvement. It enables the optimisation of a whole chain considering the related costs and benefits.

To get the most out of such an exercise, a MAP model should be used that includes a keeping quality or quality change model. Assessing the benefits and losses in terms of product quality gives much more insight than just the observation that the MA conditions dropped below or above their target levels.

The question that should always be asked is how these deviations affect the quality and keeping quality. The product quality gives static information on the status of the product at a certain moment, for instance at the point of sale. Keeping quality provides dynamic information on how long a product can be stored, kept for sale, transported to distant markets or, after sale, remain acceptable at the consumer's place.

14.3.4 Sensitivity studies

Generally, an MA package is developed based on some average product characteristics, assuming an average amount of product packaged, using the specifications of an average sample of film and assuming the MA pack will be handled and stored at certain average conditions. As the average MA pack does not exist, the question arises how the non-average package will behave at non-average conditions and how far the MA conditions will diverge from the ideal target levels. Sensitivity studies are ideal to test how sensitive a system of MAP is to changes in one or more parameters or conditions. Using a MAP model, sensitivity studies can be conducted easily by running the model multiple times, using a range of values for the different conditions under study. This will help to identify which aspects of MAP should be controlled more strictly because of their potential impact on the system as a whole.

The results strongly depend on the MA pack under study. For instance, depending on the gas exchange rate the same change in packaged biomass will have different effects on the steady state MA conditions. So, it cannot be stated in general that MAP is insensitive to a change in biomass. Also, the gas conditions in an MA pack where film and produce have comparable temperature dependencies are insensitive to temperature. Using this same pack for packaging a produce with a different temperature dependency can result in gas conditions extremely sensitive to temperature. Even if the MA conditions in an MA pack are insensitive to temperature due to the balanced combination of film and product, this does not mean that the quality of the packaged product is insensitive to temperature. These are just two different ways of assessing MAP. If a good quality change model is lacking, the 'optimum MA conditions' are the only criteria to apply when judging MA packs. When a good quality change model is available, sensitivity studies can be performed on the really important product quality.

14.4 Pros and cons

Applying models to improve MAP has, like every technique, its pros and cons. Some of the advantages are already implicitly mentioned in the previous sections describing the areas of application. By applying models, the development phase of MAP can be shortened. Huge experiments can be done behind the laptop checking all possible situations that would take weeks to test

in practice. With a good conceptual model in mind and the mathematical equivalent at the fingertips, developing MAP can be lifted beyond the phase of '*pack-and-pray*'.

When developing a model, continued balancing should be going on between the completeness and relevance of the described phenomena and the level of detail and complexity of the model needed to realise this. For scientific purposes, the ultimate model would be a mechanistic one describing all relevant underlying processes. For practical purposes, one should start from such a detailed mechanistic approach and simplify as far as possible without affecting the explanatory power of the model for that specific dedicated application. In the practice of postharvest physiology, that detailed mechanistic model is not available and the best one can do is to develop a hypothetical mechanism in agreement with the observed phenomena and in agreement with current general physiological and biochemical concepts. Such a mechanistic model can still be extremely valuable to develop concepts by verifying or falsifying hypotheses. Developing, for instance, a quality change model forces the expert to formulate a conceptual model and to realise where the gaps in his knowledge are. This is probably the most valuable and general profit of developing mechanistic models as it enhances the understanding of a complex system and directs future research to fill the gaps.

In spite of the advantages, one should remain aware of some potential traps. One of them is the risk of forgetting about real life, simply because not everything goes according to the model. The books can prescribe transport at 2ºC but cannot prevent the driver from turning off the cooler unit when delivering early in the morning in an urban region, not wanting to wake up its inhabitants. Kiwifruit could last another week according to the developed firmness model, but in practice are already lost due to spoilage.

This brings us to the fact that you cannot get anything out of a model you did not include to start with. When condensation is not included it is impossible to assess sudden temperature drops on their potential to induce condensation with all the consequences for the omnipresent microbes. If a quality change model leans heavily on one single limiting quality attribute, the user of the model should be aware of specific situations turning another quality attribute into the limiting factor. At the same time it should be recognised that it is impossible to include everything into the model as it would be impossible to validate it completely.

To end with, one should always be alert when applying models outside the range they were validated. Especially in case of empirical models, this can result in unrealistic predictions. A model can easily process unrealistic data without getting into a moral conflict. The user should always stay alert to recognise such anomalies.

14.5 Future trends

Some of the trends needed to safeguard the future of MAP are very basic while others are on the level of refining existing knowledge. One of the most

embarrassing gaps in the current knowledge is a good database on permeance data of packaging films that includes their temperature dependency. The packaging film industry should develop a standard certificate for this that comes along with each film they produce for MAP.

To enable a fundamental approach to MAP the gas exchange of the different products should also be systematically characterised as a function of at least O_2, CO_2 and temperature. This knowledge, essential for the success of MAP, is still very fragmented.

To improve the models on MAP of minimally processed produce involving high humidity levels, a better understanding is needed of the effect of humidity and condensation on film permeances. Lots of work has still to be done to integrate the expertise from microbiology within the field of MAP.

Although models on gas exchange are becoming well established, models on how the physiology underlying quality is linked to the metabolism are not readily available. Their development is hampered by gaps in the knowledge of postharvest physiology. However, to assess MA packages on the quality of their actual turnout, quality change models are needed. The ultimate goal would be to develop generic models that can be validated to a wide range of commodities.

The last issue that needs to be covered in the near future is the characterisation of biological variation and its impact on product behaviour in general and on MAP in particular. Although this issue is important for the postharvest industry as a whole, MAP would benefit quite a lot from a more fundamental approach.

14.6 References

1. ROBERTSON D, AGUSTI J, *Software Blueprints: Lightweight Uses of Logic in Conceptual Modelling*, Addison Wesley Longman Inc., 1999.
2. TIJSKENS L M M, POLDERDIJK J J, 'A generic model for keeping quality of vegetable produce during storage and distribution', *Agricultural Systems*, 1996 **51** 431–52.
3. ZAGORY D, KADER A A, 'Modified atmosphere packaging of fresh produce', *Food Technology*, 1988 **42** 70–7.
4. GORRIS L G M, PEPPELENBOS H W, 'Modified atmosphere and vacuum packaging to extend the shelf life of respiring food products', *Horttechnology*, 1992 **2**(3) 303–9.
5. KADER A A, ZAGORY D, KERBEL E L, 'Modified atmosphere packaging of fruits and vegetables', *Crit. Rev. Food Sci. Nutr*, 1989 **28** 1–30.
6. KADER A A, 'Modified atmospheres during transport and storage', Chapter 11 in: A A Kader (Ed.), *Postharvest Technology of Agricultural Crops*, 2nd edition, Publication 3311, Universtity of California, California, USA, 1992.
7. VERMEIREN L, DE VLIEGHERE F, VAN BEEST M, DE KRUIJF N, DEBEVERE J, 'Developments in the active packaging of foods', *Trends in Food Science & Technology*, 1999 **10** 77–86.

8. KADER A A, 'A summary of CA requirements and recommendations for other fruits than pomme fruits', *6th Intl. Controlled Atmosphere Research Conference*, Ithaca, NY, 1993.
9. BRASH D W, CHARLES C M, WRIGHT S, BYCROFT B L, 'Shelf-life of stored asparagus is strongly related to postharvest respiratory activity', *Posth. Biology Technol.*, 1995 **5** 77–81.
10. BURTON W G, 'Biochemical and physiological effects of modified atmospheres and their role in quality maintenance', in: H O Hultin and M Milner (Eds), *Postharvest Biology and Biotechnology*, FNP, Westport, CT, 1978.
11. GIESE J, 'How Food Technology covered modified atmosphere packaging over the years', *Food Technology*, 1997 **51**(6) 76–7.
12. CLELAND A C, *Food Refrigeration Processes – Analysis, Design and Simulation*, London, Elsevier Science Publishers, 1990.
13. GAFFNEY J J, BAIRD C D, CHAU K V, 'Methods for calculating heat and mass transfer in fruits and vegetables individually and in bulk', *Trans. ASHRAE*, 1985 91(2B) 333–52.
14. RATTI C, RAGHAVAN G S V, GARIÉPY Y, CRAPISTE G H, 'Long term storage of vegetables under modified atmospheres', Paper 96F-040, *AgEng*, Madrid, 1996.
15. HERTOG M L A T M, PEPPELENBOS H W, TIJSKENS L M M, EVELO R G, 'Modified atmosphere packaging: optimisation through simulation', in: J R Gorny (Ed.), *Proc. 7th Intl. Controlled Atmosphere Research Conference*, Davis, CA, 1997.
16. BEAUDRY R, LAKAKUL R, 'Basic principles of modified atmosphere packaging', *Washington State University Tree Fruit Postharvest Journal*, 1995 **6**(1) 7–13.
17. PEPPELENBOS H W, *The use of gas exchange characteristics to optimize CA storage and MA packaging of fruits and vegetables*, Thesis Wageningen Agricultural University, ISBN 90-5485-606-8, 1996.
18. BLACKMAN F F, *Analytic Studies in Plant Respiration*, Cambridge University Press, 1954.
19. VAROQUAUX P, ALBAGNAC G, NGUYEN C, VAROQUAUX F, 'Modified atmosphere packaging of fresh beansprouts', *J. Sci. Food Agric.*, 1996 **70** 224–30.
20. BURG S A, BURG E A, 'Gas exchange in fruits', *Physiol. Plant.*, 1965 **18** 870–84.
21. CAMERON A C, YANG F S, 'A simple method for the determination of resistance to gas diffusion in plant organs', *Plant Physiol.*, 1982 **70** 21–3.
22. CHAU K V, ROMERO R A, BAIRD C D, GAFFNEY J J, 'Transpiration coefficients for certain fruits and vegetables', *Trans. ASHRAE*, 1988 **94** 1553–62.
23. DE VRIES H S M, WASONO M A J, HARREN F J M, WOLTERING E J, VAN DER VALK H C P M, REUSS J, 'Ethylene and CO_2 emission rates and pathways in harvested fruits investigated, in situ, by laser photothermal deflection and photoacoustic techniques', *Posth. Biology Technol.*, 1996 **8** 1–10.
24. BANKS N H, DADZIE B K, CLELAND D J, 'Reducing gas exchange of fruits with surface coatings', *Posth. Biology Technol.*, 1993 **3** 269–84.

25. MAGUIRE K M, BANKS N H, OPARA L U, 'Factors affecting weight loss of apples', *Horticultural Reviews*, 2001 **25** 197–234.
26. KANELLIS A K, *Biology and Biotechnology of the Plant Hormone Ethylene*, Dordrecht, Boston: Kluwer Academic Publishers, 1997.
27. BEEVER D J, HOPKIRK G, 'Fruit development and fruit physiology', in: I J Warrington and G C Weston (Eds), *Kiwifruit, Science and Management*, Ray Richards publisher in association with NZSHS, Auckland, New Zealand, 1990.
28. SLOOF M, TIJSKENS L M M, WILKINSON E C, 'Concepts for modelling the quality of perishable products', *Trends in Food Science & Technology*, 1996 **7**(5) 165–71.
29. BENNIK M H J, *Biopreservation in modified atmosphere packaged vegetables*, Thesis Wageningen Agricultural University. ISBN 90-5485-808-7, 1997.
30. CHAMBROY Y, GUINEBRETIERE M H, JACQUEMIN G, REICH M, BREUILS L, SOUTY M, 'Effects of carbon dioxide on shelf-life and post harvest decay of strawberries fruit', *Sciences des Aliments*, 1993 **13** 409–23.
31. TIJSKENS L M M, KONOPACKI P, SIMCIC M, HRIBAR J, 'Biological variance in agricultural products. Colour of apples as an example', IIR Conference: *Improving Postharvest Technologies of Fruits, Vegetables and Ornamentals*, 19–21 October 2000, Murcia, Es. (in press).
32. MAGUIRE K M, BANKS N H, 'Harvest date, cultivar, orchard, and tree effect on water vapour permeance in apples', *J. Amer. Soc. Hort. Sci.*, 2000 **125** 100–4.
33. VAN DER SMAN R G M, *Latice Boltzmann schemes for convection-diffusion phenomena; application to packages of agricultural products*, Thesis Wageningen, ISBN 90-5808-048-X, 1999.
34. DOYON G, GAGNON J, TOUPIN C, CASTAIGNE F, 'Gas transmission properties of Polyvinyl Chloride (PVC) films studied under subambient and ambient conditions for modified atmosphere packaging applications', *Packaging Technology and Science*, 1991 **4** 157–65.
35. EXAMA A, ARUL J, LENCKI R W, LEE L Z, TOUPIN C, 'Suitability of plastic films for modified atmosphere packaging of fruits and vegetables', *J. Food Sci.*, 1993 **58**(6) 1365–70.
36. FISHMANN S, RODOV V, BEN-YEHOSHUA S, 'Mathematical model for perforation effect on oxygen and water vapor dynamics in modified-atmosphere packages', *J. Food Sci.*, 1996 **61**(5) 956–61.
37. YANG C C, CHINNAN M S, 'Modeling the effect of O_2 and CO_2 on respiration and quality of stored tomatoes', *Transactions of the ASAE*, 1988 **31** 920–5.
38. MANNAPPERUMA J D, SINGH R P, 'Modelling of gas exchange in polymeric packages of fresh fruit and vegetables', in: R P Singh, F A R Oliveira (Eds), *Process Optimisation and Minimal Processing of Foods*, CRC Press, Boca Raton, 1994.
39. CHEVILLOTTE P, 'Relation between the reaction cytochrome oxidase-oxygen and oxygen uptake in cells in vivo. The role of diffusion', *J. Theor. Biol.*, 1973 **39** 277–95.

40. LEE D S, HAGGAR P E, LEE J, YAM K L, 'Model for fresh produce respiration in modified atmospheres based on principles of enzyme kinetics', *J. Food Sci.*, 1991 **56** 1580–5.
41. RATTI C, RAGHAVAN G S V, GARIÉPY Y, 'Respiration model and modified atmosphere packaging of fresh cauliflower', *J. Food Engin.*, 1996 **28** 297–306.
42. SONG Y, KIM H K, YAM K L, 'Respiration rate of blueberry in modified atmosphere at various temperatures', *J. Amer. Soc. Hort. Sci.*, 1992 **117** 925–9.
43. CAMERON A C, BEAUDRY R M, BANKS N H, YELANICH M V, 'Modified-atmosphere packaging of blueberry fruit: Modeling respiration and package oxygen partial pressures as a function of temperature', *J. Amer. Soc. Hort. Sci.*, 1994 **119** 534–9.
44. HERTOG M L A T M, PEPPELENBOS H W, EVELO R G, TIJSKENS L M M, 'A dynamic and generic model on the gas exchange of respiring produce: the effects of oxygen, carbon dioxide and temperature', *Posth. Biology Technol.*, 1998 **14** 335–49.
45. TIJSKENS L M M, 'A model on the respiration of vegetable produce during postharvest treatments', in: G R Fenwick, C Hedley, R L Richards, and S Khokhar (Eds), *Agri-food Quality, an Interdisciplinary Approach*, The Royal Society of Chemistry, Cambridge, UK, 1996.
46. MAKINO Y, IWASAKI K, HIRATA T, 'Oxygen consumption model for fresh produce on the basis of adsorbtion theory', *Transactions of the ASAE*, 1996 **39**(3) 1067–73.
47. MAKINO Y, IWASAKI K, HIRATA T, 'A theoretical model for oxygen consumption in fresh produce under an atmosphere with carbon dioxide', *J. Agric. Engng Res.*, 1996 **65** 193–203.
48. ANDRICH G, ZINNAI A, BALZINI S, SILVESTRI S, FIORENTINI R, 'Modelling of aerobic and anaerobic respiration rates of golden delicious apples as a function of storage conditions', *Acta Horticulturae* 1998 **476** 105–12.
49. BANKS N H, CLELAND D J, YEARSLEY C W, KINGSLEY A M, 'Internal atmosphere composition – a key concept in responses of fruits and vegetables to modified atmospheres', *Proceeding of the Australasian Postharvest Conference*, 1993.
50. MAGUIRE K M, LANG A, BANKS N H, HALL A, HOPCROFT D, BENNET R, 'Relationship between water vapour permeance of apples and micro-cracking of the cuticle', *Posth. Biology Technol.*, 1999 **17** 89–96.
51. MAGUIRE K M, BANKS N H, LANG A, 'Sources of variation in water vapour permeance of apple fruit', *Posth. Biology Technol.*, 1999 **17** 11–17.
52. TIJSKENS L M M, EVELO R G, 'Modelling colour of tomatoes during postharvest storage', *Posth. Biology Technol.*, 1994 **4** 85–98.
53. SCHOUTEN R E, VAN KOOTEN O, 'Keeping quality of cucumber batches: is it predictable?', *Acta Horticulturae*, 1998 **476** 349–55.
54. POLDERDIJK J J, BOERRIGTER H A M, TIJSKENS L M M, 'Possibilities of the model on keeping quality of vegetable produce in controlled atmosphere

and modified atmosphere applications', *Proceedings of the 19th International Congress of Refrigeration*, volume II, 1995.

55. HERTOG M L A T M, BOERRIGTER H A M, VAN DEN BOOGAARD G J P M, TIJSKENS L M M, VAN SCHAIK A C R, 'Predicting keeping quality of strawberries (cv. 'Elsanta') packaged under modified atmospheres: an integrated model approach', *Posth. Biology Technol.*, 1999 **15** 1–12.
56. TIJSKENS L M M, VAN SCHAIK A C R, HERTOG M L A T M, DE JAGER A, 'Modelling the firmness of Elstar apples during storage and transport', *Acta Horticulturae*, 1997 **485** 363–71.
57. HERTOG M L A T M, TIJSKENS L M M, 'Modelling modified atmosphere packaging of perishable produce: keeping quality at dynamic conditions', *Acta Alimentaria*, 1998 **27** 53–62.
58. WIJTZES T, *Modelling the microbial quality and safety of foods*, Thesis Wageningen Agricultural University, ISBN 90-5485-578-9, 1996.
59. YEARSLEY C W, BANKS N H, GANESH S, 'Temperature effects on the internal lower oxygen limits of apple fruit', *Posth. Biology Technol.*, 1997 **11** 73–83.
60. CAMERON A C, PATTERSON B D, TALASILA P C, JOLES D W, 'Modeling the risk in modified-atmosphere packaging: a case for sense-and-respond packaging', in: G D Blanpied, J A Bartsch and J R Hicks (Eds), *Proc. 6th Intl. Controlled Atmosphere Research Conference*, Ithaca, NY, 1993.
61. BEAUDRY R M, CAMERON A C, SHIRAZI A, DOSTAL-LANGE D L, 'Modified-atmosphere packaging of blueberry fruit: effect of temperature on package O_2 and CO_2', *J. Amer. Soc. Hort. Sci.*, 1992 **117**(3) 436–41.

15

Modelling thermal processes: cooling and freezing

Q. T. Pham, University of New South Wales, Sydney

15.1 Introduction

Refrigeration is the most widely used method for preserving the quality of fresh foods, especially meat, fish, dairy products, fruit, vegetables and ready-made meals. Refrigeration can be costly in terms of equipment and energy, and if not done correctly will fail to achieve its objectives and lower the quality and safety of the product. To ensure that refrigeration is effective, we need to be able to calculate processing times, product temperatures, heat loads and water diffusion into and out of the product. These are influenced by both operating conditions (environmental temperature, humidity, air velocities, etc.) and product parameters (type, size, shape, composition). Not surprisingly a lot of research has been carried out to quantify these relationships.

Early works on food refrigeration models concentrated on the prediction of freezing or chilling time based on the mean or centre temperature of the food. Plank's[1] simplified equations for the freezing time of foods were presented in 1913, and more recent efforts have concentrated on removing his rather limiting assumptions on the properties and geometry of the product. Other workers developed models to predict moisture loss from food undergoing refrigeration. For a long time, processing time and moisture loss were the major concern of researchers. With the advent of computers, numerical methods were introduced into food research and the whole temperature field inside the product could be modelled. It was quickly realised that this would enable the prediction of many other quantities of interest related to the quality of the food such as shelf-life, microbial growth, appearance, flavour and texture. At the same time, the use of computers in plant design stimulated research into the prediction of dynamic heat load from food, and integrated models are being developed to take into

account the interactions of the food with the processing equipment.[2] More powerful computers enable the effect of variations in the product and the environment to be explored.

There is not enough space in this chapter to cover all the developments in the field, so a decision has been made to concentrate on some important problems that have occupied researchers in recent years. Our concern will be with heat and mass transfer aspects only, and the modelling of product quality will not be looked at; it can be assumed that chemical and biological changes that happen concurrently do not significantly affect temperature and moisture changes in food, at least at low temperatures (clearly this would not be the case in, say, the cooking of meat or the baking of bread), so a quality model can be readily 'grafted' on a heat and mass transfer model without changing the latter. A large amount of work has been done on the development of simple analytical models for predicting cooling and freezing times: these will also not be covered here, since there have been several reviews of this topic.[3–5]

Even with these self-imposed limitations, the simulation of food cooling and freezing is a vast area with many problems and challenges for the modeller. Calculation of the heat removal process may be complicated by phase change, during which the product's thermal properties undergo large changes over a small temperature range. The product undergoing cooling often has very complex shape and composition. Heat transfer may be coupled with moisture transfer and the equations governing the two processes should be solved simultaneously. The heat transfer coefficient is often difficult to determine for the infinite variety of real-life situations such as packaged products, cryogenic cooling, highly turbulent flow, swirling and non-parallel flow, etc., and within the same chiller the heat transfer coefficient will vary greatly from place to place. The heat transfer coefficient may also vary along the surface of a product due to the complex turbulent flow pattern and the development of the flow boundary layer, a problem still largely unsolved even by the most advanced computational fluid dynamics packages. Food products often have inconsistent compositions, shapes and sizes, which lead to variable thermal behaviour and quality after processing, so that it is difficult to quantify accurately the effects of different processing practices.

Refrigeration system designers want fast, reasonably accurate product load calculation methods as their computers may have to carry out this calculation hundreds or thousands of times in a simulation or optimisation task. For such purposes, simplified methods using ordinary differential equations (ODEs) have been developed. In other situations, an ability to predict accurately the whole temperature and moisture fields in a product is desirable as it would allow the food technologist to optimise the economic and product quality factors. Thus, microbial growth, physical change (such as weight loss), biochemical changes (such as those determining the tenderness of meat or ripeness and flavour of fruit) and subjective factors (such as surface appearance) are often highly sensitive to temperature and moisture changes. Such a predictive ability is as yet beyond us, although steady progress is being made with recent advances in computer software and hardware.

The variety of problems faced by the modeller is too large to be adequately treated in this chapter, therefore we will concentrate on only a few specific aspects that are of particular importance to the modeller and/or the industrial user of simulation software, and discuss these topics in some depth. The chapter starts with methods for modelling foods with complex shapes, from highly simplified models to detailed numerical models. It goes on to discuss models that are used to predict the dynamic heat load during chilling and freezing. In many situations involving fresh foods, water loss (or gain) will accompany heat transfer, and models for these situations will be described. Underlying all models is the need for accurate estimation of transfer coefficients for heat and mass, which will be discussed. A few typical industrial applications of modelling are described, involving the use of models in plant control and in process optimisation. Finally, the areas that most need further investigation are identified and discussed.

15.2 Modelling product heat load during cooling

In order to design an efficient refrigeration system, product heat load must be known. With batch processes, the heat load varies strongly with time, displaying a peak at the beginning of cooling and whenever there is a sharp drop in environment temperatures. It is important to be able to calculate the variations in heat load with time, since overdesign is inefficient and costly, while underdesign may cause hygiene problems and failure to comply with specifications.

A refrigeration plant may process a large number of products of different shapes, sizes and compositions, entering the chillers or freezers at different times and under different conditions. The system may need to be optimised, which involves repeated calculations. The refrigeration system designer needs a simple method that can be applied repeatedly to calculate the product heat load. The rigorous solution to the heat transfer equation for a solid body, even in the absence of any complication such as variable thermal properties or complex geometry is an infinite series expression which is unsuitable for routine use. Various attempts have therefore been made to derive simpler and faster approximate solutions.

Ordinary differential equations (ODE) models, lumped parameter models or stirred tank models are those where an object undergoing thermal changes is represented by a small number of components, each of which is at uniform temperature. Because of the simplification of the physical situation that this entails, ODE models usually have to incorporate empirical parameters to improve prediction accuracy. ODE models are often used to calculate the variation of product heat loads with time because they are much faster than more rigorous partial differential equation (PDE) models such as finite differences. ODE models are *not* suitable for product temperature calculations, except in the case of very low Biot numbers (uniform product temperature).

15.2.1 The single tank model

The simplest ODE model is obtained by considering the object being cooled as a 'stirred tank' at uniform temperature. This is permissible when the Biot number, which measures the ratio of internal to external resistance to heat transfer, is small ($\mathrm{Bi} \ll 1$), resulting in a nearly uniform internal temperature. Since most foods have medium-to-low thermal conductivity (of the order $1\,\mathrm{Wm^{-1}K^{-1}}$ or less), the above condition is satisfied only for small pieces of food (a few mm across) being cooled slowly, smallish-sized food wrapped in insulating packaging, or liquid food undergoing stirring.

15.2.2 The tank network model

A slightly more complex model is the tank network model[6] in which a solid food of complex shape such as a beef leg or carton undergoing chilling is represented by a set of stirred tanks connected by heat conductances (Fig. 15.1). The heat flows between the tanks and from each tank to the environment are proportional to the respective temperature differences, while within each tank the temperature is uniform. The thermal conductances of the heat transfer paths and thermal capacities of the tanks are determined by curve-fitting empirical or numerically generated data. For each tank i the temperature T_i is calculated from

$$m_i c \frac{dT_i}{dt} = K_{i,a}(T_a - T_i) + \sum_{j \neq i} K_{i,j}(T_j - T_i) \tag{15.1}$$

where K_{ij} is the thermal conductance (in $\mathrm{WK^{-1}}$) between tanks j and i and $K_{i,a}$ that between tank i and the environment.

With n tanks there are $n + (n-1) + \ldots + 1 = n(n+1)/2$ thermal resistances and n thermal capacities, so a small number of tanks (two or three) is sufficient in most cases to correlate the data. Energy conservation requires that the sum of thermal capacities must equal the total thermal capacity of the physical object, which is usually known, hence in practice the number of curve-fitted variables is reduced by one.

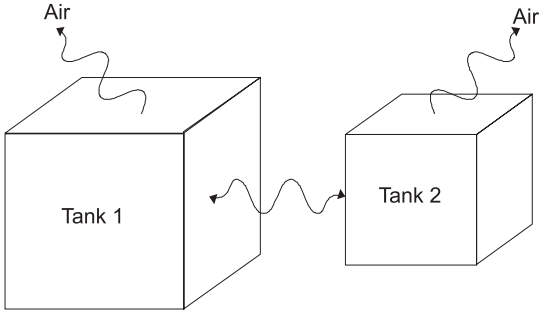


Fig. 15.1 Tank network model with two tanks.

Superficially the tank network model resembles finite volume or finite difference models which are discretisations of the partial differential equations. However, an important difference is that while the parameters of the finite difference model are derived from the physical situation, each node representing a specific location within the object, no such physical association is assumed in the tank network model. Also, while each node of a finite difference model exchanges heat only with its neighbours, each tank in the network model exchanges heat with all other tanks and the environment. The model is reminiscent of neural networks used in artificial intelligence.

To test the adequacy of the tank network model, Davey and Pham[6] partitioned their beef carcass chilling dataset into two groups, using the first group to find the model's parameters and regress them against the operating variables (carcass size, fat cover, air velocity), then using the resulting model to predict the dynamic heat load in the second group of data. They found that with as few as two tanks (four curve-fitted parameters) the predictive accuracy was very good (average error 1.5%, well within experimental uncertainty) and significantly better than either their finite difference model or finite element model. The model was also successful in predicting the effects of variations in environmental temperature (which is as expected, since these effects were linear) and heat transfer coefficients (which was somewhat more surprising).

15.3 Modelling product heat load during freezing

The ODE derived by Plank[1,7] for calculating the freezing front position in simple shapes can be used as a basis for calculating dynamic product load in more complex shapes. Cleland[8] proposed a single parameter model based on Plank's equation that will predict the dynamic heat load during freezing. This model does not adequately take into account the change in shape of the freezing front during the freezing of an irregular shaped object. Lovatt *et al.*'s[9] two-parameter model is the most accurate to date, and a slightly simplified version of this model (obtained by lumping the freezing and post-cooling heat load together) will be presented here. By considering the velocity of the freezing front and associated latent heat release, an approximate equation for heat load is obtained:

$$Q = -\Delta H_2 \frac{dr_f}{dt} \frac{dV_f}{dr_f} \tag{15.2}$$

where ΔH_2 is the latent plus post-cooling heat, dr_f/dt is the freezing front velocity and dV_f/dr_f the change of frozen volume with frozen depth. dr_f/dt can be derived from

$$\frac{dr_f}{dt} = \frac{-(T_a - T_f)}{\Delta H_2 {r_f}^{E-1} \left[\dfrac{1}{hR^{E-1}} - \dfrac{{r_f}^{2-E} - R^{2-E}}{k_f(2-E)} \right]} \tag{15.3}$$

where R is the product's smallest half-dimension, h the heat transfer coefficient, k_f the frozen thermal conductivity and E (the EHTD) is defined as the ratio of the freezing time of an infinite slab of same thickness as the smallest dimension of the object to the freezing time of the object. dV_f/dr_f is a function of geometry and can be calculated from

$$\frac{dV_f}{dr_f} = Z\frac{V}{R}\left(\frac{r_f}{R}\right)^{Z-1} \tag{15.4}$$

where V is the product volume and Z a geometric parameter. Equation (15.2) can be shown to hold rigorously for the three basic shapes (infinite slabs, infinite cylinders and spheres) when sensible heat effects are neglected, and will also hold for ellipses and ellipsoids if the freezing front remains self-similar. For simple shapes, E and Z are given by:

slab: $E = Z = 1$

infinite cylinder: $E = Z = 2$

sphere: $E = Z = 3$

For other shapes, there has been numerous analytical and empirical formulae proposed for E.[5] Z can be found from

$$Z \approx AR/V \tag{15.5}$$

For very irregular shapes with protrusions AR/V may be larger than 3, which would give unrealistic results, hence the following upper limit is imposed:

$$Z \leq 3 \tag{15.6}$$

Lovatt *et al.*'s method has been tested against finite difference computations and experimental data for the freezing of cartons and lamb carcasses. Agreement to within about 10% is obtained for most of the cooling process (and certainly during the first half, where most of the heat release occurs).

15.4 Modelling foods with complex shapes

Most heat transfer modelling works in food refrigeration have concentrated on the cooling of solid foods with complex shapes, in particular meat products and to a lesser extent horticultural products. Animal carcasses or parts thereof have some of the most complex shapes that have been modelled, not excluding airplanes, boats or cars. The first task of the modeller is therefore to decide how to take into account these shapes. In the early days when computing power was limited and the main preoccupation was to bring the thermal centre down to a given temperature, an equivalent shape approach was the most common – thus an animal carcass or leg could be replaced by an equivalent slab, cylinder or sphere whose dimension was a function of product weight.[10, 11] An alternative approach is to use empirical geometric factors, by which the cooling or freezing

time must be multiplied. The most well known of these geometric factors is the equivalent heat transfer dimensionality (EHTD) introduced by Cleland and co-workers,[12] which was given a rigorous analytical basis by McNabb *et al.*[13, 14]

In modelling a complex shape by a simpler shape, an *equivalent size* for the latter must be defined. The equivalent size can be worked back from the EHTD, for which analytical and empirical formulae are available for a wide variety of situations.[3, 15] The application of this method requires complete information on the geometry of the object being modelled. Crocombe *et al.*[16] reported a laser scanner with integrated software to calculate geometric factors and EHTD directly from scanned data. Alternatively the equivalent size can be found directly from cooling data by an error minimisation method. Pham and Coulter[17] used direct curve fitting of temperature data for the centre and surface of pig legs to calculate their equivalent diameters, using an evolutionary error minimisation method. Because the differences between surface and centre temperature are highly dependent on the product's size, the equivalent diameter could be obtained with precision to give a model with very good prediction accuracy, partly because errors in the other inputs (heat transfer coefficients, thermal properties, initial temperature profile) are automatically compensated for.

To obtain more accurate predictions, the equations of change for heat transfer, mass transfer and fluid flow must be solved by a numerical method. The most popular methods for the rigorous solution of the Fourier heat conduction equation are finite difference and finite element, while the finite volume method is widely used in computational fluid dynamics code. To assist in the understanding of the modifications necessary to deal with freezing problems, a brief, user-oriented review of these methods will be given (a full description of the methods and their theoretical bases can be found in Chapter 4).

The *finite difference method* represents temperatures in an object by a set of nodal temperatures. For example, in one dimension, the Fourier conduction equation

$$\rho c \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(k \frac{\partial T}{\partial x}\right) \tag{15.7}$$

is discretised into

$$\rho c (T_i^{New} - T_i^{Old}) = \frac{\Delta t}{\Delta x^2}[k_-(T_{i-1}^m - T_i^m) + k_+(T_{i+1}^m - T_i^m)] \tag{15.8}$$

where the subscript i denotes the ith node (or control volume) and the superscript denotes the time level (*Old* is at beginning of a time step, *New* is at the end, m is some intermediate value). ρ denotes the density, c the specific heat, k the thermal conductivity, k_+ its value between nodes i and $i+1$, k_- that between nodes i and $i-1$. The temperatures T^m on the right-hand side represent some weighted average nodal temperatures over the time interval, and different *stepping schemes* have been used with different weighting methods. Two of the most common are the explicit or Euler scheme, which calculates T^m at the beginning of the time interval ($T_i^m = T_i^{Old}$), and the Crank-Nicolson or central

difference scheme, which uses the mean between old and new temperatures ($T_i^m = (T_i^{Old} + T_i^{New})/2$).

The *finite element method* differs from finite difference in one major respect: whereas finite difference assumes that thermal capacities are lumped at the nodes, in finite element the thermal capacity is normally distributed over the volume according to some weighting rule. A finite element model starts by division of the object into elements, each containing a number of nodes (each vertex is a node, but there may be more). Heat conduction within each element is first considered separately to obtain a set of equations relating the temperatures of all the nodes belonging to the element. The procedure is repeated for all elements and the equations obtained are assembled into a matrix equation for the nodal temperature vector **T**:

$$\boldsymbol{C}d\boldsymbol{T}/dt = \boldsymbol{K}\boldsymbol{T}^m + \boldsymbol{f}^m \tag{15.9}$$

$\boldsymbol{C}$ is called the global capacitance matrix, $\boldsymbol{K}$ the global conductance (or stiffness) matrix and $\boldsymbol{f}$ the thermal load vector. The above matrix equation is then solved by a matrix solving procedure.

The *finite volume method* is widely used nowadays in commercial computational fluid dynamics (CFD) codes. In this method the conservation equations are solved over finite control volumes of arbitrary shapes. The volume integrals are approximated in terms of the volume-centred value of the field variables, while surface integrals are calculated by some averaging method. The final set of discrete equations to be solved are similar to those obtained for finite difference and finite element methods and solved in the same way.

To model simultaneously product temperature, heat load and weight loss during beef chilling, Davey and Pham[18] used a multi-region finite difference approach. A beef side was divided into seven regions: leg shank, rump, loin, ribs, shoulder, foreleg and neck, each represented by a simple shape such as an infinite cylinder or slab, with dimensions found from a regression equation. With such a technique it was possible to calculate the heat load during the first two hours of a twenty-hour chilling cycle with an average error of 12.6% and the weight loss with an average relative error of 1.25%.

Finite element and unstructured mesh finite volume techniques can be used to model any shape, but two-dimensional (2D) models are much easier to handle than three-dimensional (3D) models, which demand special graphics software and a large amount of time and effort. 2D models have been used for beef sides[19–21] and lamb loins.[22] In such models, heat conduction along the third dimension is ignored, a valid approximation for elongated shapes. Full 3D cooling models have been reported for lamb,[23] pork,[24] bread and turkey,[25] and beef side.[26] The recent advent of CFD packages and huge computer memories have allowed very detailed 3D representations of complex shapes: while a maximum grid size of 800 pseudo-cubic elements was specified for the PC-version of the BERTIX and BAKTIX programs in 1993, a 100 000 node finite volume representation for a beef side was recently constructed using the Gambit software on a PC.[26]

In spite of the ability to model in 3D, it is likely that 1D and 2D models will remain popular for some years to come for industrial simulation and for many research applications. The number of dimensions required to represent an object can often be reduced by using axial or spherical symmetry, in combination with appropriate approximations. For elongated shapes, longitudinal conduction and end effects may be neglected. The errors involved in such approximations are often smaller than those caused by uncertainties and statistical variations in shape, size and operating conditions. It is likely that full 3D models will be used mostly as references against which simpler models can be compared and calibrated (of course, the 3D models must themselves be thoroughly checked against experiment).

15.5 Numerical solution of the heat conduction equation with phase change

Special problems arise in the application of finite difference and finite element methods to freezing and thawing, due to the release of latent heat over a small range of temperatures (Fig. 15.2) and to a step change in the thermal conductivity over the same range (Fig. 15.3). These changes make the problem highly non-linear and special techniques are required for their solution.

Variations in thermal conductivity can be dealt with in a mathematically rigorous manner in both finite difference and finite element methods by using the Kirchhoff transformation[27]

$$\phi = \int^{T} k, dT \tag{15.10}$$

The heat conduction equation then becomes

$$\rho c \frac{\partial T}{\partial t} = \frac{\partial^2 \phi}{\partial x^2} \tag{15.11}$$

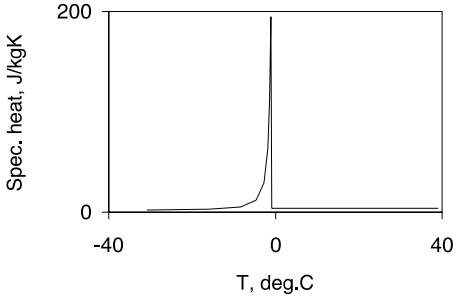


Fig. 15.2 Apparent specific heat of typical water-rich food around freezing point.

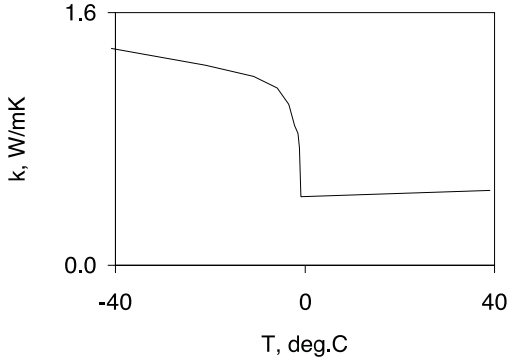


Fig. 15.3 Thermal conductivity of typical water-rich food around freezing point.

In spite of its mathematical elegance, the Kirchhoff function formulation does not guarantee better accuracy than, say, the simple approximation of an arithmetic mean *k*-value, i.e.

$$k_{ij} = [k(T_i) + k(T_j)]/2 \tag{15.12}$$

This is because there is no guarantee that the Kirchhoff function is linear when there is a phase change front in the vicinity – in fact, if sensible heat effects are neglected then the temperature profile will be a broken line and so will the Kirchhoff function's profile. Since the finite difference procedure (and the simplest version of finite element) implicitly assumes a linear Kirchhoff function profile between nodes, it will cause an error. This error is reduced to acceptable levels by using small nodal spacings (normally, ten to twenty nodes over each dimension of the object).

15.5.1 Treatment of phase change in finite difference models

The steep change in enthalpy v. temperature, which can also be visualised as a sharp, narrow peak in the effective specific heat at the freezing point, is a more difficult problem. Two main approaches have been proposed to take into account the phase change: front tracking methods and fixed grid methods. In front tracking methods, a node or a control volume boundary is placed at the phase change front, and the temperature fields in the frozen and unfrozen regions are solved separately (but connected by thermal equilibrium at the front) while the movement of the front is calculated from a heat balance. As the front moves so does the grid. In fixed grid methods, the front is not explicitly considered; if its position is required, it is calculated from the nodal temperatures or enthalpies. Front tracking methods are difficult to implement in complex shapes and in any case for foodstuffs there may not be a clear 'phase change front', since latent heat is released over a temperature range. It has also

been shown[28, 29] that fixed grid methods can give good prediction of front location. Therefore, the rest of this discussion will be concerned with fixed grid methods only.

A large amount of latent heat is released over a small temperature range during freezing. This latent heat can be treated either as a heat source, which has to be accounted for separately, or as a peak in the 'apparent specific heat' curve (Fig. 15.2). With either approach, it is easy to miss this peak when a nodal temperature crosses the freezing point too quickly. An effective remedy is the use of the enthalpy method, in which the rate change of nodal enthalpy is the calculated variable:

$$\frac{\partial H}{\partial t} = \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) \tag{15.13}$$

where

$$H = \int_{T_{ref}}^{T} \rho c.dT \tag{15.14}$$

Each calculation cycle yields a set of nodal enthalpies H_i from which the nodal temperatures T_i are back-calculated (or simultaneously calculated in an iterative procedure), thus it is impossible to 'jump' over the latent heat peak. Equation (5.13) can be solved explicitly using the Euler method[30] or implicitly. If an implicit stepping method is used, both unknown nodal temperatures T_i and nodal enthalpies H_i appear in the equation, hence an iteration has to be performed at each time step until H_i and T_i agree.

To avoid the need for iteration, Pham[28] proposed the 'temperature-enthalpy correction method', a hybrid temperature-enthalpy method (Fig. 15.4). At each

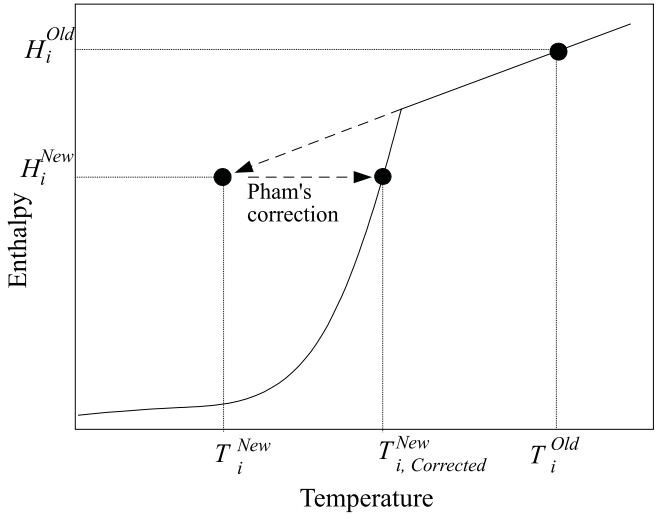


Fig. 15.4 Illustration of the temperature-enthalpy correction method.[28]

time step, the conduction equation (expressed in terms of $\partial T_i/\partial t$) is solved by an implicit method such as Crank-Nicolson (the original paper used Lees' three-level method) in the usual manner. The heat gained by the node is assumed to be correct, hence the new nodal enthalpy is

$$H_i^{New} = \rho c_i (T_i^{New} - T_i^{Old}) + H_i^{Old} \tag{15.15}$$

and the resulting nodal temperature array is corrected according to

$$(T_i^{New})_{Corrected} = T(H_i^{New}) \tag{15.16}$$

In Equation (15.15), c_i is the apparent specific heat used for the node during the temperature calculation. (Pham also suggested a method to get a good first estimate of this nodal apparent specific heat, but this was later found not to have a large bearing on the method's speed.) The temperature correction step could be considered as the first step in an iterative solution, but is sufficient on its own to bring about a dramatic improvement in accuracy and/or computing time compared to previous methods.[31]

15.5.2 Treatment of phase change in finite element

In contrast with finite difference, finite element methods make use of some distribution (weight) function to calculate the nodal capacitances: the heat gain over each time interval $\rho c \Delta T dv$ of each infinitesimal volume in the element is distributed to the nodes of that element, node i receiving a fraction w_i which depends on the location of the infinitesimal volume. To calculate the distribution of the heat gain over the whole element requires a numerical integration of $c\Delta T$ over the same area, which can be performed only very roughly (using a small number of sampling points). It can be seen that if a freezing front is passing through the element, the integration-by-sampling procedure can easily miss it.

To circumvent this difficulty, Pham[29] suggested the use of lumped capacitances, in which all the thermal capacitance of the element is concentrated at the nodes, rather than being distributed over the element. (For 'simplex' elements, i.e. triangles in 2D or tetrahedrons in 3D, this simply means attributing the thermal capacitances equally to each vertex.) This enables the latent heat peak problem to be dealt with in the same manner as in the finite difference method, for example by the use of the explicit enthalpy method or Pham's temperature-enthalpy correction method. Integration of the heat gain over the element is no longer necessary. Subsequently Comini *et al.*[32] showed that Pham's temperature-enthalpy correction method can also be applied to distributed-capacitance (i.e. conventional) finite elements, although the physical interpretation is less clear.

Using well-known test problems, Pham[33] compared ten of the most advanced fixed-grid finite element methods to date (after eliminating dozens of others) in terms of accuracy (agreement with analytical solution where available), time interval for convergence to within 1% of the converged solution (i.e. the solution obtained as $\Delta t \to 0$), heat balance error (percentage difference between heat

flows through boundaries and total heat gain of product – a measure of whether the latent heat load peak has been missed), and computing time as measured by the number of matrix inversions required. The test problems use both materials with a sharp phase change (heat released over 0.01 K) and a material with food-like properties. He concluded that the (non-iterative) lumped capacitance method with Pham's temperature-enthalpy correction performed best in terms of most of the above criteria. A further advantage of non-iterative methods is that the heat balance, which is easily checked, can serve as a useful indication of whether the time step is sufficiently small: a heat balance error of less than 1% generally indicates that convergence has been reached. Iterative methods tend to be time-consuming, and because they ensure a good heat balance at all time steps, this cannot be used as a check on accuracy or convergence to the low-Δt limit.

15.6 Modelling combined heat and mass transfer

Simultaneous heat and mass transfer arises when there is water movement as well as heat flow between product and surroundings. This frequently occurs in the cooling and freezing of unwrapped or loosely wrapped products such as meat, fruit and vegetables, where evaporative cooling can be similar or larger than sensible heat transfer, and also during immersion or spray cooling/freezing of unwrapped product when water or solutes diffuse from the freezing medium into the product.

During cooling, water evaporates from the surface, causing it to dry and water to diffuse from inside the product. If freezing occurs, internal diffusion stops, but the ice at or near the surface will sublime or 'freeze-dry' and the ice front gradually recedes into the product. These phenomena have important practical implications.

- Temperature changes and product heat loads are strongly affected by evaporative cooling effects. Indeed, evaporative cooling may be more important than sensible heat (cooling due to temperature difference alone) at least during the initial period, when warm, wet product is exposed to cold dry air.
- Water loss from unwrapped product is an important economic loss; a 1% loss (in terms of total product weight) means a 1% loss in sales value, much more in terms of loss of profit. Typical losses in meat chilling, freezing and cold storage are about 2–3% of product weight.
- Product appearance and surface feel or texture is strongly affected by water evaporation. Food may suffer from 'freezer burn' appearance, a glassy or desiccated look due to freeze drying and subsequent structural changes in the surface tissue.
- Most importantly, microbial growth is strongly dependent on both surface temperature and surface water activity, the latter being determined by a

balance between evaporation and internal water diffusion. It has been found experimentally[34, 35] that surface water activity during meat chilling changes in a complex manner, first falling as water evaporates strongly from a warm surface, then rising again as the surface cools and is re-wetted by moisture diffusing from inside.

15.6.1 Combined heat and mass transfer in food with impermeable skin

With some fruit such as apples and with moist food wrapped in semi-permeable films, the product is covered by a relatively impermeable skin which, together with the air film resistance, accounts for all the resistance to mass transfer. The moisture of the product under the skin can be assumed to be constant. This situation can be described by a 'lumped parameter' model. The product can be incorporated in a generalised model describing the whole food/package/stack system as a network of heat and moisture sinks linked by heat and mass transfer paths.[36]

15.6.2 Combined heat and mass transfer in non-porous food

With products that have no low-permeability skin or wrap, the partial differential equations describing heat and mass transfer in the solid must be solved simultaneously. An explicit (Euler) time-stepping scheme is the easiest to use and can account automatically for interactions between moisture diffusion and heat conduction, provided the time step is sufficiently small. Implicit stepping schemes often use large time steps and some iteration is necessary.

Moisture diffusion inside the product is governed by the PDE

$$\frac{\partial w}{\partial t} = \frac{\partial}{\partial x}\left(D_w \frac{\partial w}{\partial x}\right) \tag{15.17}$$

which can be discretised in a manner similar to the heat conduction equation. In this equation, w is the moisture content and D_w the diffusivity of moisture in the food. At the surface, the heat conducted to the surface is balanced by evaporation and convective cooling:

$$H_{fg} k_y (Y_a - a_w Y^{sat}(T_s)) + h(T_a - T_s) + k \left.\frac{\partial T}{\partial x}\right|_s = 0 \tag{15.18}$$

where H_{fg} is the latent heat of evaporation, k_y the mass transfer coefficient based on Y, Y the absolute humidity, Y^{sat} the saturation humidity, a_w the surface water activity, h the heat transfer coefficient and the subscripts a and s refer to air and surface respectively. The water diffusing to the surface must be balanced by evaporation:

$$k_y (Y_a - a_w Y^{sat}(T_s)) + \rho_{dm} D_w \left.\frac{\partial w}{\partial x}\right|_s = 0 \tag{15.19}$$

where the surface water activity is related to surface moisture by the equilibrium relationship

$$a_w = a_w(w_s, T_s) \tag{15.20}$$

and the convective heat and mass transfer coefficients are related by the Chilton-Colburn analogy:[37]

$$\frac{h}{k_y c_{pa}} = \left(\frac{Sc}{Pr}\right)^{2/3} \tag{15.21}$$

Radford *et al.*[38] were the first to present a finite difference model of simultaneous heat and mass transfer in thin slabs of meat. For large products such as beef carcasses, the numerical solution of the heat and mass transfer equations is difficult to carry out efficiently because the thermal and moisture diffusion coefficients in the food differ by several orders of magnitude (typically $1.0\times10^{-7}\,m^2/s$ for thermal diffusivity and $1.0\times10^{-10}\,m^2/s$ for moisture diffusivity). Hence heat diffuses much faster and on a much larger scale than moisture. For instance, during meat chilling, heat is lost throughout a block of meat while moisture loss is confined to a surface layer a few mm thick. This means that when using a numerical method such as finite difference, the grid for mass transfer needs to be much finer than that for heat transfer. Because of this difficulty, an empirical approach has often been used, preventing accurate computation of the cooling rate, evaporation rate and surface moisture. Usually a constant surface water activity is assumed. Davey and Pham[18] modelled a complete beef side assuming a constant surface water activity of 85%.

Pham and Karuri[35] used separate discretisation grids for temperature and moisture calculations. The temperature grid covers the whole object while a one-dimensional moisture grid covers the near-surface regions (up to a depth of 20 mm), which are affected by mass transfer. This approach gave reasonable predictions for the heat load and weight loss from a beef carcass during chilling, and also correctly predicted the surface drying and re-wetting process during chilling (Fig. 15.5). A major problem is the lack of accurate data on the equilibrium relationship $a_w(w, T)$ of fresh foods.

15.6.3 Combined heat and mass transfer in porous food

In most food, the interaction between heat and mass transfer in the product is largely one-way, moisture diffusion being strongly influenced by the temperature field but not the other way round. In porous food such as bread or fermented dough, there is strong two-way interaction, since heat transfer inside the product may be enhanced by evaporation-condensation effects:[39] moisture evaporates from the warmer parts of each pore and condenses on the cooler parts, as in a heat pipe. If one is concerned only with heat transport, an effective thermal conductivity can be used, where the increase in thermal conductivity *of the pores* due to evaporation-condensation can be calculated from:[40, 41]

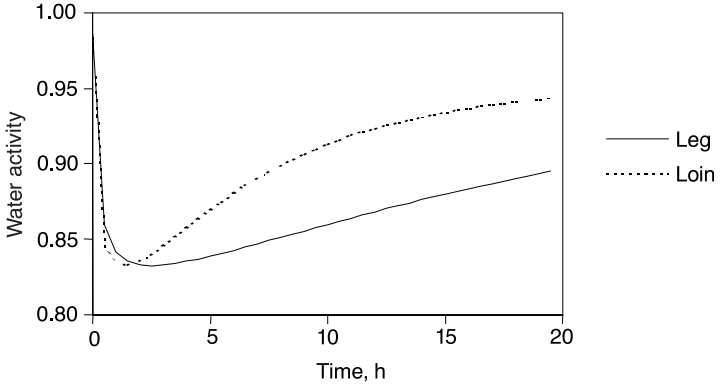


Fig. 15.5 Predicted surface water activity for rump and loin sections.[35]

$$k_{evap-cond} = a_w \nu H_{fg} D_{wa} \frac{d\rho_w^{sat}}{dT} \tag{15.22}$$

where ν is a mass flow correction factor ($\approx$1 at medium to low temperatures), L the latent heat of vaporisation, D_{wa} the diffusivity of water vapour in air and ρ_w^{sat} the saturated water vapour density. In the temperature range 0–40ºC the conductivity of the air-filled pores can be increased by a factor 2 to 9 by this contribution.[41] Because the pores form only a fraction of the total product volume, the resulting enhanced pore thermal conductivity must be combined with that of the solid matrix by means of a composite model such as the Maxwell-Eucken model or the EMT model.

A more rigorous approach is to solve both the heat and mass transfer equations. The program BAKTIX[42] uses this approach to calculate temperature and moisture profiles in bread being cooled and frozen. Water is assumed to move in the vapour phase, as governed by the equation

$$\dot{m} = \frac{D_{wa}}{\mu} \nabla \rho_w \tag{15.23}$$

where $\dot{m}$ is the water flux and μ a mass transfer resistance factor. The heat flux due to mass transfer consists of a sensible heat and a latent heat component:

$$q = k\nabla T + \dot{m}(c_{pw}T + H_{fg}) \tag{15.24}$$

where c_{pw} is the specific heat of water vapour. Using this approach, BAKTIX was able to predict qualitatively the movement of moisture towards the crust of bread during cooling, an important phenomenon which may cause a layer of ice to form under the crust during freezing and subsequent crust detachment.

15.6.4 Combined heat and mass transfer from frozen food

Food with a frozen surface loses moisture not by internal diffusion and evaporation from the surface but by sublimation from a receding ice front. The

ice front can be assumed to have the vapour pressure of pure ice, but as it recedes into the product a dehydrated layer forms which will slow down the sublimation rate.[43] Two phase change phenomena are involved, sublimation and freezing. The problem was modelled by Campanoñe *et al.*[44, 45] for 1D geometries using a front-tracking finite difference method. The dehydrated zone is modelled by a flexible grid with distance increments increasing proportionately to the depth of the freezing zone. The undehydrated zone (both frozen and unfrozen) was modelled by a fixed grid, except that the last node moves with the sublimating interface (and hence the last space increment of the undehydrated zone decreases with time). An apparent heat capacity method appeared to have been used to deal with the freezing front.

15.6.5 Combined heat and mass transfer from food undergoing cooling and freezing

In the problems considered in the previous sections, moisture is in either unfrozen or frozen form. In practical freezing, two distinct mechanisms apply: during the precooling period, moisture diffuses from the inside to the surface where it evaporates; once the surface has frozen, moisture is immobilised as ice, which sublimes and recedes under a dehydrated layer. The two mechanisms are of similar importance: during the cooling and freezing of lamb carcasses, for example, about 1% of the total product weight is lost during each period.[46] Different solution procedures must be applied in each period. Although this poses no fundamental problem, a full numerical model has yet to be implemented and tested for that situation.

15.7 Estimation of transfer coefficients

All heat and mass transfer models require the heat/mass transfer coefficient as an input. The importance of this parameter is measured by the Biot number, $Bi = hR/k$. To be efficient, a food cooling process should be designed for a Biot number higher than 1, say by increasing the air velocity in blast chillers and freezers or by using immersion or spray chilling, and thus it is quite important to know the transfer coefficient accurately. Unfortunately, this is one of the quantities least amenable to prediction and most subject to variations in practice.

Most food cooling models use a mean transfer coefficient over the whole product. This is unrealistic as in practice the coefficient varies from place to place with the development of the flow boundary layer, being higher where the latter is thinner (at the spot where the cooling air stream hits the product) and decreasing downstream. Minima and maxima may occur as the boundary layer is disturbed, detaches or reattaches, or recirculation zones occur near the surface. Where natural convection is the dominant effect or interacts with forced convection, the situation is even more complex since the effect of natural convection will gradually decrease as the product cools.

The effect of air turbulence has been a neglected area, due probably to the difficulty of measuring turbulence parameters in industrial situations. Yet turbulent intensity has a large effect on heat transfer, an increase in turbulence intensity from 6–31% having the same effect as an increase in air velocity from 0.5 m/s to 1.5 m/s during the cooling of pork.[47]

In the past, transfer coefficients have been obtained mostly from empirical relationships or direct experimentation. Recently computational fluid dynamics (CFD) is gaining attention as a tool for taking into account the variations in heat transfer coefficients. CFD has been used to simulate the airflow in and/or around display cabinets, cold stores, freezers, transport containers, and around products of various shapes. Some commercial CFD codes have an automatic grid refinement feature, the grid at locations where the velocity gradient is highest being automatically refined during computation to increase the accuracy of the prediction.[48] Perhaps the most complex food shape yet modelled by CFD is a beef side, modelled with a grid consisting of 100 000 nodes[26] (Fig. 15.6).

In principle, a CFD program can solve the whole problem in one go, giving both the temperature field inside the product and the velocity and temperature field outside it. This approach has been tested on a cylindrical piece of meat inside a chilling cabinet and produced good agreement between measured and predicted meat temperatures.[49] There are practical difficulties with this approach, however, due to the present limitations of CFD packages. In the modelling of the cooling of a beef side mentioned above, it was found difficult to model radiative heat transfer and simultaneous heat and mass transfer in the product and the air. Even without the mass transfer equation, it took a week of computation (on a Pentium II-300 MHz), with frequent manual interventions to vary the time step, to simulate a twenty-hour chilling process. A quicker if less rigorous approach is to use a steady-state version of CFD to compute surface transfer coefficients, then use these values in a heat-conduction only calculation of product temperatures; as long as the transfer coefficient is not very dependent on product temperature, the approach can give good results[50] (Fig. 15.7).

CFD programs solve the partial differential equations that describe the mass, momentum and energy balance, and for laminar flows exact solutions can be obtained. For turbulent flows the situation is less clear. The velocity fluctuations must be averaged, which give rise to a turbulent viscosity μ_t that depends on the intensity and scale of turbulence. Additional partial differential equations with empirical parameters must be solved in order to calculate this turbulent viscosity and its companions, the turbulent thermal and mass diffusivities. The most popular turbulence model is the $k - \varepsilon$ model, in which conservation equations for the turbulence intensity k and turbulent energy dissipation rate ε are solved. There is a lack of reliable model parameters for swirling flows, flows with recirculation or boundary layer detachment and natural convection, precisely the situations for which knowledge is most needed. Pham and Nguyen[50] found significant differences in the calculated heat transfer coefficient, especially at low velocities, when the standard $k - \varepsilon$ model and the so-called RNG model are used to calculate turbulence effects. It can be said, therefore, that while CFD can

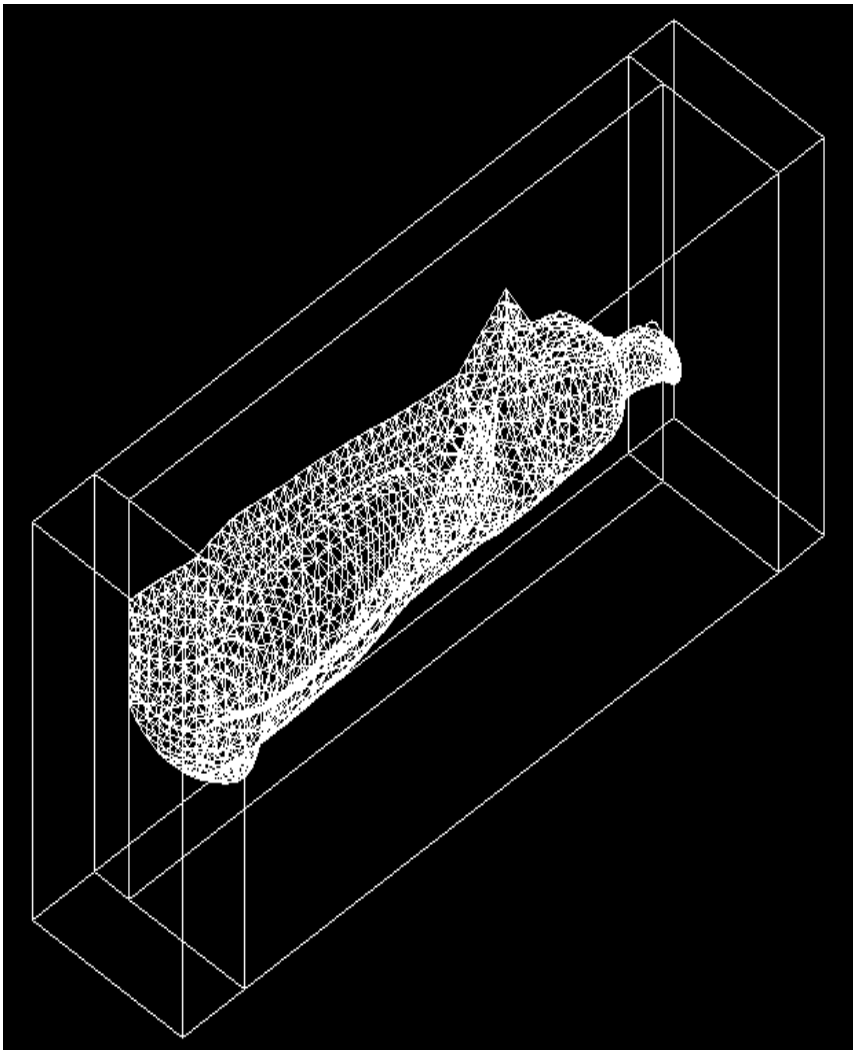


Fig. 15.6 Grid representation of a beef side for a CFD calculation.[26] The grid shown consists of about 10 000 nodes but will be automatically refined to 100 000 nodes by the CFD program.

provide a qualitative picture of the flows, precise heat transfer coefficient calculation for most situations is still beyond its capabilities.

The effect on heat transfer of voidage in loose packages of refrigerated products is a problem of some importance. If the air gap is very small, heat flows through it by pure conduction, but for gaps more than a few millimetres thick, convection sets in and increases the heat transfer. Both CFD[51] and experimental techniques[52] have been applied to determine the magnitude of these effects in cartons of meat.

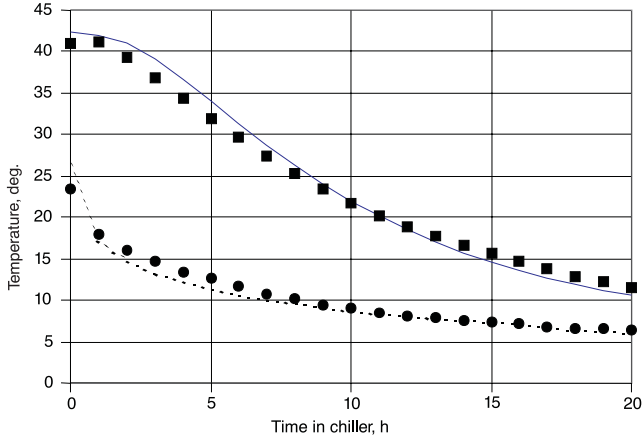


Fig. 15.7 Leg centre and surface temperatures of a beef side. ----- calculated using CFD-generated heat transfer coefficients, ■ ● measured.[50]

15.8 Application of models

15.8.1 Model-based control of food refrigeration processes

The ultimate purpose of modelling is to control and optimise real processes. Examples of such applications are relatively few in food refrigeration, but can be expected to become more and more common with recent advances in computer speed and software. Wee *et al.*[53] presented a computerised fan speed control system for a batch lamb freezer that has been implemented in New Zealand's meat plants. Data on the range of carcass weight entering the freezer, the loading time and the product wrapping used are entered manually (a process which can also be automated), and the freezer air temperature is monitored continuously. The fan speed is adjusted continuously to ensure that the product is frozen to specification (−18ºC deep leg) at the end of the allocated time, but no earlier, to ensure maximum tenderness of the meat and compliance with regulations. The calculation of the required fan speed is based on a Plank-type analytical model of the freezing time, modified to take into account variations in boundary conditions.[54]

15.8.2 Optimisation of food refrigeration processes

Experimental optimisation of food processing operations is time consuming and often difficult due to the variability of the product, errors in measurements and random variations in conditions. Computer models that can predict accurately the effects of operating parameters are therefore a powerful tool for optimisation. Such models usually make use of numerical methods that are time consuming and subject to rounding errors, making them difficult to

optimise by conventional mathematical methods. However, rapid advances in computer speed and the development of new stochastic optimisation methods are helping to overcome these obstacles.

The cooling of foods must satisfy often contradictory objectives. From the hygiene point of view, faster cooling is always better than slow cooling. Loss of water from unwrapped product is also minimised by fast cooling, due to the rapid chilling and drying of the surface. However, energy consumption is increased, while food quality factors can be affected in different ways. With fresh meat, fast cooling may lead to 'cold shortening'[55] and toughness, and there is a limited operating window which will produce meat of acceptable tenderness and hygiene. Other food quality aspects benefit from faster cooling or freezing. Mallikarjunan and Mittal[56] used Box's method to optimise the chilling of beef with respect to weight loss and chilling time. They suggested a two- or three-period chilling regime.

Pham and Lovatt[57] used an evolutionary optimisation method (genetic algorithm) to optimise the thawing of meat blocks and the chilling of beef carcasses. The problem was to design a temperature regime to chill a beef carcass within sixteen hours, keeping the potential growth of *E. coli* on the surface of the leg (the slowest cooling part) to within three generations (about one log increase), while allowing enough time to maximise the tenderness of the faster-cooling loin muscle. The air temperature is allowed to vary between +15ºC and −10ºC, due to regulatory and design constraints. The *E. coli* growth rate is assumed to follow Gill's[58] model, which describes the variation in lag time and growth rate as functions of temperature. The temperature dependence of the rate of development of tenderness is given by an Arrhenius-type equation.[59, 60] The objective function is the sum of the tenderness score and a stiff penalty function for exceeding the three-generation *E. coli* growth constraint. Both the leg and the loin are represented by finite-difference models. The numerical models produce small rounding errors, which make derivative-based optimisation methods unsuitable, hence a stochastic genetic-algorithm type method proved ideal for the task. The optimisation produced a complex air temperature regime, gradually decreasing from 15ºC to 0.8ºC then rising to 5.3ºC in the last hours. The tenderness score showed a significant improvement compared with conventional cooling regimes.

A model-based optimising controller for beef chilling is described by Lovatt and Pham,[61] in which the computer calculates the weight loss, meat ageing (development of tenderness) and potential for microbial growth. Each of these factors is assigned a unit cost or value. The controller then sets the air temperature, air velocity and relative humidity to maximise the total value of the product. These variables are not held constant, but change over time for two reasons. First, the optimum temperature, velocity and humidity profiles are not generally flat. Second, the profiles later in the process depend upon what happened earlier in the process. Thus, if there is some sort of disruption during the run, the controller will take that into account in its decision-making process for the latter part of the run.

Recently, there have been developments in multi-objective optimisation techniques[62] using evolutionary algorithms, which would be ideal for food refrigeration processes where weight loss, tenderness, microbial growth and other factors must be optimised simultaneously.

15.9 Summary and future developments

The cooling and freezing of solid foods is a challenging topic for modellers. In the last few decades, they have taken advantage of advances in mathematical techniques and computer technology to solve many problems, but much more research is still needed in a number of areas. Modellers would like to see their models used in the prediction, control and optimisation of process objectives. However, such applications have not been as widespread in industry as hoped for, and even when used they should be regarded with caution, due to weaknesses in the models. Further research needs to be carried out on 'weak links' in the modelling of food refrigeration processes, some of which are listed below.

15.9.1 Heat transfer coefficient estimation

The knowledge of heat transfer coefficients for complex flow patterns and complex geometries is still very inadequate. Data have been reported only for a few very simple situations, and mostly on average heat transfer coefficients only. Yet accurate knowledge of local heat transfer coefficients is essential for the calculation of product temperature and moisture near the product's surface, which in turn governs essential variables such as microbial growth and appearance. In most reported experimental work, the effect of turbulence has been ignored, although it could be a major factor.[63] The measurement of turbulence has until now been difficult but new, reasonably-priced turbulence sensors are coming on the market. Ideally, it should be possible to calculate heat transfer coefficients from first principles using CFD programs, but we lack a sufficiently accurate and fast model of turbulence for doing this. Although the fundamental fluid flow equation (Navier-Stokes equation) can be used in principle to model turbulence of any kind, it requires massive computing power even for the simplest problems, and is unlikely to be used in industry in the foreseeable future. Thus the prediction of heat and mass transfer coefficients is likely to remain a long-term problem in the modelling of food refrigeration processes.

15.9.2 Food properties estimation

There is now a good collection of data and calculation methods for the thermal properties (calorimetric properties, thermal conductivity) of a range of foods. However, water-related properties (moisture diffusivity, equilibrium isotherms)

are still insufficiently documented and there is a lack of general methods for their prediction. As a result, the prediction of weight loss and surface water activity, and hence microbial growth, is still fraught with uncertainties.

15.9.3 Modelling complex product shape

The availability of fast computers and advanced CAD-type geometric modelling software should make 2D and even 3D models more popular in the next few years, as part of CFD models or otherwise. Although building a 2D or 3D model requires special skills and software, there will be scalable models of standard products such as beef carcasses, bread loaves, etc., developed by specialists, models which the user can tailor by changing a few inputs such as product grade or dimensional ratios.

15.9.4 Modelling of mass transfer controlled freezing

In very fast freezing situations such as the immersion or cryogenic freezing of small foodstuffs, mass transfer (ice nucleation and crystal growth) rather than heat transfer may be the rate controlling factor. The microstructure of the frozen food and hence its quality indicators such as appearance, drip and denaturation will be highly sensitive to the rate of cooling. In some cases, slow cooling results in smaller ice crystals due to the effect of supercooling.[64] Cooling food below their glass transition temperature will exhibit massive increases in solution viscosity, which will dramatically reduce recrystallisation and other diffusion-controlled processes, including quality deterioration.[65] There has been little quantitative modelling work done to predict the freezing rate and quality changes in mass transfer-controlled situations.

15.9.5 Handling of statistical variations

Food products and processes are often subjected to much wider variations in shape, size, compositions and operating conditions than other manufactured products. This has traditionally imposed limitations on the usefulness of accurate models. Given the rapid increase in computer speed, the application of statistical simulation techniques such as Monte Carlo would appear desirable for many processes. While a 2D finite element simulation may still take several minutes to run and a 3D CFD simulation may take hours or days, a 1D finite difference simulation takes only a few seconds on a fast PC and thus Monte Carlo followed by statistical analysis is a feasible approach. The quality factors of the processed food can then be characterised fully by frequency distributions. We can expect that the statistical approach will be used more frequently in the evaluation of processes and the setting up of regulations.

15.10 References

1. PLANK R, 'Die Gefrierdauer von Eisblocken', *Zeitschrift für die gesamte Kalte Industrie*, 1913 **20**(6) 109–14.
2. LOVATT S J, 'A unified strategy for modelling meat refrigeration process indicators', *19th Int. Congress of Refrigeration*, The Hague, **IIIa** 653–9, 1995.
3. CLELAND A C, *Food Refrigeration Processes Analysis, Design and Simulation*, London, Elsevier, 1990.
4. ASHRAE, *Handbook of Refrigeration*, Atlanta, ASHRAE, 1998.
5. PHAM Q T, Prediction of cooling/freezing/thawing time and heat load.' In: *Advances in Food Refrigeration*, ed. Da-Wen Sun, Leatherhead, Leatherhead Food RA, 2001.
6. DAVEY L M, PHAM Q T, 'A comparison of three models for predicting heat and weight loss from beef sides during chilling'. *20th Int. Congress of Refrigeration*, Sydney, 1999, Paper 681.
7. PLANK R, 'Beitrage zur Berechnung und Bewertung der Gefriergeschwindigkeit von Lebensmitteln', *Zeitschrift für die gesamte Kalte Industrie*, 1941 **3**(10) 1–16.
8. CLELAND A C, 'Use of dynamic simulation in the design of food freezing equipment', In: *Food Engineering and Process Applications*, Vol. 2, ed. Le Maguer, M and Jelen, P, London, Elsevier, 1986.
9. LOVATT S J, PHAM Q T, LOEFFEN M P F, CLELAND A C, 'A new method of predicting the time-variability of product heat load during food cooling, Part 1: Theoretical considerations', *J. Food Eng*, 1993 **18** 13–36.
10. EARLE R L, FLEMING A K, 'Cooling and freezing of lamb and mutton carcasses: 1. Cooling and freezing rates in legs', *Food Technol.* 1967 **21** 79–84.
11. BAILEY C, COX R P, 'The chilling of beef carcasses', *Proc. Inst. Refrigeration*, 1976 **72** 76–90.
12. CLELAND A C, EARLE R L, 'Simple method for prediction of heating and cooling rate in solids of various shapes', *Int. J. Refrigeration*, 1982 **5** 98–106.
13. MCNABB A, WAKE G C, HOSSAIN M D M, LAMBOURNE R D, 'Transition times between steady states for heat conduction, Part I: General theory and some exact results', *Occasional Publications in Mathematics & Statistics*, No. 20, Massey University, 1990.
14. HOSSAIN M D M, CLELAND D J, CLELAND A C, 'Prediction of freezing and thawing times for foods of regular multidimensional shape by using an analytically derived geometric factor', *Int. J. Refrigeration*, 1992 **15**(4) 227–34.
15. LIN Z, CLELAND A C, SELLARACH G F, CLELAND D J, 'A simple method for prediction of chilling times: extension to three-dimensional irregular shapes', *Int. J. Refrigeration*, 1996 **19** 107–14.
16. CROCOMBE J P, LOVATT S J, CLARKE R D, 'Evaluation of chilling time shape

factors through the use of three-dimensional surface modelling', *20th Int. Congress of Refrigeration*, Sydney, 1999, Paper 353.
17. PHAM Q T, COULTER S, 'Modelling the chilling of pig carcasses using an evolutionary method', *Proc. 19th Int. Congress of Refrigeration*, The Hague, **IIIa**, 676–83, 1995.
18. DAVEY L M, PHAM Q T, 'Predicting the dynamic product heat load and weight loss during beef chilling using a multi-region finite difference approach', *Int. J. Refrigeration*, 1997 **20** 470–82.
19. ARCE A, POTLURI P L, SCHNEIDER K C, SWEAT V E, DUTSON T R, 'Modelling beef carcass cooling using a finite element technique', *Trans. A.S.A.E.*, 1983 **26** 950–4, 960.
20. MALLIKARJUNAN P, MITTAL G S, 'Heat and mass transfer during beef carcass chilling – modelling and simulation', *J. Food Eng.*, 1994 **23** 277–92.
21. DAVEY L M, PHAM Q T, 'Development of a multi-layered two dimensional finite element representation of a beef side, to enable dynamic product heat load and weight loss prediction during beef chilling', *Int. J. Refrigeration*, 2000 **23**(6) 444–56.
22. PHAM Q T, MERTZ I, WEE H K, 'Prediction of temperature changes in lamb loins by the finite element method', *Proc. 18th Int. Congress of Refrigeration*, Montreal, 1991, **IV** 1859–62.
23. CLELAND D J, Prediction of Freezing and Thawing Times for Foods, Ph.D. thesis, Massey University, New Zealand, 1985.
24. VAN GERWEN R J M, 'Modelling temperature distribution in products moving in cold chains'. *Int. Inst. Refrigeration Meeting*, Wageningen, Netherlands, 1988.
25. VAN DER SLUIS S M, 'Simulation of heat and mass transfer in products', *Proc. 19th Int. Congress of Refrigeration*, The Hague, **IIIa** 692–8, 1995.
26. NGUYEN A, PHAM Q T, 'A computational fluid dynamic model of beef chilling', *20th Int. Congress of Refrigeration*, Sydney, 1999, Paper 310.
27. KIRCHHOFF J C, Vorlesungen über die Theorie der Wärme, 1894. Cited in Carslaw H S and Jaeger J C *Conduction of Heat in Solids*, 2nd edn, Oxford, Clarendon Press, 1959, p. 11.
28. PHAM Q T, 'A fast, unconditionally stable finite-difference method for heat conduction with phase change', *Int. J. Heat Mass Transfer*, 1985 **28** 2079–84.
29. PHAM Q T, 'The use of lumped capacitances in the finite-element solution of heat conduction with phase change', *Int. J. Heat Mass Transfer*, 1986 **29** 285–92.
30. EYRES N R, HARTREE D R, INGHAM J, JACKSON R, SARJANT R J, WAGSTAFF J B, 'The calculation of variable heat flow in solids', *Trans. Royal Soc.* 1946 **A240** 1.
31. PHAM Q T, 'A comparison of some finite-difference methods for phase-change problems', *Proc. 17th Int. Congress of Refrigeration*, Vienna, **B** 267–71, 1987.

32. COMINI G, DEL GIUDICE S, SARO O, 'Conservative equivalent heat capacity methods for non-linear heat conduction', In: Lewis R W and Morgan K (eds) *Numerical Methods in Thermal Problems*, Swansea, Pineridge, 1989 **6**, Part 1, pp. 5–15.
33. PHAM Q T, 'Comparison of general-purpose finite element methods for the Stefan problem', *Numerical Heat Transfer Part B – Fundamentals*, 1995 **27** 417–35.
34. HERBERT L S, LOVETT D A, RADFORD R D, 'Evaporative weight loss during meat chilling', *Food Technol. Australia*, 1978 **30**, 145–8.
35. PHAM Q T, KARURI N, 'A computationally efficient technique for calculating simultaneous heat and mass transfer during food chilling', *20th Int. Congress of Refrigeration*, Sydney, 1999.
36. TANNER D J, CLELAND A C, ROBERTSON T R, OPARA L U, 'A generalised model for prediction of mass loss from packaged horticultural produce during storage', *Acta Horticulturae*, 1998 **476** 113-20.
37. PERRY R H, GREEN D W, MALONEY J O, *Perry's Chemical Engineers' Handbook*, 7th edn, New York, McGraw-Hill, 1997, pp. 12–13.
38. RADFORD R D, HERBERT L S, LOVETT D A, 'Chilling of meat – math model for heat and mass transfer'. *Int. Inst. Refrigeration Bulletin*, 1976 Annex 76–1 323–30.
39. DE VRIES D A, 'The theory of heat and moisture transfer in porous media revisited', *Int. J. Heat Mass Transfer*, 1987 **30**, 1343–50.
40. DE VRIES D A, 'The thermal conductivity of granular materials', *Bull. Int. Inst. Refrigeration*, 1952, Annex 1952–1 115–31.
41. DE VRIES D A, 'Simultaneous transfer of heat and moisture in porous media', *Trans. Am. Geophysical Union*, 1958 **39**(5) 909–16.
42. VAN DER SLUIS S M, 'Cooling and freezing simulation of bakery products', *Int. Inst. Refrigeration Meeting Comm. B1, B2, D1, D2/3*, Palmerston North, New Zealand, 1993.
43. PHAM Q T, WILLIX J, 'A model for food desiccation in frozen storage', *J. Food Sci.*, 1984, **49** 1275–81, 1294.
44. CAMPAÑONE L A, SALVADORI V O, MASCHERONI R H, 'Modelling and simulation of heat and mass transfer during freezing and storage of unpackaged foods', *Int. Inst. Refrigeration Meeting of Commissions B1, C1 and C2*, Orsay, 1998.
45. CAMPAÑONE L A, SALVADORI V O, MASCHERONI R H, 'A finite difference method to solve coupled heat and mass balances with simultaneous surface dehydration during freezing', *Latin American Applied Research*, 1998, **28** 83–8.
46. LONGDILL G R, PHAM Q T, 'Weight losses of New Zealand lamb carcasses from slaughter to market', *Int. Inst. Refrigeration Bull.*, Annex 1, 1982, 123–30.
47. KONDJOYAN A, DAUDIN J D, 'Influence of airflow properties during chilling and chilled storage of foodstuffs', *19th Int. Congress of Refrigeration*, The Hague, Netherlands, **II** 202–9, 1995.

48. FLUENT INC., *Fluent User's Guide Manual*, Vols. 1, 2, 3, 4, releases 4.5 & 5.0, Fluent Inc., Lebanon, NH, USA, 1998.
49. HU Z, SUN D-W, 'Temperature distribution of cooked meat joints in an air-blast chiller during cooling process: CFD simulation and experimental verification.' *20th Int. Congress of Refrigeration*, Sydney, 1999, Paper 169.
50. PHAM Q T, NGUYEN A V, 'Mean convective heat transfer coefficient of beef carcases computed by a computational fluid dynamic model', *FoodSim 2000*, Nantes, France, 26–27 June 2000.
51. WEE H K, PHAM Q T, 'Numerical model of unsteady-state heat transfer with convection and phase change in a carton of meat'. *Chemeca (Australasian Chemical Engineering Conference)*, Auckland, New Zealand, 1989.
52. NORTH M F, LOVATT S J, CLELAND A C, 'Methods for evaluating the effect of large voids on food package cooling time', *20th Int. Congress of Refrigeration*, Sydney, 1999.
53. WEE H K, LOEFFEN M P L, PHAM Q T, 'Fan speed control in a blast freezer using model reference control', *Proc. 18th Int Congress of Refrigeration*, Montreal, Canada, **III** 997–1000, 1991.
54. LOEFFEN M P F, EARLE R L, CLELAND A C, 'Two simple methods for predicting food freezing time with time variable boundary conditions', *J. Food Sci.*, 1981 **46**, 1032–4.
55. LAWRIE R A, *Meat Science*, 5th edn, Oxford, Pergamon Press, 1991.
56. MALLIKARJUNAN P, MITTAL G S, 'Optimum conditions for beef carcass chilling', *Meat Sci.*, 1994 **39** 215–23.
57. PHAM Q T, LOVATT S J, 'Optimisation of refrigeration processes by stochastic methods', *Food Australia*, 1996 **48**(2) 64–9.
58. GILL C O, 'Assessment of the hygienic efficiencies of processes for cooling meat at slaughtering plants', *Agriculture Canada Technical Bulletin* 1993–10E, 1993.
59. GRAAFHUIS A E, LOVATT S J, DEVINE C E, 'A predictive model for lamb tenderness', *Proc 27th New Zealand Meat Industry Research Conf. (Hamilton)*, 143–7, 1992.
60. DRANSFIELD E, WAKEFIELD D K, PARKMAN I D, 'Modelling post-mortem tenderisation – I. Texture of electrically stimulated and non-stimulated beef', *Meat Sci.*, 1992 **31** 57–73.
61. LOVATT S J, PHAM Q T, 'Integrated mathematical models to simultaneously predict and control product quality and refrigeration system performance', *New Zealand Inst. Food Sci. Technol./Australian Inst. Food Sci. Technol. Conference*, 1995.
62. ZITZLER E, DEB K, THIELE L, 'Comparison of multiobjective evolutionary algorithms: Empirical results', *Evolutionary Computation*, 2000 **8**(2) 1–24.
63. KONDJOYAN A, DAUDIN J D, 'Heat & mass transfer coefficient at the surface of elliptical cylinders placed in a turbulent air flow', *J. Food Eng.*, 1993 **20** 339–67.

64. REID D S, 'Factors which influence the freezing process: an examination of new insights', *20th Int. Congress of Refrigeration*, Sydney, 1999.
65. REID D S, HSU J, JU J, 'The importance of the glass transition temperature to the temperature dependence of the stability of frozen foods', *19th Int. Congress of Refrigeration*, The Hague, **II** 331–8, 1995.

16

Modelling thermal processes: heating

S. Bakalis, P. W. Cox and P. J. Fryer, University of Birmingham

16.1 Introduction

16.1.1 Types of thermal process

Thermal processing is a ubiquitous operation in food processing. Many of the commonest food processing operations, such as canning, baking and pasteurisation, rely on heating:

- In a number of cases the effect of heat is intended for preservation alone, i.e. to kill bacteria and inactivate enzymes, such as in the pasteurisation of milk and the sterilisation of canned food. In this case the aim is to deliver the required microbial kill with as little damage to the structure of the food as possible.
- In the processing of foods, such as meats and vegetables, heat acts also to develop taste and flavour, so that in addition to sterilisation heat is required to carry out physical changes to the food.
- However, in many other situations food is heated to develop the structure of the material, such as in baking of bread or biscuits, where heating acts both to change the starch structure and function and also to develop the bubble structure within the material (for examples, see Campbell *et al.*, 1999)
- There are also situations, such as drying and frying, in which heat transfer is accompanied by mass transfer, and the two effects must be considered as coupled; evaporation of moisture in drying requires heat transfer to provide the necessary heat.

Each of the above points has been the subject of whole books! – to attempt a full summary is impossible. The modelling of each type of process has followed the same general trend:

- purely empirical correlations or graphical solutions
- models based around very simple approximations, such as simple geometries (spheres or cylinders) and uniform physical properties
- finite difference models, in which the basic equations are discretised on a simple grid system
- finite element (FE) and finite volume models, in which more complex geometries can be used.

As computer power has increased, FE models have become easier to solve, the codes have become more robust and more accurate, and the computers on which the codes run have become cheaper.

For simplicity, this review will concentrate on the case of heating for microbial cook alone, and will not consider taste or texture development. It will show how modelling is being used and how modelling techniques can be applied in this area. The problem with using heat for sterilisation is in demonstrating that the material is safe; models are critical in minimising the amount of experiments that have to be done and in convincing regulatory bodies.

16.1.2 Basic equations

Microbiological and quality kinetics

The rate of thermal processing is commonly quantified through the integrated lethality of a thermal process calculated using equation (16.1) (from the work of Ball, 1923):

$$F = \int_0^t 10\left(\frac{T(t) - T_{ref}}{z}\right)dt \qquad (16.1)$$

where z is the increase in temperature that gives an increase in rate of a factor of 10, and T_{ref} is a reference temperature. F has the units of time; it is the length of time that the food would have to be held at the reference temperature to obtain the same effect as the actual process with T. The thermal history of the processed product may equally be applied to product sterility or nutritional quality (where z will become a rate of cooking and not microbial lethality). Equation (16.1) is difficult to justify other than as an experimental fit; it is a local approximation to the Arrhenius expression, only accurate over a narrow temperature range, and it is not clear whether the death of microorganisms follows the Arrhenius kinetics.

Thermal transport equations

Heat is transferred by three mechanisms: conduction through solids or stationary liquids or gases, convection through flowing fluids, and radiation. For conduction and convection the rate of heat transfer is proportional to the temperature difference, whilst for radiation it is the difference between the fourth power of the temperatures. There are a multitude of good books on heat transfer, amongst them Incorpera (1981), Özişik (1993) and Carslaw and Jaeger (1980). The reader is also referred to Chapter 4 for a more comprehensive

overview of models for heat and mass transfer and methodologies for their numerical solution.

For products where conduction is the sole mechanism of heat transfer the temperature profile may be estimated from the partial differential equation:

$$\rho c_p \frac{\partial T}{\partial t} = \nabla(\lambda \nabla T) \tag{16.2}$$

the solution of which requires knowledge of the spatial variation, and temperature dependence of the thermal conductivity, λ, the density, ρ, and the specific heat, c_p, of the product. At the edges of the solid, different boundary conditions may apply. The simplest is constant temperature; however, a heat transfer boundary condition is often necessary, in which the flux to the surface is given by, for example, a convective heat transfer coefficient, or by radiation. The overall rate of heating of a solid will depend on consecutive processes; heat must move to the product and then within it. The relationship between external and internal thermal transport can be estimated using the Biot number:

$$Bi = \frac{hd}{\lambda} \tag{16.3}$$

where h is the interfacial heat transfer coefficient and d some characteristic dimension of the body being heated. The higher the Biot number, the greater is the effect of heat transfer coefficient; in practice, a $Bi > 10$ implies that the slowest heat transfer process will be conduction within the solid particle. For a low Biot number (<0.1) the process is controlled externally, with the solid essentially isothermal.

The heating of fluids is more complex because of fluid motion, so that both thermal and fluid transport equations need to be solved. Solution of the Navier-Stokes equation is needed for the flow field. Simplified equation sets are often used; for example, in a tubular geometry the partial differential equations describing the heat and momentum transport are (Bird *et al.*, 1964):

Equation of continuity:

$$\frac{1}{r}\frac{\partial}{\partial r}(rv) + \frac{\partial u}{\partial z} = 0 \tag{16.4}$$

Equations of motion:

$$\rho\left(v\frac{\partial u}{\partial r} + u\frac{\partial u}{\partial z}\right) = -\frac{\partial P}{\partial z} + \frac{1}{r}\frac{\partial}{\partial}\left(r\mu\left(\frac{\partial v}{\partial z} + \frac{\partial u}{\partial r}\right)\right) + 2\frac{\partial}{\partial z}\left(\mu\frac{\partial u}{\partial z}\right)$$

$$\rho\left(v\frac{\partial v}{\partial r} + u\frac{\partial v}{\partial z}\right) = -\frac{\partial P}{\partial r} + \frac{2}{r}\frac{\partial}{\partial}\left(r\mu\frac{\partial v}{\partial r}\right) + \frac{\partial}{\partial z}\left(\mu\left(\frac{\partial v}{\partial r} + \frac{\partial u}{\partial z}\right)\right) \tag{16.5}$$

Equation of energy:

$$\rho C_p\left(v\frac{\partial T}{\partial r} + u\frac{\partial T}{\partial z}\right) = \lambda\left\{\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{\partial T}{\partial r}\right) + \frac{\partial^2 T}{\partial z^2}\right\} \tag{16.6}$$

The assumptions used in deriving these equations are:

- that the flow is axisymmetric
- there is negligible thermal generation by viscous dissipation
- the effects of natural convection are also negligible
- the liquid is homogeneous
- constant density, specific heat and thermal conductivity

The types of boundary conditions usually applied are: for both velocity and temperature, a known profile at the inlet of the heater, with a known temperature profile at the wall of the heater and cooler, with a no-slip boundary for the velocity. In the case where a holding tube is used an adiabatic boundary condition is applied at the wall.

The next sections review the modelling of heat transfer to packed and flowing foods using both conduction/convection and heat generation models.

16.2 Processing of packed and solid foods

16.2.1 Introduction

The classical method of thermal preservation is canning, still the basis of a very large industry, despite the reduced quality of many canned products. Canned food is not sterile when it is packed; filled cans are exposed to a temperature-time profile sufficient to give a safe product. However, any process within which heat is applied externally will cause the centre of the product to lag the surface and bulk product temperatures. Therefore, estimates of temperatures at the slowest heating point are required if equation (16.1) is to be used effectively and, more importantly, safely. The heating rate at the slowest heating point may also be used to ensure that a product is not over processed and therefore experience an impractical, in terms of quality and process costs, thermal history. Estimates of F values using equation (16.1) to describe integrated lethality were first proposed by Ball (1923). Within this work with canned low acid foods (pH < 4.5), the proposed reference temperature (T_{ref}) was 121.1ºC and a Z value of 10 was used. The Z value was determined from the slope of a decimal reduction time against temperature graph for the thermally tolerant spores of the spoilage pathogen *Clostridium botulinum*. A thermal process was then considered to be safe if the slowest heating point of a can reached an F value of three minutes.

Mathematical modelling allows estimation of temperature at the slowest heating point. In addition, using a model it is possible to explore processing variation upon the calculation of a desired F value and therefore process operation. This is particularly true if a new product or processing method is to be established, especially where *a priori* knowledge of the product and process variability is not available.

Although tables exist to predict the measured temperature responses and thermal diffusivities of various foods and packages (e.g., in Tucker and Holdsworth, 1991), many practical and theoretical investigations have involved

the examination of simplified regular geometries, such as cylinders or spheres (e.g. Kim and Teixeira, 1997). It was proposed that for conducting solids, and for non-flowing conducting liquids, a geometric simplification was valid as the thermal diffusivity (α) need only be estimated across the diameter of a cylinder describing the shortest chord of the original solid. Paramount to successful modelling is availability of accurate physical data with which models can be constructed and used for process simulations. In addition to static isothermal estimates of parameters, such as density, viscosity, thermal conductivity and specific heat capacity, knowledge of how these parameter change with heating and process time must also be used. For example, it is difficult to estimate the thermal diffusivity (α) and the surface heat transfer coefficient h in the same experiment; α should be inferred from heating curves where an infinite surface heat transfer coefficient can be assumed, i.e. with condensing steam (Kim and Teixeira, 1997).

16.2.2 Conduction in simple solids: solutions of the conduction equation

Before personal computers became commonplace, calculating the integrated lethality was complex and time consuming even for geometrically simplified conduction cooked products. In place of the repetitive manual calculations 'simpler' methods were used, for example, F value estimates could be made by first measuring temperature (using a thermocouple), calculating the lethality and then examining the area under a lethality rate versus process time graph, which could be directly correlated with F (for examples of this see Lopez, 1987). However, for products where conduction is the sole mechanism of heat transfer the temperature profile can be estimated using equation (16.2). Graphical solutions for this partial differential equation can be produced for various situations of simple-shaped solids and constant thermal diffusivity, e.g. the Heisler and Gurney-Lurie charts, examples of which can be seen in Toledo (1991). These charts plot dimensionless temperature (θ) against Fourier number ($\alpha t/R^2$) for the inverse of various Biot numbers ($1/Bi$). Solutions for the appropriate geometry allows the temperature at the geometric centre to be estimated. However, these methods employ simplifying assumptions:

- internal heat transfer is solely by conduction: external heat transfer is by uniform external heat transfer coefficient or wall temperature
- only simple shapes (spheres, cubes, cylinders) can be solved
- the product is homogeneous and isotropic
- the initial temperature is uniform.

Akterian (1999) summarised the application of equation (16.2) to conductive heat transfer in terms of a partial differential equation, incorporating a shape factor G to account for a product with symmetrical geometry (equation 16.7a). Equation (16.7b) describes the boundary conditions for surface convection and equation (16.7c) is the boundary condition for the line of symmetry.

$$\frac{\partial T}{\partial t} = \alpha\left(\frac{\partial^2 T}{x\partial^2} + \frac{G}{x}\frac{\partial T}{\partial x}\right) \tag{16.7a}$$

$$\frac{\partial T}{\partial x} = -\frac{h}{\lambda}(T - T_m) \tag{16.7b}$$

$$\frac{\partial T}{\partial x} = 0 \tag{16.7c}$$

where T_M is the temperature of the heating medium.
Such a simplified modelling approach has been employed to examine the thermal inactivation of bacteria in model foods, for example Bellara *et al.* (1999) modelled the growth of pathogenic bacterial species after thermal treatment within agar cylinders. Investigations modelled the inactivation of the bacteria and explored the notion that slow heating rates (≤ 2ºC/min) may increase thermotolerance in potentially pathogenic bacteria (Mackey and Derrick, 1987; Quintavalla and Campanini, 1991). The model could predict successfully the rate and position of bacterial destruction across the radius of the model sausage, using equation (16.8) and predict the reduction in bacterial numbers, *N*.

$$\frac{dN}{dt} = \frac{2.303}{D_{ref}} 10^{(T-T_{ref})/z} N \tag{16.8}$$

Hendrickx *et al.* (1993) describe how thermal processing of solids might be optimised, using computer-based finite difference calculations. Early attempts using digital computers include Timbers and Hayakawa (1967), Hayakawa (1969) and Teixeira *et al.* (1969). Hendrickx *et al.* (1993) then extended the discussion and development of a model for variable sterilisation temperatures. Numerical simulations could be used to examine and explore varying factors such as the target *F* value, *Z* value, product quality (Z_q), initial product temperature and retort come up time. However, the model of Hendrickx *et al.* (1993) still used simple product geometries, i.e. infinite cylinder, slab and spheres, and the assumed simplifications, detailed above, for the exploration of various process parameters.

16.2.3 More complex models: non-uniformity and convective flows

Real products are rarely of a regular geometry, have thermal properties which vary with temperature and have different heat resistances along the boundary. For example, in retorts, where condensing steam is used as the heating method, condensation may adversely affect the uniformity of heat transfer to the product surface; heat transfer to a dry surface will be very high, but the presence of a film of liquid will reduce the heat transfer rate (Verboven *et al.*, 1997).

Non-isotropic aspects of conductive cooking have been addressed, for example, by Pan *et al.* (2000) in the modelling of the cooking of frozen hamburgers. Their approach, which involved unequal cooking to both the major external surfaces of the patty, considered the enthalpy changes associated with the melting of ice and fat as well as resulting mass transfer effects. The base

equation used in the model (equation 16.9) was considered to be valid throughout the whole structure of the hamburger patty regardless of whether it was frozen or not, or whether it was a surface crust or the bulk of the patty.

$$\frac{\partial H}{\partial t} = \frac{\partial}{\partial x}\left(\lambda(H)\frac{\partial T(H)}{\partial x}\right) \tag{16.9}$$

where λ and T are both functions of the enthalpy. It was claimed that consideration of both enthalpy changes and of mass transfer effects made the proposed final model more accurate than obtainable with a heat transfer model alone.

FE methods of examining heat transfer now allow partial differential equations to be explored in complex geometries. For example, Tewkesbury *et al.* (2000) used a computational model to predict the cooling of chocolate within a polycarbonate mould. The use of commercially available software, in this case FIDAP (Fluid Dynamics International Inc., Evanston, Illinois), allowed the authors to model the conduction cooling of chocolate through the mould, and took into account the different thermophysical properties of each component, particularly the change in the effective specific heat capacity of chocolate as a function of temperature and cooking rate. The model developed by Tewkesbury *et al.* (2000) gave a solution for a non-isotropic system, although the properties of each of the components were constant, and was able to predict accurately the heat conduction between the chocolate, the mould and to the environment.

FE techniques no longer require very complicated packages; Fig. 16.1 shows the solution of the conduction equation for heat transfer for a complex product

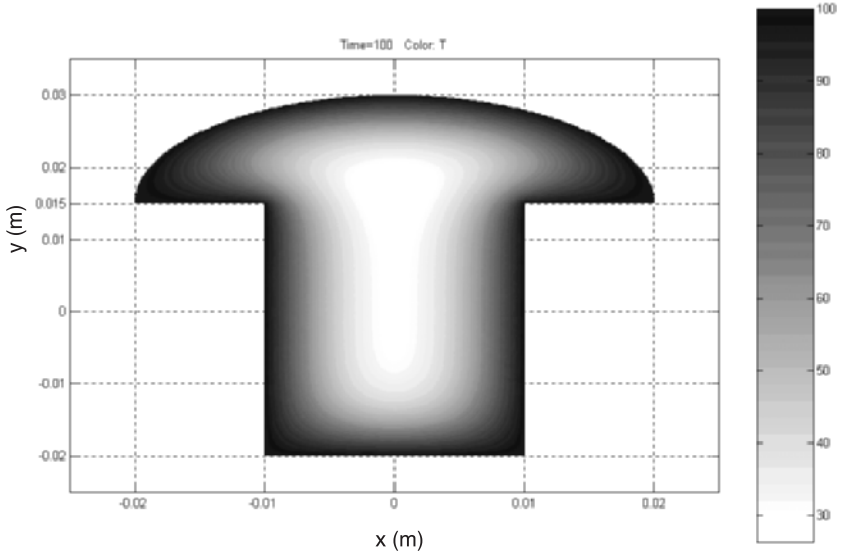


Fig. 16.1 Partial differential equation solution for heat transfer into a complex product. The key indicates the final temperature after 100 seconds.

geometry using the PC package Matlab (The MathWorks Inc., Natick, MA). The two-dimensional solution (i.e. assuming a transverse section of an infinite slab) assumes that the product has the thermal properties of water, an initial temperature of 20ºC and a boundary temperature of 100ºC. The simulation is over 100 seconds and the temperature profile takes only five seconds to calculate.

Importantly, for a large proportion of commercially processed products, heating is rarely by pure conduction. A prime example of this is canning, where liquids or soft solids of varying viscosities are routinely processed in batch retorts; heat transfer occurs by a combination of conduction and natural convection within the can. Computational fluid dynamics (CFD) can be used to solve for the flow field. A large body of literature exists to describe this complex situation. Reviewed here are several recent papers that demonstrate progression to the current limit of understanding of such systems. All of the papers cited contain useful broad introductions to the area.

Kumar *et al.* (1990) and Kumar and Bhattacharya (1991b) describe the heating of canned viscous liquids that have temperature dependent viscosity. Within their CFD model an initial conductive heat transfer phase reduces the viscosity of the liquid at the periphery of the can. This less viscous and more buoyant hot liquid rises and then re-circulates to the centre of the can. The equations that describe this action (continuity – equation (16.4), energy conservation – equation (16.6) and momentum in the radial direction – equation (16.5)) and the effects of these phenomena are discussed more fully below in the context of fluid movement within pipes. In naturally convecting cans the momentum equation for the liquid movement incorporates a natural convection term (i.e. the density driving force ρg; Abdul Ghani *et al.*, 1999a):

$$\rho\left(\frac{\partial u}{\partial t}+v\frac{\partial u}{\partial r}+u\frac{\partial u}{\partial z}\right)=-\frac{\partial p}{\partial z}+\mu\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)+\frac{\partial^2 u}{\partial z^2}\right]+\rho g \quad (16.10)$$

Through their model Abdul Ghani *et al.* (1999a) provided an estimate of the axial velocity of fluid elements. In carboxy-methyl cellulose it was estimated to be 10^{-4}–10^{-5} ms^{-1} and for water: 10^{-2}–10^{-1} ms^{-1}. The differences in magnitude were explained by the ratio of the buoyancy force to the viscous force in each case (expressed as the Grashof number). However, the rates quoted for water have not been observed by these authors when using a positron emission tomography method to study can heating, albeit at lower temperatures than those reported (unpublished results: Bakalis, Cox and Fryer; also see section 16.5).

Additionally, Abdul Ghani *et al.* (1999b) introduced a rate for bacterial inactivation within convecting liquids, combining convection and reaction terms within a single equation:

$$\frac{\partial C}{\partial t}=v\frac{\partial C}{\partial r}+u\frac{\partial C}{\partial z}=D_e\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial C}{\partial r}\right)+\frac{\partial^2 C}{\partial z^2}\right]-k_t C \quad (16.11)$$

All cans contain some headspace, i.e. air above the surface of the fluid. As Kumar *et al.* (1990) pointed out headspace may perform one of two roles. When water is heated within a can containing a headspace the free space may quickly

become saturated and give a high heat transfer rate as if it were condensing steam. In contrast, the space above a viscous liquid may not become wholly saturated and therefore the heat transfer rate would be reduced ($< 70\,\mathrm{Wm^{-2}K^{-1}}$). Abdul Ghani *et al.* (1999a) suggest that the headspace acts as an insulator, and that natural convection within a can may induce a temperature difference of approximately $10-12^{\circ}\mathrm{C}$ depending on the rate of liquid movement and headspace insulation. A secondary point was made that the bottom surface of a can – resting on a retort crate, for example – may not receive as much heat as the side wall increasing the temperature difference. As a consequence of convection and fluid displacement the slowest heating point is no longer the geometric centre; it was found to be within the bottom tenth of the can height and not on the axis of symmetry.

Differences in external heat transfer have prompted the simulation of the temperature distributions throughout batch retorts. As pointed out above, the boundary conditions experienced by a product may have a profound effect upon the estimation of the integrated lethality, and reports of temperature deviation of up to $10^{\circ}\mathrm{C}$ within a process can be found (Akterian *et al.*, 1998). Additionally, model loads have been investigated, albeit as conduction-heated products, within the virtual processes. Akterian (1999) and Akterian *et al.* (1998) suggested the use of previously calculated sensitivity functions, these factors correct for arbitrary fluctuations in the heating medium during different parts of the heating holding and cooling cycles. Incorporating these functions with an estimate of the bulk temperature deviation, estimated from the expected lethality rate, allowed for the process to be controlled using a simple microprocessor controller. Ryckaert *et al.* (1999) linked a process model to a PID controller that operated an industrial oven. The modelling of the heat balance with the heated chamber was used with temperature estimates from 32 thermocouples throughout the oven volume. The modelled temperature distribution was then used to tune the PID terms and so improve the oven control through the different heat input rates that correspond to different parts of the process cycle.

Varga *et al.* (2000) used an FE method to model horizontal cascading water retorts, and verified their model with industrial scale measurements. It was stated that a reliable mathematical solution required a model that should be combined with an appropriate statistical method to estimate real-world variability, as well as an understanding of the quantitative implications of the variance upon the system. Similarly, Verboven *et al.* (2000) used CFD to examine temperature deviations within an industrial scale forced convection oven. The model could predict oven temperatures to within $5^{\circ}\mathrm{C}$ of measured values (albeit in a trial with a reduced product load), although it was pointed out that because of the assumptions made about wall effects, in order to make the computation time non-prohibitive, at present the model was of qualitative use only.

What remains as a clear goal is the accurate modelling of the inside of an oven or retort with a full load of convection heated products with temperature dependent physical properties. This aspiration, although not yet realised in the

literature, may soon be reached as computation becomes quicker and models more accurate and descriptive.

16.3 Continuous heating and cooling processes

Continuous thermal processing was developed to solve the problems associated with traditional batch processing, such as low heat transfer rates, the long processing times needed to achieve the required lethality and the resulting high nutritional losses. The advantage of continuous thermal processing lies in the fact that the activation energy of microbial death is higher than that of the reactions occurring during thermal destruction of nutrients. As a result, high temperature short time (HTST) processes offer the potential to give the same level of sterility for a reduced quality loss. It is also possible to get higher heat transfer rates in fluids flowing through heat exchangers. Aseptic processes first appeared in 1927 while the first patent was granted to Ball in 1936 (Ramaswamy *et al.*, 1995). Commercial application of the technology was used extensively only after the introduction of aseptic packaging and flexible packages, in the 1960s – however, packing methods for high-solids fraction foods are still not well developed.

Depending on the nature of the processed food the physics of the process are quite different. Therefore, simulations will be examined separately for liquid and particulate foods. In both cases a variety of approaches have been used. Historically, early investigators used over-simplifying assumptions in order to derive analytical solutions; more recently researchers have used discretised techniques to solve the partial differential equations describing the problem.

16.3.1 Sterilisation of liquid foods

Continuous sterilisation of foods is commonly used for products like milk, juices, sauces and soups. The application of this technology has gained a lot of attention with the introduction of aseptic packaging, which resulted in self-stable products in convenient flexible packages.

In a typical sterilisation process the product passes through a heater, in order to inactivate the bacteria and is then cooled. In many cases sterility is achieved in a holding tube placed between the heater and the cooler. Although a number of heat exchangers are used in the industry, most of the published mathematical models use a simplified pipe geometry. The process variables in this case are the product flow rate (typically 100 l/min), the pipe diameter (typically 0.03 m), the temperature of the heater (typically 140ºC) and the physical properties of the fluid (rheological properties, specific heat capacity and thermal conductivity). In most cases relevant to industrial applications the flow is laminar.

The concept of F value, originating from traditional canning, is also used in aseptic processing to describe microbial death. For the non-isothermal case of continuous flow the equation for the total lethality F at a radial position r over an axial length L, can be modified as follows:

$$F(r,L) = \int_0^t 10^{(T(r,t)-T_0(r,t))Z} dt = \int_o^L 10^{(T(r,t)-T_0(r,t))Z} \left(\frac{1}{\mathrm{d}z/\mathrm{d}t}\right) \mathrm{d}z$$

$$= \int_0^L 10^{(T(r,z)-T_0(r,z))Z} \left(\frac{1}{u(r,z)}\right) \mathrm{d}z \qquad (16.12)$$

To derive this equation a variable transformation was performed. This equation assumes that bacterial spores remain on the streamlines throughout the process. The lethality given by equation (16.12) is used for the slowest heating zone, i.e. the centre line. A bulk (volume average) lethality is often estimated as follows:

$$\hat{F}(z) = \frac{\int_0^R F(r,z)u(r,z)2\pi dr}{\int_0^R u(r,z)2\pi dr} = \frac{\int_0^R F(r,z)u(r,z)2\pi dr}{\dot{Q}} \qquad (16.13)$$

For first order kinetics of nutrient destruction the same equations can be used to calculate the nutrient retention. The lethality and the average lethality can be estimated by using the appropriate Z value. Typically for nutrient retention bulk lethality is estimated. If the velocity and temperature fields are known, one can use the above equation to estimate the microbial death and the nutrient retention.

Viscosity is a strong function of temperature: the flow pattern and thus the residence time at different radii depends on the temperature distribution. There has been no reported analytical solution for the above system of equations for temperature dependent viscosity for a pipe flow. Simpson and Williams (1974) developed an analytical method for the design of aseptic processes by neglecting the effect of temperature on viscosity. The authors suggested a total dimensionless length ($\zeta = z\alpha/(R^2 U_{avg})$ of 1.2, with 0.8 for the heating section and 0.4 for the cooling (a holding section was not considered). Based on their calculations this value is accurate within $\pm 2\%$ for all power law fluids with $0.3 \leq n \leq 1$ and $-4 \leq \psi < 4$, where ψ characterises the strength of dependence on temperature. The dimensionless length was estimated to give appropriate inactivation of *Clostridium botulinum* spores at the centreline. It is important to note that the structure of the dimensionless length ζ, i.e. the length of the pipe used during aseptic processing, should increase proportionally to the square of the pipe radius. This relationship does not consider any effects of the wall temperature.

Kumar and Bhattacharya (1991a) used a commercial finite element program (FIDAP) to simulate the process for a shear thinning fluid with temperature dependent viscosity. Lethality was calculated along the centreline. The length of the heater was selected through a trial so a sterility of at least 6 min was reached at the exit. The heater was followed by a 10 m cooling section. The mesh used consisted of 25 nodes in the radial direction not equally distributed. In the axial direction 100 nodal points were used for the first two metres of the heating section, for the rest of the pipe the spacing between two nodes was equal to 0.025 m. Nine node isoparametric elements were used. The variables considered

in this study were the wall temperature in the heating section, the tube diameter and the flow rate. As expected the length of the heating section increased with an increase in flow rate, for a given temperature and diameter but decreased with an increase in temperature.

Jung and Fryer (1999), again using FIDAP, demonstrated how to optimise the quality of food in a sterilisation process for a Newtonian and a shear thinning fluid. Both lethality and quality were estimated as bulk averages. The agreement of numerical predictions with various analytical solutions was tested. A constant length of 12 m was considered for the heater while the length of the holding section was estimated so that a final sterility of 3 min was reached at the exit. Fifty elements were used in the radial direction. The authors concluded that the common practice of the food industry to estimate lethality, assuming a Newtonian isothermal flow could lead to significant overprocessing. In addition, the conventional High Temperature Short Time treatment could fail under some circumstances. A heater temperature of about 170ºC was found to be optimal for the conditions studied; this lies above the usual operating temperature for food processes.

Recently Liao *et al.* (2000), again using FIDAP, investigated the sterilisation of a starch suspension. The effect of gelatinisation on viscosity was included. Microbial death was calculated at the shortest heating zone, while a volume average was used for the nutrient retention. The authors did not include any heat needed for the starch gelatinisation in the energy equation. In addition, even at temperatures higher than the gelatinisation temperature, where the viscosity is relatively low, velocity distributions appear to be fairly uniform, indicating high viscosity values.

The problem of sterilisation of liquid foods in tubular heat exchangers is well understood. Future areas of interest include studies of foods with complex rheological properties, such as slip on the wall and yield stress. In addition, application of the existing models to industrial situations for optimisation is expected to minimise cost, and improve the nutrient content and quality of the processed foods.

16.3.2 Sterilisation of foods containing particulates

Especially since approval has been granted by the FDA, aseptic processing of foods containing particulates has become of great interest to the food industry. There are a number of aseptically processed foods, such as soups, that contain solid fractions up to 60% with sizes typically between 3 and 20 mm (Lareo *et al.*, 1997a,b). Aseptic processing is expected to lead to a new range of self-stable products in flexible packages that are preferable to the institutional and food service sector when compared to the commonly used #10 cans.

Although the flow of a single particle has been studied extensively, particulate flows are not very well understood (Lareo *et al.*, 1997c). Complicated particle-particle interactions and the often-complex rheological properties of the liquid, result in a non-uniform and often unpredictable flow. In order to study thermal treatment of foods, data such as the heat transfer coefficient between the fluid and

the particle are needed. Although a number of correlations for the heat transfer coefficient have been published in the literature (Barigou *et al.*, 1998) the accuracy of these models for food systems is questionable. Furthermore, particles, as with the fluid, experience a wide range of residence times. In order to ensure a safe product the time temperature history of the fastest moving particle should be considered. Locating this particle in real-life cases is not trivial. As a result it is often a common practice in the food industry to assume a laminar particle flow, which results in overprocessing.

In order to model particulate flows a set of equations describing the motion of the fluid and the particles has to be solved simultaneously. The motion of fluid is described from the continuity and momentum equations (equations (16.4) and (16.5)), while the motion of particles is given from a force balance on each particle. A description of the forces acting on the particles is given by Sastry *et al.* (1989). The energy equations for the particles and the fluid have to be solved to estimate temperature at various locations. One has to keep in mind that the process is transient. From the above it is clear that solving the full problem for realistic process conditions poses a tremendous computational challenge.

A number of investigators have tried to simulate the process, often using over-simplifying assumptions. In an early attempt, Manson (1974) assumed infinite fluid-particle heat transfer coefficient and demonstrated the importance of the residence time distribution. Larkin (1989) suggested a modification of Ball's method, in order to predict the sterility in the middle of a particle. Sastry (1986) developed energy balances for a continuous steriliser. Thermal balances were used to estimate the temperature profile along the tube, assuming that solid and liquid velocities were the same, and the effect of these profiles on different particles (size and residence time) was studied. The author concluded that most of the lethality is taking place in the holding tube and that increasing particle size requires a longer tube length.

Mankad *et al.* (1995) presented a model to study the importance of slip velocity, i.e. the difference between the velocity of the liquid and the solid. The model was based on energy and mass balances for the liquid and the solid phases. The resulting equations are as follows:

Energy balance for the liquid:

Heat change by thermal diffusion — Heat exchange with the wall

$$\underbrace{\frac{\partial T_f}{\partial t}}_{\text{Accumulation of heat}} = \overbrace{\alpha_{ax}\frac{\partial^2 T_f}{\partial x^2}}^{\text{Heat change by thermal diffusion}} - \underbrace{v_f\frac{\partial T_f}{\partial x}}_{\text{Heat change by convective fluid flow}} + \frac{1}{(C\rho)_f(1-\phi_{act})}\overbrace{\left[\frac{2h_w}{R_t}(T_w - T_f) - \underbrace{h_p a(T_f - T_s)}_{\text{Heat exchange with a particle}}\right]}^{\text{Heat exchange with the wall}} \quad (16.14)$$

Energy balance for the particle (heat conduction):

$$\frac{\partial T_s}{\partial t} = \alpha_s \frac{1}{R_p^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial T_s}{\partial r} \right) \tag{16.15}$$

The equations were solved as an initial boundary value problem, for steady state conditions using implicit and explicit finite difference methods. The effect of the slip velocity and fraction of the delivered solids on the length of the heating tube required to heat the solid and the liquid to 100ºC and the residence time were examined. The authors showed that increase of slip velocity increased the fluid-particle heat transfer and reduced the required heating tube length while increasing the delivered solids concentration had the opposite effect. Limitations of the model arise from the assumption of uniform radial profile for the velocity and the temperature. Furthermore, the authors acknowledged the difficulty in predicting accurately the heat transfer coefficient between the particle and the fluid. This model was extended by Mankad and Fryer (1997) to account for more realistic flows occurring in particulate foods. The flow was divided in two regimes, a sedimented bed with a low velocity and a low fraction region above it. The temperature and the velocity of the fluid were uniform in each zone. The model reveals the importance of slip velocity upon the temperatures of the two phases. Flows where the velocity differences were minimised appear to be best in terms of process time.

Recently Sandeep *et al.* (2000) simulated an isothermal two-phase flow in straight and helical holding tubes. The motion of fluid was described by the continuity and the three momentum equations. A source term added in the momentum equations accounted for the effect of the particle on the fluid. The translation of particles was predicted from three linear dynamics equations and the rotation from three angular dynamics equations. The equations were solved using a finite difference scheme (fourth-order four-stage explicit Runge-Kutta). Although the authors used a shear-thinning fluid (CMC) the form of Navier-Stokes used was for constant properties. No boundary conditions were used for the fluid particle interface. In general, it is not clear if the authors differentiated between the nodes used for the fluid and the particles, and as a result it could be possible for the fluid and the particles to share the same physical space. This limits the application to low particle concentrations. The authors demonstrated that the existence of particles enhanced the secondary flow that improves the mixing characteristics and product uniformity. An increase of the particle diameter or flow rate appeared to narrow the residence time distribution (RTD). The geometrical characteristics of the tubes appeared to have minimal effect on the RTD (less than one second for both average residence time and difference between fastest and slowest moving particle).

Aseptic processing has great industrial potential. The limited understanding of the physics makes optimisation very challenging, in practice, over processed products, with inferior quality characteristics often being produced. The problem of sterilisation of liquid foods in tubular heat exchangers is well understood. Future areas of interest include studies of foods with complex rheological

properties, such as slip on the wall and yield stress. Experimental investigation of particulate flows will give greater insight to the phenomena occurring and provide means to validate existing numerical simulations. One area of research that shows promise is the use of magnetic resonance imaging (MRI) to provide temperature maps of foodstuffs in the sterilisation of foods containing particulates (Sun *et al.*, 1993, 1994; Hulbert *et al.*, 1997; Kantt *et al.*, 1998). Increasing computational power will provide the means to handle multiple particles flowing under non-isothermal conditions. In addition, application of existing models to industrial processes for optimisation will minimise cost, and improve the nutrient content and quality of the processed foods.

16.4 Heat generation methods: ohmic and microwave heating

Conventional thermal processes are limited to the standard three ways of heating, by convective, conductive and radiative transport. In practice, conduction is a slow process, which limits the practical heating rates of solid foods and of foods in packages, such as cans. Alternative methods have been sought in which heat is supplied using other techniques; here, *heat generation* methods are briefly considered. In these, heat is generated by the material in situ as the result of interaction with an external field. This field can be applied by shear (for example, the viscous heating of a solid during shear in an extruder barrel) or by some external electric field. The two best-studied examples of heat generation processes are those of microwave and ohmic heating, in which external electric fields are used.

In microwave heating a high frequency field is passed through the food, stimulating the vibrational frequencies of chemical bonds to heat the material: details of the process are found in the excellent review of Metaxas (1996). In ohmic heating, an electric current is passed through a food material which then heats as a result of its inherent electrical resistance; a review is given by Fryer and Davies (2001).

Modelling these processes is useful for process and product designers: the need is to show that uniform heating can be provided to commercial products, to ensure safety and optimise product quality. In both cases, the conventional thermal conduction equation within a solid:

$$\rho c_p \frac{\partial T}{\partial t} = \lambda \nabla^2 T \tag{16.16}$$

must be modified by inclusion of a source term Q, the amount of heat generation per unit volume:

$$\rho c_p \frac{\partial T}{\partial t} = \lambda \nabla^2 T + Q \tag{16.17}$$

Within a liquid, a convective mixing term must be included; where there is fluid motion or relative motion between particles and liquids the full Navier-Stokes equation must be solved in the appropriate geometry. This is complex in itself,

even if the heating term were constant. However, the heating term results from the presence of the field; variations in local field strength can thus result in different local heating rates. It is necessary to solve simultaneously for both the electric field and for the thermal field that results; this is difficult both theoretically and computationally.

The source term in microwave heating is due to the interaction of the external field and the material; this is a function of the field frequency and strength and the ability of the food to absorb the microwave energy. The two Maxwell's equations used in deriving the field equations for microwaves are Ampere's law:

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} + \frac{\partial \mathbf{D}}{\partial t} \tag{16.18}$$

relating the variation of the magnetic field **H** to the electric flux density **D** and the electric field **E**, and Faraday's law

$$\nabla \times \mathbf{E} = \frac{\partial \mathbf{B}}{\partial t} \tag{16.19}$$

which relates **E** to the flux density **B**. For a dielectric material this becomes the wave equation:

$$\nabla^2 \mathbf{E} = \mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{16.20}$$

and the power dissipated per unit volume is

$$Q = \frac{1}{2}\sigma_e |E|^2 \tag{16.21}$$

The source term in electrical resistance (ohmic) heating is of the same form but is here due to the resistance of the food:

$$Q = \sigma E^2 = \sigma(\nabla V)^2 \tag{16.22}$$

where E is the voltage gradient and σ the electrical conductivity. To calculate this requires solution of Laplace's equation for the voltage field within a system in which the electrical conductivity varied with position:

$$\nabla \cdot (\sigma \nabla V) = 0 \tag{16.23}$$

throughout the material. In both cases the equations are coupled through the temperature dependence of the physical properties of the system such as electrical conductivity and permittivity.

Temperature variation results from variation in the field: in microwave heating there is a penetration depth as a result of absorption of the external field, which is generally written as

$$D_p = (4.8/f)\sqrt{\varepsilon'/\varepsilon''} \tag{16.24}$$

where the penetration depth is in centimetres when the frequency f is in GHz, and ε' and ε'' are the dielectric constant and the dielectric loss factor respectively.

In electrical heating the heating variation occurs as a result of inhomogeneous distribution of the electric field, which is distorted by the presence of different electrical conductivities, such as conductors and insulators. Current flows around an insulator, and through a conductor; this distorts the uniformity of the field.

A number of attempts have been made to model these processes and to demonstrate the complexity of the heating patterns that can result; these are summarised below.

16.4.1 Microwave heating

Excellent discussions of the problems inherent in modelling microwave processes, and in validating them for commercial production, are given by Bows (2000) and Bows *et al.* (1999). The interaction between the applied field and the food material is a key issue. Multimode resonant applicators, used in all domestic ovens and almost all industrial applications produce one resultant heating pattern from the complex interaction of the field, material and applicator. Strategies to minimise temperature differences (rotating turntables, mode stirrers or moving the product) are not always sufficient to overcome undesirable effects. These can include runaway heating when thawing frozen foods (Buffler and Stanford, 1995), centre focusing in spheres and cylinders (Ohlsson and Risman, 1978) and edge or corner overheating in trays (Bows and Richardson, 1990). For frozen food heated in a domestic oven, thermal runaway heating is observed when thawed areas preferentially absorb a greater proportion of the microwave field than the remaining frozen areas. Thermal runaway is chiefly a property effect; as a material's dielectric properties change (with temperature for some food materials, but particularly on thawing), the electric field pattern also changes. Bows *et al.* (1999) describe a novel method of microwave heating, referred to as phase control. The method uses constructive interference techniques: when the microwave fields interacting within a product are coherent (vector addition heating), many heating patterns can be generated at an instant in time, and more controlled heating of food can be carried out.

A 3D finite element time domain code, using edge elements, was used to simulate phase controlled heating using the method developed by Dibben and Metaxas (1994). The model solves Maxwell's equations in 3D and includes a waveguide input feed with a surface excitation plane. The time domain was used because it was found that the conductivities of foods caused ill-conditioned matrices in the frequency domain.

Good agreement between the images and experiments was found; however, the code was complex, with a mesh of 73 500 tetrahedral elements in the foodstuff and 135 000 tetrahedral elements in the whole solution domain. A Silicon Graphics Indigo 2 XZ with an R4000 processor and 128 MB RAM took 85 hours of CPU time to obtain a complete solution. This is a valuable tool but, unlike conventional heating, requires advanced computers to solve the model. Nott *et al.* (1999) used MRI to map quantitatively in three dimensions the

complex temperature distributions induced by microwave heating of food materials: results are compared with infrared thermal images. MRI is limited to non-metallic systems such as paper, plastic, ceramic and glass, but this is an approach which allows non-invasive sampling.

16.4.2 Ohmic heating

Although ohmic heating is thought of as a novel process, its use in food processing goes back a century (de Alwis and Fryer, 1990). Over the last twenty years, interest in the process has followed the development of a commercial unit by APV Baker (Parrott, 1992). In this process, food is pumped past a series of electrodes connected to three-phase supply at 50–60 Hz. Heating rates of the order of 1ºC/s are possible, in field strengths on the order of 10 V/cm. The process has found use in sterilisation of high-solids fraction (30–40% solids) mixtures as well as the production of high-value materials such as pasteurised fruit pieces for yoghurts.

Numerical models of electrical heating have concentrated on a set of problems of different length and timescales:

- Heating rates within a solid-liquid mixture (such as de Alwis *et al.*, 1989; Palaniappan and Sastry, 1991).
- The types of temperature pattern found inside solid-liquid mixtures (such as Fryer *et al.*, 1993; Kemp *et al.*, 1999).
- Predicting the temperatures of mixtures undergoing ohmic heating (such as Zhang and Fryer, 1993, 1994; Benabderrahmane and Pain, 2000).

Laplace's equation for the electric field can only be solved analytically in very simplistic cases, such as for an isolated sphere or infinite cylinder of constant physical properties in a uniform field. de Alwis *et al.* (1989) show that for such cases the ratio of the heating rate in a solid and undisturbed liquid is given by:

$$R_Q(sphere) = \frac{9\sigma_s\sigma_L}{(\sigma_s + 2\sigma_L)^2}; \quad R_Q(cylinder) = \frac{4\sigma_s\sigma_L}{(\sigma_s + 2\sigma_L)^2} \qquad (16.25)$$

The heating rate in the cylinder can thus never exceed that of the surrounding fluid, whilst that in the sphere can exceed that of the liquid for $1 < \sigma_s/\sigma_L < 4$. That ohmic heating can result in solids that overheat the liquid was shown by de Alwis *et al.* (1989). In general, for non-uniform shapes and non-uniform physical properties, computer models are needed to solve for the heating patterns. The first interaction between the particle size and shape and the electric field was shown by de Alwis and Fryer (1990), using a code written especially for the problem. It is now possible to use commercial codes, such as ANSYS (Zhang and Fryer, 1994) or FIDAP, with which Kemp *et al.* (1999) validate a computational model for the heating pattern around an insulating particle. Here, both simulation and experiment show that it is possible to have over- and underheating in the same particle; underheating results from the region of low

electric field behind the insulating particle, and overheating is due to the high conductivity of the second particle. Local heating and cooling effects have been shown by thermocouple and in elegant MRI experiments by Ruan *et al.* (1999).

Various models have been proposed to identify the coldest spot in an ohmic formulation: finding this is key to deciding what a process should be. In conventional processing, of course, the coldest spot will occur in the solid; with the correct choice of solid and liquid electrical conductivity, the particles will overheat the liquid, and the coldest spot will be found in the fluid. The amount of fluid mixing is then critical; high viscosity fluids have a higher range of temperature differences between them than less viscous ones (see Fryer *et al.*, 1993).

Models for the whole process have been developed which involve a series of simplifications. Zhang and Fryer (1993, 1994) used a finite-element model to predict the behaviour of a sphere in a well-mixed fluid, and then used that model as the basis for a Fortran model of a flowing mixture.

Heat generation techniques have found some applications in industry – obviously microwave ovens have achieved significant penetration into the kitchen as a result of the rapidity of heating possible. Developments in modelling will be used (i) to make industrial application easier, and (ii) to make products which are of higher quality when heated in the domestic microwave oven.

16.5 Developments in the field

The modelling of thermal processing of foods is an active research subject around the world. The subject is developing in a number of areas:

16.5.1 Ease of solving models

Advances in computing continue to make it more straightforward to run the types of programs that are needed to solve these problems. As shown above, FE software can now run efficiently on a PC, whereas only a few years ago it required workstation or mainframe capabilities. It is likely that problems that currently are at the limits of computing power, such as the efficient modelling of microwave heating, will be simple to solve in a few years time. This means that models could be used as the basis for real-time control systems: at the moment most run too slowly.

16.5.2 Realistic physical properties

Many of the papers described above have treated systems with simplistic physical properties, such as constant thermal diffusivity. In practice, many physical properties vary with temperature; the strong variation of viscosity and electrical properties has been shown in some of the papers described above to lead to very strongly coupled problems, where the thermal and other fields have

to be solved together. To do this accurately requires accurate data: the more accurate the data, the better the fit of the model to reality. In some cases, it is the lack of data rather than the lack of appropriate theory that is limiting the accuracy of the models.

16.5.3 New types of model

The models described above have used finite element, volume or difference approaches to solve heat transfer models. Other types of approach are being used in some areas, such as discrete element and lattice Boltzmann modelling of single- and two-phase flows in complex geometries, and cellular automata models for the interaction of microbes and foods. These different types of model have some advantages, particularly where FE meshes are complicated and change with time, as in particle flows. At the moment, they are difficult to use, and commercial codes are not well developed; as these types of models become simpler to use they will be applied to food problems.

16.5.4 Kinetic models for food processes

Given accurate physical property data and computer codes, thermal models are capable of predicting the temperatures throughout food solids and liquids. However, these predictions cannot be used fully if information on the rates of the processes affected by temperature are not available. The microbiological models used in many thermal models are probably simplistic; if equation (16.1) does not describe the effect of temperature on a microbial population then the results of the model will not apply. More research is needed to determine models for microbial growth and inactivation, and on the rates of development of texture and flavour in foods. This type of data, combined with effective thermal models, will lead to the production of models which can predict food quality and safety, and act as the basis for optimising production.

Some of our recent work (Bakalis, Cox and Fryer, unpublished results) has studied ways of validating flow and heating models. The School of Physics at the University of Birmingham, UK, has developed a unique way of following flows in opaque fluids using tracers, the so-called Positron Emitting Particle Tracking (PEPT, Parker *et al.*, 1993). In this technique a radioactive tracer particle is monitored during its passage through a system. The tracer emits positrons, which then generate a pair of back-to-back gamma rays on collision with an electron (an extremely rapid and geometrically close annihilation event). By detecting the trajectories of the gamma rays (with a 180^{o} of separation) the position and velocity of the tracer can be followed by triangulating their position at many times per second. The technique can also be used in pilot-scale equipment as the gamma rays can penetrate reasonable thicknesses of metal. This technique allows flow patterns to be followed in real systems.

As an example of how it can be used in foods, PEPT was used to examine particle paths in a canning process. A typical metal can was filled to 90% of its

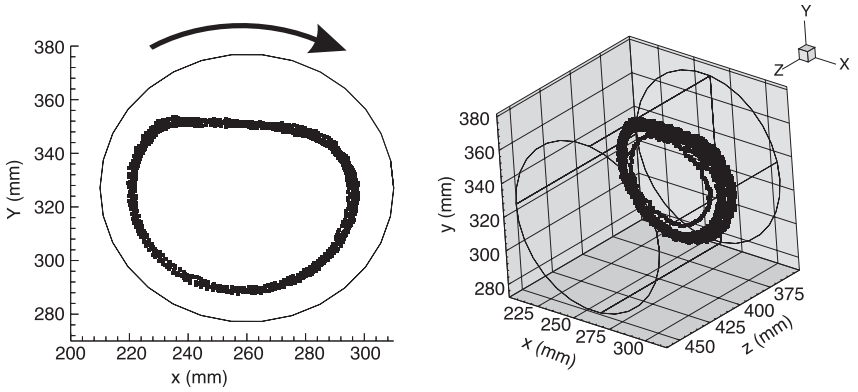


Fig. 16.2 PEPT traces of a metal can filled with soup.

volume with vegetable soup obtained from a local market, and a 600 micron wide tracer particle was inserted in a piece of potato from the soup and was followed as the can was rotated axially at a speed of 25 rpm. Fig. 16.2 represents the particle path over a period of three minutes. It can be seen that the particle follows a D shape trajectory with the headspace affecting the flow. This pattern is very different when compared to a fully filled can, where the particles are moving in circles. Overall PEPT appears to be a promising technique that can give invaluable insight on thermal processing and provide means to validate numerical simulations. This type of data, coupled with the modelling techniques described here, will lead to a better predictive understanding of food processes.

16.6 References

ABDUL GHANI A G, FARID M M, CHEN X D, RICHARDS P (1999a). Numerical simulation of natural convection heating of canned food by computational dynamics. *J. Food Eng.*, **41**, 55–64.

ABDUL GHANI A G, FARID M M, CHEN X D, RICHARDS P (1999b). An investigation of deactivation of bacteria in a canned liquid food during sterilization using computational fluid dynamics. *J. Food Eng.*, **42**, 207–14.

AKTERIAN S G (1999). Online strategy for compensating for arbitrary deviations in heating medium temperature during batch thermal sterilization processes. *J. Food Eng.*, **39**, 1–7.

AKTERIAN S G, SMOUT C, HENDRICKX M E, TOBBACK P P (1998). Application of sensitivity functions for analysing the impact of temperature non-uniformity in batch sterilisers. *J. Food Eng.*, **37**, 1–10.

BALL C O (1923). Thermal process time for canned foods. *Bulletin of the National Resources Council*, **7**, 1.

BARIGOU M, MANKAD S, FRYER P J (1998). Heat transfer in two-phase solid-liquid food flows: A review. *Trans. I.ChemE.*, **76**, C, 3–29.

BELLARA SR, FRYER PJ, MCFARLANE CM, THOMAS CR, HOCKING PM, MACKEY BM (1999). Visualization and modelling of the thermal inactivation of bacteria in a model food. *Appl. Env. Microbiol.*, 65, 3095–9.

BENABDERRAHMANE Y, PAIN J-P (2000). Thermal phase behaviour of a solid/liquid mixture in an ohmic heating sterilizer – slip phase model – sensitivity to process parameters. *Chem. Eng. Sci.*, **55**, 1371–84.

BIRD RB, STEWART WE, LIGHTFOOT, EW (1964). Transport Phenomena. John Wiley & Sons, Inc.

BOWS JR (2000) A classification system for microwave heating of food. *Int. J. Food Sci. Technol.*, **35**, 417–30.

BOWS JR, RICHARDSON PS (1990). Effect of component configuration and packaging materials on microwave reheating a frozen 3-component meal. *Int. J. Food Sci Technol*. **25**, 538–50.

BOWS JR, PATRICK ML, JANES R, DIBBEN DC (1999). Microwave phase control heating. *Int. J. Food Sci. Technol.*, **34**, 295–304.

BUFFLER CR, STANFORD MA (1995). Effects of dielectric and thermal properties on the microwave heating of foods. *Microwave World*, **16**, 5–10.

CAMPBELL GM, WEBB C, PANDIELLA SS, NIRANJAN K (1999). Bubbles in Food. Eagan Press.

CARSLAW HS, JAEGER JC (1980) Conduction of Heat in Solids. Oxford University Press.

DE ALWIS AAP, FRYER PJ (1990). A finite element analysis of heat generation and transfer during ohmic heating of food. *Chem. Eng. Sci.*. **45**, 1547–60.

DE ALWIS AAP, HALDEN K, FRYER PJ (1989). Shape and conductivity effects in the ohmic heating of foods. *Chem. Eng. Res. Des.*, **67**, 159–68.

DIBBEN DC, METAXAS AC (1994). Finite element time domain analysis of multimode applicators using edge elements. *J. Microwave Power and Electromagnetic Energy*, **29**, 242–51.

FRYER PJ, DAVIES LJ (2001). Modelling electrical resistance ('ohmic') heating in foods. To be published in Irudayaraj, J. (ed.), *Food Process Operations Modelling: design and analysis*. Marcel Dekker.

FRYER PJ, DE ALWIS AAP, KOURY E, STAPLEY AGF, ZHANG L (1993). Ohmic heating of solid-liquid foods; heat generation and convection effects, *J. Food. Eng.*, **18**, 101–25.

HAYAKAWA K (1969). New parameters for calculating mass average sterilizing value to estimate nutrients in thermally active foods. *Can. Inst. Food Technol. J.*, **2**, 165–72.

HENDRICKX M, SILVA C, OLIVEIRA F, TOBBACK PP (1993). Generalised (semi-)empirical formulae for optimal sterilisation temperatures of conduction-heated foods with infinite surface heat transfer coefficients. *J. Food Eng.*, **19**, 141–58.

HULBERT GJ, LITCHFIELD JB, SCHMIDT SJ (1997). Determination of convective heat transfer coefficients using 2D MRI temperature mapping and finite element modelling. *J. Food Eng.*, **34**, 193–201.

INCORPERA FP (1981). Fundamentals of Heat Transfer. Wiley.

JUNG A, FRYER PJ (1999). Optimising the quality of safe food: Computational modelling of a continuous sterilisation process. *Chem. Eng. Sci.*, **54**, 717–30.

KANTT CA, SCHMIDT SJ, SIZER CE, PALANIAPPAN S, LITCHFIELD JB (1998). Temperature mapping of particles during aseptic processing with magnetic resonance imaging. *J. Food Sci.*, **63**, 305–11.

KEMP M, DAVIES L, FRYER PJ (1999). The geometry of shadows: effects of inhomogeneities in electrical field processing. *J. Food Eng.*, **40**, 245–58.

KIM KH, TEIXEIRA AA (1997). Predicting internal temperature responses to conduction-heating of odd shaped solids. *J. Food Process Eng*., **20**, 51–63.

KUMAR A, BHATTACHARYA M (1991a) Numerical analysis of aseptic processing of a non-Newtonian liquid food in a tubular heat exchanger. *Chem. Eng. Comm.*, **103**, 27–51.

KUMAR A, BHATTACHARYA M (1991b). Transient temperature and velocity profiles in a canned non-Newtonian liquid food during sterilisation in a still cook retort. *Int. J. Heat Mass Transfer*, **34**, 1083–96.

KUMAR A, BHATTACHARYA M, BLAYLOCK J (1990) Numerical simulation of natural convection heating of canned thick viscous food products. *J. Food Sci.*, **55**, 1403–11, 1420.

LAREO C, BRANCH CA, FRYER PJ (1997a). Particle velocity profiles for solid-liquid food flows in vertical pipes. 1. Single particles. *Powder Tech*, **93**, 23–34.

LAREO C, NEDDERMAN RM, FRYER PJ (1997b). Particle velocity profiles for solid-liquid food flows in vertical pipes. 2. Multiple particles. *Powder Tech*, **93**, 35–45.

LAREO C, FRYER PJ, BARIGOU M (1997c). The fluid mechanics of two-phase solid-liquid food flows: A review. *Trans. I.ChemE.*, **75**, C, 73–105.

LARKIN JW (1989). Use of a modified Ball formula method to evaluate aseptic processing of foods containing particulates. *Food Technol.*, **43**, 124–31.

LIAO HJ, RAO MA, DATTA AK (2000). Role of thermo-rheological behaviour in simulation of continuous sterilization of a starch dispersion. *Trans. I.ChemE.*, **78**, C, 48–56.

LOPEZ A (1987). Heat process determinations. In: *A Complete Course in Canning and Related Processes*, Vol 2, 21–30. The Canning Trade Inc., Baltimore.

MACKEY BM, DERRICK CM (1987). Changes in the heat resistance of *Salmonella typhimurium* during heating at rising temperatures. *Lett. Appl. Microbiol.*, **4**, 13–16.

MANKAD S, FRYER PJ (1997). A heterogeneous flow model for the effect of slip and flow velocities on food sterilser design. *Chem. Eng. Sci.*, **52**, 1835–43.

MANKAD S, BRANCH CA, FRYER PJ (1995). The effect of particle slip on the sterilization of solid-liquid food mixtures. *Chem. Eng. Sci.*, **50**, 1323–36.

MANSON JE (1974). Thermal process simulation for aseptic processing of foods containing discrete particulate matter. *J. Food Sci.*, **39**, 1084–9.

METAXAS AC (1996). *Foundations of Electroheat: A Unified Approach.* John Wiley & Sons, Inc.

NOTT KP, BOWS JR, PATRICK ML (1999). Three-dimensional MRI mapping of microwave induced heating patterns. *Int. J. Food Sci. Technol.*, **34**, 305–17.

OHLSSON T, RISMAN P O (1978). Temperature distribution of microwave heating, spheres and cylinders. *J. Microwave Power*, **13**, 303–10.

ÖZIŞIK M N (1993). *Heat Conduction*, 2nd edition. John Wiley & Sons, Inc.

PALANIAPPAN S, SASTRY S K (1991). Modelling of electrical conductivity of liquid-particle mixtures. *Trans. I.ChemE.*, **69**, C, 167–74.

PAN Z, SINGH R P, RUMSEY T R (2000). Predictive modelling of contact-heating process for cooking a hamburger patty. *J. Food Eng.*, **46**, 9–19.

PARKER D, BROADBENT C J, HAWKESWORTH M R, MCNEIL P (1993) Position emission particle tracking – a technique for studying flow within engineering equipment. *Nuc. Inst. Meth. A.*, 236–92.

PARROTT D L (1992). Use of ohmic heating for aseptic processing of particulates. *Food Technol.*, **46**, 68–72.

QUINTAVALLA S, CAMPANINI M, (1991). Effect of rising temperature on the heat resistance of *Listeria monocytogenes* in meat emulsion. *Lett. Appl. Microbiol.*, **12**, 184–7.

RAMASWAMY H S, ABDELRAHIM K A, SIMPSON B K, SMITH J P (1995). Residence time distribution (RTD) in aseptic processing of particulate foods: a review. *Food Res. Int.*, **28**, 291–310.

RUAN R, CHEN P, CHANG K, KIM H-J, TAUB I A (1999). Rapid food particle measurement temperature mapping during ohmic heating using FI-ASH MRI. *J. Food Sci.*, **64**, 1024–6.

RYCKAERT V G, CLAES J E, VAN IMPE J F (1999). Model-based temperature control in ovens. *J. Food Eng.*, **39**, 47–58.

SANDEEP K P, ZURITZ C A, PURI V M (2000). Modelling non-Newtonian two phase flow in conventional and helical tubes. *Int. J. Food Sci. and Tech.*, **35**, 511–22.

SASTRY S (1986). Mathematical evaluation of process schedules for aseptic processing of low-acid foods containing discrete particulates. *J. Food Sci.*, **51**, 1323–8.

SASTRY S, HESKITT B, BLAISDELL J (1989). Experimental and modelling studies on convective heat transfer at the particle-liquid interface in aseptic processing systems. *Food Tech.*, **43**, 132–6.

SIMPSON S G, WILLIAMS M C (1974). An analysis of high temperature/short time sterilization during laminar flow. *J. Food Sci.*, **39**, 1047–54.

SUN X Z, LITCHFIELD J B, SCHMIDT S J (1993). Temperature mapping in a model food gel using magnetic resonance imaging. *J. Food Sci.*, **68**, 168–72, 181.

SUN X Z, SCHMIDT S J, LITCHFIELD J B (1994). Temperature mapping in a potato using half Fourier transform MRI of diffusion. *J. Food Proc. Eng.*, **17**, 423–37.

TEIXEIRA A A, ZINSMEISTER J W, ZINSMEISTER G E (1969). Computer simulation of variable retort control and container geometry as a possible means of improving thiamine retention in thermally processed foods. *J. Food Sci.*, **40**, 656–9.

TEWKESBURY H, STAPLEY A G F, FRYER P J (2000). Modelling temperature distributions in cooling chocolate moulds. *Chem. Eng. Sci.*, **55**, 3123–32.

TIMBERS G E, HAYAKAWA K (1967). Mass average sterilizing value for thermal process. Part 1: Comparison of existing procedures. *Food Technol.*, **21**, 1069.

TOLEDO R T (1991). Thermal process calculations. In: *Fundamentals of Food Process Engineering*, 2nd edition, 315–97. Van Nostrand Reinhold.

TUCKER G S, HOLDSWORTH S D (1991). Mathematical modelling of sterilisation and cooking processes for heat preserved foods. *Trans. I.ChemE.*, **69**, C, 5–12.

VARGA S, OLIVERIA J C, SMOUT C, CHANTAL C, HENDRICKX M E (2000). Modelling temperature variability in batch retorts and its impact on lethality distribution. *J. Food. Eng.*, **44**, 163–74.

VERBOVEN P, NICOLAÏ B M, SCHEERLINCK N, DE BAERDEMAEKER J (1997). The local surface heat transfer coefficient in thermal food process calculations: a CFD approach. *J. Food Eng.*, **33**, 15–35.

VERBOVEN P, SCHEERLINCK N, DE BAERDEMAEKER J, NICOLAÏ B M (2000). Computational fluid dynamics modelling and validation of the temperature distribution in a forced convection oven. *J. Food Eng.*, **43**, 61–73.

ZHANG LI, FRYER P J (1993). Models for the electrical heating of solid-liquid food mixtures. *Chem. Eng. Sci.*, **48**, 633–43.

ZHANG LI, FRYER P J (1994). Food processing by electrical heating; the sensistivity of product sterility and quality to process parameters. *AIChEJ*, **40**, 888–98.

Part V

Applications: safety and quality in the food chain

Introduction

In the last part of this book on food process modelling, the application of the models and techniques discussed in the previous parts are used in a larger frame of view. To apply these models in real world systems, these models and their applications have to be considered in their interaction with real world systems. All these applications have in some way to do with product quality and product safety, all driven by consumer behaviour, wishes and acceptance.

To increase the application of models dedicated to problems local to a product attribute or a situation of processing, we need to put these dedicated models into a larger framework. Not only these local aspects have to be considered, but also their interactions with the ever-changing conditions of handling and distribution, and the ever-changing response over users and consumers.

In Chapter 17 the problems with quality, consumer's perception and consumer's acceptance are decomposed into a conceptual model of how quality works in general for each individual. Out of this decomposition comes a clear view on interactions between product properties and consumer's perception and acceptance. Based on these concepts, a common language on quality needed to improve communication throughout the food supply chain, from grower/producer up to and including the consumer can be deduced.

Chapter 18 devotes its attention to the difficulties of food spoilage and food safety. Predictive microbiology is a discipline of microbiology, rapidly increasing in importance and application. Just like quality, safety is such a paramount property of our food, that it is absolutely necessary to be accepted by consumers. However, unlike quality, consumers cannot perceive food safety

until it is too late. That is why predictive modelling is becoming so important in modern food processing and engineering. Predictive microbiology is moving from the largely empirical type of modelling to the more fundamental type of modelling, making adequate use of theoretical concepts with respect to growing rate and induced lag phases.

In Chapter 19, the fundamental knowledge of kinetic modelling is transformed into a practical and applicable system of temperature time indicators (TTI). The currently most common available type of TTIs are reviewed and described, along with their historical development. Based on fundamental but simplified kinetics of quality changes, the action, application and problems of TTIs are described in detail. Based on the results, a system for optimised distribution and stock rotation system is developed and presented: LSFO (least shelf-life first out).

In Chapter 20, simplified models on product quality and keeping quality are used in a broader application targeted to the optimisation of distribution centres. In this application the technical possibilities and resources available to distribution centres are taken into account to maintain product quality as good as possible while meeting the demands on delivering amount and delivery time.

In Chapter 21 the models on quality and safety and the techniques used to build them, are put into a larger perspective of possible and desirable application in marketing and consumer behaviour, mainly focused on food supply chains, their management and possibilities for optimisation.

Pol Tijskens

17

Modelling food quality

E. C. Wilkinson and L. M. M. Tijskens, ATO-BV, Wageningen

17.1 Introduction

Quality is becoming increasingly important as a determinant of food choice. As disposable income in the developed world increases, the influence of price on food choice decreases and factors such as healthfulness, convenience and quality become more important. Quality is particularly important in determining repeat purchases of food products (Steenkamp and van Trijp, 1996). Competing on quality rather than on price often has advantages from the point of view of companies. It creates customer loyalty, raises barriers to competition and reduces price elasticity (Steenkamp, 1990). However, a quality-based strategy can only succeed if the company knows what the customer understands by a good quality product and if the company can translate these demands into (technical) product specifications. In this chapter an approach to food quality modelling is described which will allow companies to *understand* the relationship between perceived quality and product characteristics, to *predict and control* quality during production and distribution, and to *optimise* quality during product development.

However, first it is necessary to define what is meant by quality. Since Roman times it has been well known that quality and taste are something personal and specific to every individual human being: '*de gustibus et coloribus non est disputandum*', or 'do not discuss colour and taste'. Although the information, contained within the product, upon which we all judge the quality of that product, is the same for everybody, the interpretation and the appreciation can be very different for different people. So a workable definition of quality has to cover both aspects: the general information in the product and the specific effect the product properties exert on different people.

A frequently used definition is that given by Juran (1974), 'fitness for use'. Kramer and Twigg (1970) provide as definition of quality: 'the composite of those characteristics that differentiate individual units of a product, and have significance in determining the degree of acceptability of that unit by the buyer'. Steenkamp (1990) gives a more elaborate definition: 'Perceived quality is an idiosyncratic value judgement with respect to the fitness for consumption which is based upon the conscious and/or unconscious processing of quality cues in relation to relevant quality attributes within the context of significant personal and situational variables.' Implicit in all three definitions is that quality is the result of the interaction between the person and the product. It will depend not only on the characteristics of the product (colour, sugar content, etc.) but also of the person (age, culture, etc.) and of the context (meal, dish, etc.). There is no one objective measure of quality and this is emphasised by authors who use terms like 'perceived quality' (Steenkamp, 1990) or 'assigned quality' (Sloof *et al.*, 1996).

Another implication of these definitions of quality is that modelling of food quality will require the collaboration of disciplines both in the social sciences (for example, consumer scientists, market researchers and economists) and in the natural sciences (for example, biochemists and physiologists).

17.2 Key principles and methods

17.2.1 Process of quality assignment

In order to model food quality it is important to understand how quality assignment takes place. The description that follows draws largely, though not exclusively, on the work of Steenkamp (1989, 1990), Sloof *et al.* (1996) and Sloof (1999).

Before consuming a food product (for example, when purchasing it in the shop) an individual forms an opinion as to the quality of the product, termed the *expected quality*. The individual bases the expected quality on *quality cues*. A quality cue is an attribute of the food product which can be perceived before purchase and consumption and which is believed to be indicative of its quality. The individual believes the cue to be highly correlated with product quality, a belief which can be based on personal experience or on information from acquaintances or the media. Examples of cues are country of origin ('the best olives come from Italy'), price ('good wines tend to be more expensive') or colour ('dark red meat has a better quality than pale meat').

Only when the product has been consumed can the individual form a final opinion about the quality, the *experienced quality*. This unidimensional measure of perceived quality is a function of the *quality attributes.* Quality attributes are all those product attributes, which are relevant for determining the quality. Which product attributes are quality attributes thus depends on the priorities of the individual who assigns the experienced quality. The quality attributes can be divided into experience attributes which are determined before and during usage

(flavour, ease of preparation, etc.) and credence attributes which are based on beliefs (nutritional value, production methods, food safety, etc.). The beliefs, in turn, can be based on information on the packaging, on information obtained from the media or from personal contacts.

Furthermore, a distinction can be made for both quality attributes and quality cues between intrinsic attributes/cues and extrinsic attributes/cues. An intrinsic attribute or cue is one which cannot be changed without changing the product itself. Examples are the taste, the vitamin content, the size. An extrinsic attribute or cue is one which is not part of the physical product, for example the price, the packaging, the brand or the supermarket where the product is purchased. Extrinsic attributes and cues are largely the domain of the marketing department. While not denying their importance for quality assignment, most quality modelling has concentrated on the intrinsic attributes and cues.

Some authors distinguish a third integration step in which an overall quality is assigned based on both the expected and the experienced quality (Poulsen *et al.*, 1996). This allows for situations where disconfirmation of expectations leads to an overall quality assignment which is different from the experienced quality. For example, the appearance of the packaging may lead to the expectation of a certain taste and therefore quality. If this is confirmed by experienced quality, then satisfaction ensues. However if the taste is less favourable than expected, then this will lead to negative, disconfirmation which may lead to the product being rejected in a situation where, if expectations had been lower, it would have been accepted (Andani and MacFie, 2000). It follows from this that producers need to pay attention not only to quality attributes but also to quality cues and their relationship with quality attributes. The interplay of attributes, cues and quality is illustrated in Fig. 17.1.

A concept closely related to quality is acceptability. When the individual decides on the acceptability of a product, he compares the (expected) quality to some criterion, termed the 'quality limit' (Tijskens, 2000). If the quality exceeds the quality limit, he accepts the product, otherwise he rejects the product. This quality limit is dependent on the personal preferences and situation of the

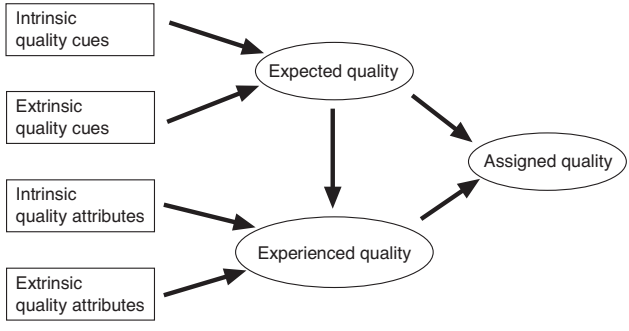


Fig. 17.1 Quality assignment based on intrinsic and extrinsic quality cues and quality attributes.

individual. Acceptability is involved in the 'keeping quality' (Tijskens and Polderdijk, 1996) which is so important for perishable products. For perishable products such as fruit and vegetables, the quality attributes change (usually deteriorate) over time. Keeping quality is a measure of the time it takes before the assigned quality falls below the quality limit at any condition during storage and transport. Shelf-life is the keeping quality under specified storage conditions (Tijskens, 2000). It follows from the above that keeping quality is no objective measure but depends on the priorities and preferences of the individual. Certainly producers should take account of differences between countries and between market segments.

For quality modelling it is necessary to consider how intrinsic quality attributes and cues are related to the product properties, defined as the physico-chemical characteristics of the product. A single attribute can be a function of several product properties. For example, the quality attribute 'perceived sweetness' can be a function not only of the amount of sugars but also of acids (Lawless and Heynmann, 1998). Several distinct steps can be identified in arriving at the quality attribute or cue (Sloof, 1999; Sloof *et al.*, 1996; Tijskens *et al.*, 1994). Firstly, the product properties form stimuli which are perceived by the human senses. For example, the taste receptors on the tongue are triggered by the acid and sugar components in the food product. Secondly, these perceptions are integrated and evaluated to form an evaluation of the intensity of the quality attribute or cue. In our example, the information from the taste receptors is combined to form an evaluation of how sweet the product is. In final appreciation step a hedonic judgement takes place. In the terms of the example, the product is judged to be not sweet enough, just right or too sweet. These steps are summarised in Fig. 17.2.

17.2.2 Quality assignment model

Introduction to quality model

If the aim of food quality modelling is summarised as the modelling of the effects of choice of cultivar or recipe, temperature and other external factors during storage and processing on perceived quality, then it is clear from the above that a number of steps are involved. As is illustrated in Fig. 17.3, the relationship between external factors and product properties, the relationship between product properties and quality attributes/cues and the relationship between quality attributes/cues and perceived quality all need to be described. This leads naturally to the decomposition of the task of food quality modelling into a number of sub-tasks (see also Chapter 2). This is best achieved by decomposing the food quality model into two main sub-tasks (Sloof, 1999):

- A quality assignment model (QAM) describing the relationship between quality attributes/cues and perceived quality. This model should also consider the influence of situation and characteristics of the individual (age, culture, etc.) on quality assignment.

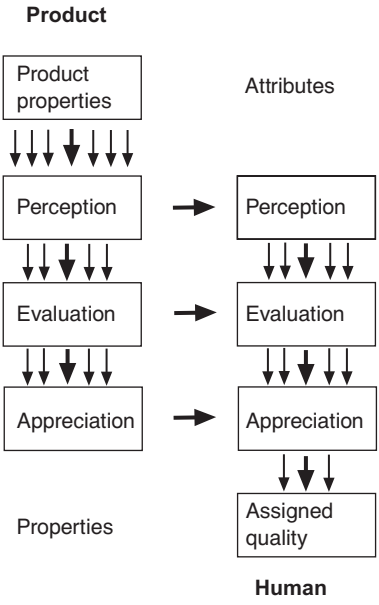


Fig. 17.2 Steps in assigning quality to a product by a human.

- A dynamic product model (DPM) describing the effect of external factors such as choice of cultivar or recipe, temperature during storage and processing on product properties and therefore quality attributes/cues.

Approaches to develop quantitative models will now be discussed for both of the sub-tasks.

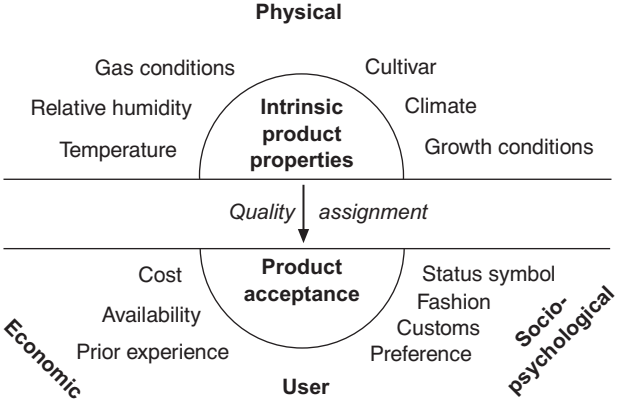


Fig. 17.3 Quality assignment as affected by product properties, economic and socio-psychological circumstances.

Product-centred approach

Molnár (1995) has developed a mathematical model to describe quality assignment. He distinguishes five categories of attributes, namely

- sensory properties
- chemical composition and physical properties
- microbiological contaminants
- toxicological contaminants
- packaging, labelling, shelf-life.

The relative importance of each category will vary according to the food product. For example, for functional foods the category 'sensory properties' may be judged relatively unimportant. Within each category, relevant attributes are identified based on the knowledge of experts. Each attribute x_i is normalised to give z_i with $1 \geq z_i \geq 0$. z_i receives the value 1 when x_i is at its optimum value and the value 0 when x_i is at its worst value or becomes unacceptable. In fact, x_i represents the evaluation of the attribute and z_i the appreciation of the attribute. To arrive at total quality, each normalised attribute is assigned a weight, and the attribute categories are also assigned weights, giving the mathematical function

$$Q = \sum_{j=1}^{n_{cat}} W_j \sum_{i=1}^{n_j} w_i z_i \tag{17.1}$$

where Q is total quality, W_j, $j = 1 \ldots n_{cat}$ are the weights for the attribute categories, w_i, $i = 1 \ldots n_j$ are the weights for the individual attributes of category j, z_i, $i = 1 \ldots n_j$ are the normalised attributes.

Molnár defines 'primary critical' attributes as attributes whose zero value indicates that the product is unfit for human consumption and that the total quality is therefore zero. Examples of primary critical attributes are sensory off-flavour attributes and food safety related attributes.

Again Molnár suggests that experts are used to determine the weights for the quality function. The fact that Molnár does not use consumers to gain information about quality attributes and the quality function is a weak point of his approach. It is the consumer who will finally assign a quality to the product and it is unlikely that the expert can fully interpret the priorities of the consumer. It is known that experts tend to consider quality as compliance with technical specifications and to lay excessive emphasis on the absence of defects compared to consumers (Tijskens *et al.*, 1994; Lawless and Heynmann, 1998).

Consumer-centred approach

A number of researchers operating in the field of marketing and consumer research have developed and quantified models for quality assignment which do place the consumer centrally (Acebrón and Dopico, 2000; Steenkamp and van Trijp, 1996; Poulsen *et al.*, 1996). The models are built around the description of the process of quality assignment given above and thus include quality cues and expected quality as well as attributes and experienced and overall quality. The

quality assignment model is quantified for a particular product by means of consumer research. Typically consumers are asked to assess uncooked, packaged products and assign scores to a pre-determined list of quality cues and to expected quality. Then they are presented with (or prepare themselves) the cooked product. Again they assess the product and assign scores to quality attributes and experienced quality. Sometimes they may be asked to assign a final overall quality score. The quality scores are related to the attribute and cue scores using a statistical technique such as linear regression, Partial Least Squares regression (Steenkamp and van Trijp, 1996) or LISREL (Poulsen *et al.*, 1996). For each step in the quality assignment process this gives one quantitative formula which describes quality assignment for all consumers. It is interesting to see that although the consumer research approach to quality assignment modelling places the consumer at the centre, no allowance is made for the subjective nature of assigned quality and for differences between consumers in their quality model.

It is not always clear in these models whether consumers are scoring the appreciation or the evaluation of the attributes and cues. In Poulsen *et al.* (1996) consumers are asked to rate the sensory attributes of cookies on a nine-point scale ranging from 'much less than ideal' through 'ideal' to 'much more than ideal'. This is clearly measuring appreciation. However, Acebrón and Dopico (2000) ask consumers to rate cooked beef on a four-point scale from very tough to very tender. This is an evaluation and it seems to be assumed implicitly that appreciation is positively and linearly related with tenderness.

There are two problems with using consumers to provide evaluations rather than appreciations. Firstly it is an unnatural task for consumers. People are naturally able to say how much they like a particular attribute, but intensive training (as a product expert, as a member of a sensory analytical panel) is required before they are reproducibly able to score how much there is present of a particular attribute. Secondly, the relationship between the evaluated amount and the appreciation is often non-linear (as reflected in the scale used by Poulsen *et al.*, 1996). The relationship between the appreciation of individual attributes and total assigned quality is probably monotone and it is plausible to assume that it is approximately linear. The modelling of linear relationships is much simpler and much less demanding in terms of the amount of data required.

It can also be questioned whether asking consumers to score both quality cues/attributes and expected/experienced quality at the same moment on the same product will give reliable results. It can lead to sociably desirable answers, with high fat or high sugar content products receiving artificially low quality scores. Also because pre-determined lists are used based on expert knowledge, it may bring the attention of the consumer to an attribute or cue which in a real-life situation would have played no part in his quality assignment. An alternative approach would be to allow the consumer to score the attributes or cues at a different moment (but for an identical product) to the moment at which the quality is scored.

Statistical techniques

Two techniques, which allow the influence of individual cues and attributes on assigned quality to be deduced indirectly are conjoint analysis and preference mapping. Both techniques were not originally intended for developing quality models but can be used for that purpose.

Conjoint analysis was developed to determine price elasticity. In conjoint analysis the respondent is presented with written descriptions of a set of products and is asked to rank or score them in order of preference (Johnson, 1974). The products are described according to a limited number of attributes with each attribute taking one of a fixed number of values according to a factorial experimental design. The respondent does not rank or assign scores to the individual attributes but rather to the (described) product as a whole. By relating the rankings or scores for the products to the attributes varied in the experimental design (for example, using regression) it is possible to estimate the effect of each attribute on preference. Conjoint analysis has been used to build quality functions for ham (Steenkamp, 1987) and flower bulbs (Wilkinson *et al.*, 1994; Wilkinson and Polderdijk, 1996).

The aim of preference mapping is to model the relationship between sensory profiles of products and consumer preferences. Consumers are presented with a set of physical products. They judge each product and assign it a score according to their preference. At the same time a trained analytical sensory panel evaluates the values of a set of sensory attributes for the same set of products to provide a sensory profile for each product.

Consumer preferences are related to sensory profiles using linear or nonlinear regression models, which allow an estimation of which sensory profiles are preferred. A strong point of preference mapping is that the preferences of each individual consumer are modelled separately. This leads naturally to a recognition of the subjective nature of preference and to segmentation, clustering of groups of consumers with similar preferences. While preference mapping is designed to elicit information about the influence of sensory attributes, the technique could in theory be extended to include the influence of other categories of attributes.

Summary – quality assignment models

To summarise, a number of approaches have been described for developing quality assignment models. The consumer research approach is the one which keeps closest to the description of the process of quality assignment. However, it could be improved by incorporating features from other approaches. These include the definition of primary critical attributes and the decomposition of attributes into categories with differing importance, as proposed by Molnár, the indirect measurement of attribute and cue appreciation as used in conjoint analysis and preference mapping, and the modelling of quality assignment at the individual consumer level, analogous to the preference mapping approach.

17.2.3 Dynamic product model

The dynamic product model aims to describe the relationship between external factors and product quality attributes or cues. It is possible to decompose the model still further to model the relationship between external factors and product properties separately, as well as the relationship between the product properties and the perceived quality attributes or cues. For example, one model could describe the relationship between storage or processing variables and the sugar content of French fries, and a second model could describe the relationship between the sugar content and the colour. In fact this is rarely done, with most authors modelling the quality attribute or cue directly (though see Hertog *et al.*, 1997).

The most common approach is to build a model for a single attribute or cue. Often these are primary critical attributes, according to the definition of Molnár, such as spoilage. Models are frequently developed to predict keeping quality as determined by spoilage. As discussed earlier, keeping quality is in fact the time until quality falls below an acceptable level, the quality limit. An example is the model described in Hertog *et al.* (1999) which describes spoilage of strawberries due to Botrytis infection as a function of temperature and gas conditions. In Zwietering (1993) and Zwietering and Rombouts (1994) kinetic models are discussed which describe bacterial growth and which can be used to predict keeping quality.

In addition to models for primary critical attributes, models have also been developed for other attributes. The choice of attribute is often influenced by the priorities of the producer rather than the consumer. Marcelis and Gijzen (1998) developed production (growth) models which predict the fresh weight of cucumber fruits. They consider this to be an important quality attribute because it determines market price. This reflects an outlook on quality from the point of view of the producer and distributor rather than the consumer. Such an outlook also leads to an emphasis on quality cues, which determine expected quality at the point of sale rather than experienced quality attributes. As an example, Vankerschaver *et al.* (1996) developed a model for the visual quality of cut endive but not for its taste.

A major application of dynamic product models is in simulation studies. They can show the effect of changes in storage and processing conditions on key quality attributes and can thus be used to optimise logistic chains or food processing. For example, Hertog *et al.* (1999) used simulations to gain insight into the effect of logistic chains and packaging on spoilage of strawberries. For true optimisation, dynamic product models need to be combined with quality assignment models. This is not yet done. However, in simulations, some direct information about assigned quality is obtained, as can be seen when the quality of a product is no longer acceptable.

17.2.4 Integration of quality assignment models and dynamic product models

While most modelling takes place within either the quality assignment sub-task or the dynamic product sub-task, the quality function deployment and the quality

guidance frameworks do overstep this boundary (Viaene and Januszewska, 1999; Steenkamp and van Trijp, 1996; Poulsen *et al.*, 1996; Bech, 2000). Both frameworks comprise a set of models, which relate consumer quality preferences via quality cues and attributes to technical product specifications. However, they are more than just a collection of models as they embody a philosophy in which product development is guided by the voice of the consumer. They are tools to improve the collaboration between marketing and R&D departments in the process of new product development.

The quality function deployment framework defines a series of research steps to carry out this goal. Firstly, potential market segments are defined and consumer demands are identified. Secondly, consumer preferences are translated into technical and sensory product specifications. These relationships are summarised visually in a set of pictorial matrices which collectively form the so-called 'House of quality' (Bech, 2000). This can be used to predict consumer perceptions of existing own and competing products and new concepts. The predictions can be validated using the results of consumer research. An example of the use of quality function deployment for the chocolate industry can be found in Viaene and Januszewska (1999).

The quality guidance framework extends the consumer research approach to modelling quality assignment (see above) with a step relating perceived quality attributes and cues to product properties. These relationships are inherently multivariate. One quality attribute may be a function of several product properties. For example, the meat quality attribute 'appearance' may be influenced by colour, amount of fat and moisture. Equally one product property may affect several quality attributes. Pâté coarseness may contribute positively to the attribute 'taste' but negatively to the attribute 'leanness' (Steenkamp and van Trijp, 1996). To describe these multivariate relationships, Partial Least Squares regression (Steenkamp and van Trijp, 1996) or LISREL (Poulsen *et al.*, 1996) can be used.

17.3 Areas of application

The two main areas of application for food quality models are in quality control and product development.

17.3.1 Quality control

In quality control an already existing product forms the basis. The aim of the producer is to ensure a constant and high quality at all times. This is achieved by monitoring the product during production and distribution and taking the necessary actions to maintain its quality. Because quality control does not have the intention of altering the existing product, quality is primarily interpreted as meaning conformance to technical specifications and avoidance of defects. This does not necessarily mean that the subjective nature of quality is ignored. Thus

in quality control of rice it is recognised that quality indices for rice intended for the Japanese market will differ to that for the American market due to differences in taste preferences and use of rice in meals – Japanese prefer sticky rice served plain; Americans prefer non-sticky rice served with a sauce (Barton *et al.*, 1998). However, it can mean that quality control tends to emphasise the avoidance of negative attributes and cues (for example, external defects) rather than maximising positive attributes and cues (for example, a good taste). Also technical specifications have a lack of flexibility which means that they do not always reflect the sometimes rapid changes in consumer preferences.

Food quality modelling can be implemented for three different purposes in quality control.

- *Prediction of key quality attributes using instrumental measurements*. Often at-line, in-line or on-line instrumental measurements are carried out routinely at production and distribution facilities. These measurements of product properties are then related to key quality attributes/cues using calibration models. Examples are instrumental measurements of cucumber colour, of pea mealiness, of the fat content of French fries and of rice protein content. There is an increasing use of non-invasive spectroscopic and other multi-parameter instruments, indicative of an increasing recognition that a given quality attribute or cue is generally a function of several product properties. The application of instrumental measurements would benefit from quality assignment models. These could be used to direct attention to those quality aspects which are truly found to be important by consumers rather than aspects which producers think consumers find important. Also the use of quality assignment models can help avoid excessive emphasis on quality cues rather than quality attributes. Moreover they could be used to determine which values of quality attributes or cues are appreciated, and by which groups (segments) of consumers. This would ultimately enable production tailored to the quality demands of specific segments in the market.
- *Prediction of effect of raw material properties on final product quality*. For producers of processed food there is a need to determine and control the effect of raw material characteristics on final product quality. With this information they can reward suppliers by paying more for 'high quality' raw materials. This is done for example in the French fries industry where the amount paid out to growers for potatoes is partly determined by the fry colour. Food producers can choose to vary their production process to compensate for variations in raw material characteristics and ensure a constant output quality. For some products (for example juice, port and blended whisky), producers can use these prediction models to optimally combine different batches of raw materials in order to obtain a constant and high quality. In most current practical applications, prediction models relating raw materials to final product quality are empirical, based on observed correlations under a given processing condition. Such prediction models gain in usefulness when they are based on dynamic product models,

which describe how processing conditions change the raw product to obtain a certain set of quality attributes (see Van Dijk and Tijskens, 2000).

- *Calculation of quality indices*. A quality index can be seen as the practical application of a quality assignment model. It provides a well-understood measure of the quality of a product which is recognised by all members of the production and trade, and allows products to be compared in an objective manner. An example of this kind of practical application of simplified quality assignment models is the Streif index for apple maturity at harvest. Developed in the early seventies (Streif, 1976), the index is defined as the ratio between the firmness of apples and the Brix refraction times the starch stage. It has obtained an increasing application to determine the harvest date for apples, optimal for subsequent storage and eating quality. It is nowadays so widely used that complete conferences (De Jager *et al.*, 1996) are devoted to its application, improvement and its calibration against sensory and expert data.

17.3.2 Product development

In new product development both quality assignment models and dynamic product models can be applied in the search for a product with the desired quality. The process of product development starts with the consumer – 'what does he want?', moves to the product – 'how do I make it?' and then back to the consumer – 'does he like what I made?'. Several iterations may be necessary. In the process, food quality modelling can be involved at all three stages.

- In the first stage, quality assignment models can be applied to determine the quality preferences of consumers. The quality assignment models can show whether there is segmentation in quality perception, which attributes (and at which levels) determine quality perception and whether available products are currently meeting quality demands.
- In the second stage, R&D is presented with a set of desired quality attributes and cues. At this stage dynamic product models can be applied to determine what modifications in processing conditions or ingredients are required to achieve the desired attribute profile.
- In the third stage, traditionally a consumer acceptance test is carried out on one or two products. However, it is possible to use the quality assignment models developed in the first stage to predict the assigned quality for a greater range of potential products as a first screening.

The use of food quality modelling for new product development as described here is largely implemented in the quality function deployment and quality guidance frameworks discussed in the previous section. However, these frameworks give limited input for product design and optimisation (stage 2) as they stop at the translation of quality demands into product properties. These frameworks do not describe how these product properties should be modified. To achieve this, fundamental knowledge has to be generated and incorporated (see Chapters 2 and 3).

17.4 Pros and cons and future trends

An advantage of food quality modelling is that it makes explicit the subjective nature of assigned quality, implying that the same product can have different perceived qualities for different people in different situations. As yet this aspect of quality modelling is not well incorporated in the methodology with most quality assignment models providing one mathematical function for all consumers, or for only one defined subgroup. The future will see more attention to the question of segmentation, in which quality assignment models are built for defined groups of consumers with similar quality demands and defined categories of situations. This will enable companies to tailor their products to the quality demands of defined user groups.

Variations not only between consumers but also between individual product units within a batch – the so-called biological variance – will also receive more attention in the future (Tijskens *et al.*, 2000). Regarding the quality assignment model, the perceived quality of a batch obviously depends on the perceived quality of the individual units, but the nature of this relationship is not so obvious. Is it a simple average, or do poor quality units have extra weight? More research is needed on this perception question. Regarding the dynamic product model, there are some interesting developments in modelling batch behaviour with a limited number of unit-dependent parameters (Nicolaï *et al.*, 1995; Nicolaï and Van Impe, 1996; Tijskens and Wilkinson, 1996; Tijskens, 2000; see also Chapter 8).

Quality assignment models are inherently very broad, covering in their fullest form intrinsic and extrinsic attributes as well as cues. This is both an advantage and a potential problem. The advantage lies in the healthy counterbalance it offers to the tendency in the past to reduce quality to one or two easily measured aspects. Thus producers and distributors in the fruit and vegetable sector have in the past tended to concentrate on a limited number of mostly external attributes which could easily be measured instrumentally or by product experts (for example, colour and firmness, but not taste). The broad range of the quality assignment models, on the other hand, presents big challenges for the consumer research methodology necessary to quantify the models. The mechanisms underlying the perception and appreciation of experience attributes such as taste are very different from those for credence attributes such as calcium content or extrinsic cues such as brand name. Sensory perception and appreciation is largely determined by the physical morphology, a result of genetic makeup and age. Belief in the importance of calcium is a cognitive attitude and much more susceptible to influence.

The extensive quality assignment models contrast too with the dynamic product models. While the quality assignment models are highly multivariate, relating many cues and attributes to several scores of quality, and usually assume simple linear relationships, the dynamic product models are predominantly univariate, but are increasingly treated in a multi-response approach (Chapter 3). They model single quality attributes using nonlinear models

describing kinetic or other chemical or physical processes. Perhaps the biggest challenge for the future is the combination of these two different styles of modelling to produce workable models which can predict quality given processing and other conditions, or alternatively which can deduce optimal processing conditions given quality demands.

17.5 References

ACEBRÓN L B, DOPICO D C, (2000). The importance of intrinsic and extrinsic cues to expected and experienced quality: an empirical application for beef. *Food Quality and Preference*, **11**, 229–38.

ANDANI Z, MACFIE H J H, (2000). Consumer preference. In Shewfelt R L, Brückner B (eds), *Fruit and Vegetable Quality: An Integrated View*, 158–77. Technomic Press, USA.

BARTON F E, WINDHAM W R, CHAMPAGNE E T, LYON B G, (1998). Optimal geometrics for the development of rice quality spectroscopic chemometric models. *Cereal Chemistry*, **75**, 3, 315–19.

BECH A, (2000). House of quality. In Shewfelt R L, Brückner B (eds): *Fruit and Vegetable Quality: An Integrated View*, 199–222. Technomic Press, USA.

DE JAGER A, JOHNSON D, HÖHN E, (eds.) (1996). *The postharvest treatment of fruit and vegetables. Determination and prediction of optimum harvest date of apples and pears*. Proceedings COST 94 conference, 164 pp., Lofthus, Norway, 9–10 June 1994.

HERTOG M L A T M, TIJSKENS L M M, HAK P S, (1997). The effects of temperature and senescence on the accumulation of reducing sugars during storage of potato (*Solanum tuberosum* L.) tubers: A mathematical model. *Postharvest Biology and Technology*, **10**, 67–79.

HERTOG M L A T M, BOERRIGTER H A M, VAN DEN BOOGAARD G J P M, TIJSKENS L M M, VAN SCHAIK A C R, (1999). Predicting keeping quality of strawberries (cv. 'Elsanta') packed under modified atmospheres: an integrated model approach. *Postharvest Biology and Technology*, **15**, 1–12.

JOHNSON R M, (1974). Trade-off analysis of consumer values. *Journal of Marketing Research*, **XI**, 121–7.

JURAN J M, (1974). Basic concepts. In Juran J M, Groyna F M Jr, Bingham R S Jr, (eds), *Quality Control Handbook*, 3rd edn, McGraw-Hill, New York.

KRAMER A, TWIGG B A, (1970). *Quality Control for the Food Industry*, AVI Pub. Co, Westport.

LAWLESS H T, HEYNMANN H, (1998). *Sensory Evaluation of Food: Principles and Practices*. Chapman & Hall, New York, USA.

MARCELIS L F M, GIJZEN H, (1998). Evaluation under commercial conditions of a model of prediction of the yield and quality of cucumber fruits. *Scientia Horticulturae*, **76**, 171–81.

MOLNÁR P J, (1995). A model for overall description of food quality. *Food Quality and Preference*, **6**, 185–90.

NICOLAÏ B M, VAN IMPE J F, (1996) Predictive food microbiology: a probabilistic approach. *Mathematics and Computer Simulation*, **42**, 287–92.

NICOLAÏ B M, VAN IMPE J F, MARTENS T, DE BAERDEMAEKER J, (1995). A probabilistic model for microbial growth during cold storage of foods. *Proceedings 19th International Congress of Refrigeration*, **I**, 232–9, August, The Hague, the Netherlands.

POULSEN C S, JUHL H J, KRISTENSEN K, BECH A C, ENGELUND E, (1996). Quality guidance and quality formation. *Food Quality and Preference*, **7**, 2, 127–35.

SLOOF M, (1999). *Physiology of Quality Change Modelling*. Dissertation, Free University of Amsterdam, No 99–1. 21–9.

SLOOF M, TIJSKENS L M M, WILKINSON E C, (1996). Concepts for modelling the quality of perishable products. *Trends in Food Science and Technology*, **7**, 165–71.

STEENKAMP J-B E M, (1987). Perceived quality of food products and its relationship to consumer preferences: theory and measurement. *Journal of Food Quality*, **9**, 373–86.

STEENKAMP J-B E M, (1989). Product quality: an investigation into the concept and how it is perceived by consumers. Ph.D, Agricultural University, Wageningen, The Netherlands.

STEENKAMP J-B E M, (1990). Conceptual Model of the Quality Perception Process, *Journal of Business Research* **21**, 309–33.

STEENKAMP J-B E M, VAN TRIJP H C M, (1996). Quality guidance: a consumer-based approach to food quality improvement using partial least squares. *Economic Review of Agricultural Economics*, **23**, 2, 195–215.

STREIF J, (1976). Einfluss der Temperatur auf verschiedene Reifemerkmale von Äpfeln (Influence of temperature on various maturity characteristics of apples.) *Erwerbsobstbau*, **18**, 168–71.

TIJSKENS L M M, (2000). Acceptability. In Shewfelt R L, Brückner B (eds): *Fruit and Vegetable Quality: An Integrated View*, 125–143, Technomic Press, USA.

TIJSKENS L M M, SLOOF M, WILKINSON E C, (1994). Quality of perishable produce. A philosophical approach. *Proceedings of the Sixth International Symposium of the European Concerted Action Program COST94 The Post-Harvest Treatment of Fruit and Vegetables – Current Status and Future Prospects*, 19–22 October 1994, Oosterbeek, The Netherlands, 493–502.

TIJSKENS L M M, POLDERDIJK J J, (1996). A generic model for keeping quality of vegetable produce during storage and distribution. *Agric. Systems*, **51**, 431–52.

TIJSKENS L M M, WILKINSON E C, (1996). Behaviour of biological variability in batches during post-harvest storage. *AAB Modelling 'Confer'ence: Modelling in Applied Biology – Spatial Aspects*, 267–68, 24–27 June, Brunel University, Uxbridge, UK.

TIJSKENS L M M, KONOPACKI P, SIMČIČ M, HRIBAR J, (2000). Biological variance

in agricultural products. Colour of apples as an example. Proceedings IIR Conference: *Improving Postharvest Technologies of Fruits, Vegetables and Ornamentals*, 19–21 October 2000, Murcia, Spain (in press).

VAN DIJK C, TIJSKENS L M M, (2000). Mathematical modeling of enzymatic reactions as related to the texture of fruits and vegetables after storage and mild preheat treatments. In Alzamora S M, Tapia S M, López-Malo A (eds.) *Design of Minimal Processing Technologies for Fruit and Vegetables*, Aspen Publishers Inc., USA, p. 127–52.

VANKERSCHAVER K, WILLOCX F, SMOUT C, HENDRICKX M, TOBBACK P, (1996). Modeling and prediction of visual shelf life of minimally processed endive. *Journal of Food Science*, **61**, 1094–8.

VIAENE J, JANUSZEWSKA R, (1999). Quality function deployment in the chocolate industry. *Food Quality and Preference*, **10**, 377–85.

WILKINSON E C, POLDERDIJK J J, SLOOF M, LOKERS R M, (1994). Q-Bulb. Proceedings VIAS Symposium, Ede, The Netherlands, 35–42.

WILKINSON E C, POLDERDIJK J J, (1996) Modelling quality perception in the distribution chain: a case study for tulip bulbs. Proceedings 2nd International Conference on *Chain Management in Agri- and Food Business*, Ede, The Netherlands, 571–7.

ZWIETERING M, (1993). Modeling of microbial quality of food. PhD Thesis, Wageningen University. ISBN: 90-5485-143-0.

ZWIETERING M H, ROMBOUTS F M, (1994). Modelling the microbiological quality of food products (in Dutch). *Voedings Middelen Technologie*, **19**, 19–22.

18

Modelling microbiological safety

J. Baranyi, Institute of Food Research, Norwich and
C. Pin, Universidad Complutense de Madrid

18.1 Introduction

The increasing demand for mathematical modelling in the interest of microbiological safety of food began with the awareness that quality control of foods based on inspection of the final product was costly, laborious and inefficient. Models are quick and economical ways to assess food safety objectively. Since the 1980s, the quantitative approach to microbial ecology in food has been studied intensively. The number of related publications has been increasing exponentially since then. The name 'predictive food microbiology' was coined in the 1930s but practical applications began to materialise only in the 1980s, with powerful desktop computing in everyday use.

Mathematical modelling is a main tool behind this development. In some ways, mathematical modelling is an art of omitting the unnecessary. The starting point of a modelling procedure is the process of abstraction where we disregard those features of the modelled phenomenon that are not important from a certain point of view. An example of this is the theory of constant *D*-value of bacterial inactivation. According to this, if the heating temperature is constant, a homogeneous cell population dies as described by first order kinetics, with constant specific death rate, *k*. This model is obtained after omitting unnecessary details, and serves as a purified version of reality, in that the population is never guaranteed to be completely homogeneous, the temperature has, in fact, a distribution in the heating menstrum, etc. In spite of this, the canning industry has been using constant *D*-values ($= \ln(10)/k$) for a long time, with satisfactory results, first based on purely empirical observations, later confirmed by mechanistic thinking.

Though modelling bacterial inactivation and survival is probably the most important topic in mathematical modelling of microbial safety, we give some

examples of mathematical techniques to study bacterial growth in this chapter. The reason for this is that bacterial inactivation in the food environment has been more thoroughly analysed in recent decades compared to the relatively new area of modelling bacterial growth. Besides, in the interest of minimally processed, safe foods, growth studies gain more and more significance. This chapter will review some results related to growth modelling.

18.2 Developing mathematical models

A mathematical model is a set of assumptions, some of which can be formulated by equations describing mathematical relations between the introduced variables. In biology, the real system is extremely complex, so its model must inevitably include simplifying idealisations. These occur, for example, when one variable measures a feature that is, in fact, a composite of several others, or when a variable is considered to be constant in time and/or homogeneous in space although it is known to be time-dependent and heterogeneous in reality. The extent of this neglect depends on theoretical and practical considerations such as the available mathematical techniques, computing power and data.

18.2.1 Empirical and mechanistic models

Mathematical models are frequently classified as mechanistic and empirical models. Empirical models are only expected to describe accurately a set of observations, without taking into account the intrinsic mechanism by which these data are generated. A mechanistic model describes rather the process, either directly observable or unobservable, that generates those data. In practice, purely mechanistic models are rare, rather a mixture of the two is applied, possibly closer to one than to the other. Examples of models involving more empirical than mechanistic elements are those used in predictive microbiology, frequently aiming at the pure collection and smoothed (i.e. 'noiseless') representation of computerised microbial data. However, with the increase of these data, more and more experience is accumulated and certain qualitative features become 'compulsory' for the models to be created.

Frequently, model development is an iterative process going through a 'learning curve', when initial, empirical models (describing observations purely quantitatively) can help to define certain qualitative features of a more mechanistic model to be developed. It is also desirable to embed the model into more general principles of science and to make it open to further developments as the quantity and quality of information about the system increases.

For example, consider the development of bacterial growth models. It was observed by Malthus (1798) that, if there is no inhibition, a homogeneous population grows according to an exponential function, which is a straight line if the size of the population is plotted in log scale. This exponential model for growth was, initially, a purely empirical model. Later, mathematical descriptions of more

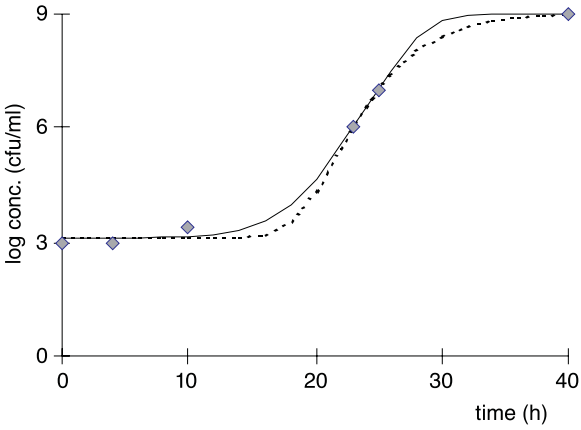


Fig. 18.1 Bacterial growth curve fitted by the Gompertz function (broken line) and by the model of Baranyi and Roberts (1995). The goodness of fit is similar but the latter model 'forces' a straight line fitting in the exponential phase where the Gompertz function has a pronounced curvature, thus overestimating the rate there.

general principles of nature confirmed this law (for a summary, see Renshaw, 1991). Today this law should be a basic requirement, a starting point, for any mathematical growth model. In other words, if the physical environment is favourable and constant with time, then the log of the population size against time should be represented by a straight line (at least in the 'happy growth region', i.e. over an adjustment period and far from overpopulation, etc; see Fig. 18.1).

This is today a basic requirement for any reasonable mechanistic model of bacterial growth (Baranyi and Roberts, 1995). There are, however, other, well-fitting, useful empirical models of bacterial curves (Gompertz, logistic, etc., see Zwietering *et al.*, 1990), which do not have this mechanistic background. The problem with those empirical models is that they are not applicable even for a slight extrapolation. For example, starting from a Gompertz function *for the logarithm of the bacterial size* (which has nothing to do with the Gompertz model for the *size* of the population), it is difficult to create a consistent mathematical model for a changing environment. Models for non-isothermal situations, in some sense, require a certain extrapolation from isotherm models. As a mechanistic principle, the isotherm situation should only be a special case of the dynamically changing temperature environment (Fig. 18.2). This requirement is satisfied, for example, by the dynamic model of Baranyi and Roberts (1995). That model is a system of differential equations whose explicit solution is guaranteed only for constant temperature profile, giving the well-known sigmoid curve for the bacterial cell concentration. Under changing temperature, the model is the same but the solution can be obtained generally by numerical methods only (step-by-step iteration on the computer).

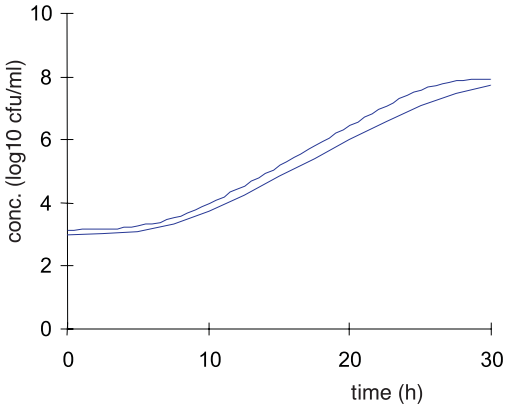


Fig. 18.2 If a model applied to isotherm environment is only a special case of the model describing the dynamically changing environment situation, then small temperature changes have a small effect on the bacterial response. The lower growth curve is a prediction obtained by the model of Baranyi and Roberts (1995) for 15ºC. The upper curve is predicted (by the same model) for the situation when the temperature increases at a constant rate of 0.1ºC/hour, from 15ºC. For this prediction, a numerical integration was necessary because of the dynamically changing environment.

18.2.2 A dynamic growth model

First, we define some notations we use frequently in this chapter. A vector will be denoted by bold: $\mathbf{v} = [v_1 \ldots v_n]$ means that the vector $\mathbf{v}$ has n entries, v_1, v_2, …, v_n. These entries can be time-(t-)dependent variables, too, in which case $\mathbf{v}(t) = [v_1 \ldots v_n(t)]$ is a time-dependent vector variable.

We will denote the time and the bacterial concentration by t and x, respectively. The $x(t)$ function is the main centre of our investigation. We consider this as a response to the environment, which we assume to be satisfactorily characterised by some main factors, $E_1(t) \ldots E_n(t)$. As can be seen, both the environment and response is meant in dynamic sense, allowing them to change with time. In the main, the subject of predictive microbiology is the mapping *Environment* $\rightarrow$ *Bacterial response*, whose most important element is the sub-mapping $\mathbf{E}(t) \rightarrow x(t)$.

18.2.3 Exponential growth

Discussing growth models commonly start with the exponential model. According to that, the specific increase or decrease, $\Delta x/x$, of the population under Δt interval is proportional to the length of the interval: $\Delta x/x = \mu \Delta t$. The proportionality factor μ is called the specific rate and it is positive for growth, negative for death.

From the above equation, the following differential equation can be obtained

$$\dot{x} = \mu x \tag{18.1}$$

where the $(\dot{})$ represents the *d/dt* operation. If and only if μ is constant, then this is equivalent to

$$x(t) = x_0 e^{\mu t} \tag{18.2}$$

where x_0 is the value of the bacterial concentration at the time $t = 0$.

It is worth considering the natural logarithm of the bacterial population:

$$y(t) = \ln x(t) \tag{18.3}$$

by which, equation (18.1) can be written as

$$\dot{y} = \mu \tag{18.4}$$

18.2.4 Sigmoid bacterial curves

Especially in food microbiology, where the bacterial concentration of interest spans from 1 up to 10^9–10^{12} cells/ml, it is important to study the entire bacterial growth curve which is commonly of sigmoid shape.

Turner *et al.* (1976) published a comprehensive study on sigmoid growth models where the size of the population follows a sigmoid pattern with time. In our situation, however, the *log-population* follows a sigmoid pattern with time and using those classical sigmoid models for the log-population would be a completely empirical approach.

In accordance with the principles mentioned in section 18.1, we start from the idealistic model described by equation (18.4). We expect that the higher the cell concentration, the smaller the specific rate. Therefore we develop the model further in such a way that it can contain the original model as a special case. We write

$$\dot{y} = \mu\, u(y) \tag{18.5}$$

with some $u(y)$ 'inhibition' function, which is between 0 and 1 and decreasing to 0 monotonically as y increases. All the well-known sigmoid growth models (Gompertz, logistic, Richards, etc.) can be re-created for $x(t)$ if $u(y)$ is chosen appropriately (see Baranyi and Roberts, 1995). The most frequently used inhibition function can be derived from the model of Richards (1959)

$$u(y) = 1 - e^{-m(y_{max} - y)} \tag{18.6}$$

where y_{max} is the natural logarithm of the maximum population density and m is a curvature parameter characterising the transition to the stationary phase. With $m = 1$, the logistic model of population growth can be obtained for $y = \ln x$.

The model described by equation (18.5) results in a bi-phasic function for $y(t)$, the shape of which goes through a transition from an increasing straight line

to a horizontal one. Microbial growth curves, however, generally go through an initial adjustment period called lag phase, lending a sigmoid shape to the growth curve. The end of the lag period, denoted by λ in what follows (see Fig. 18.4), is defined as the intercept of the initial level and the tangent drawn to the inflexion of the sigmoid curve (McMeekin *et al.*, 1993). It is easy to see (Coleman, 1978), that equation (18.5) is not suitable to describe bacterial lag, because $y(t)$ cannot have inflexion for any reasonable inhibition function. This is why we complete the model with a so-called 'adjustment function'

$$\dot{y} = \alpha(t)\mu\, u(y) \tag{18.7}$$

where $\alpha(t)$ is a monotone increasing function with values between 0 and 1. Several forms of $\alpha(t)$ can be defined; a useful one is a Michaelis-Menten type inhibition (Baranyi and Roberts, 1995):

$$\alpha(t) = P(t)/(K_P + P(t)) \tag{18.8}$$

where $P(t)$ is a critical substance necessary to initiate growth (such as a new enzyme, if the substrate has changed), and K_P is the Michaelis-Menten saturation constant. The critical substance is assumed to increase at a specific rate ν, from an initial value P_0:

$$\dot{P} = \nu P \tag{18.9a}$$

$$P(0) = P_0 \tag{18.9b}$$

Partly for numerical stability reasons, partly because the $P(t)$ substance is unknown and the solution of the system depends only on the P_0/K_P ratio, it is worth introducing the $q(t) = P(t)/K_P$ variable, by which we obtain the equations

$$\dot{q} = \nu q \tag{18.10a}$$

$$q(0) = q_0 \tag{18.10b}$$

where the q_0 parameter depends on the history of the cells (how the initial quantity of the critical substrate, P_0 relates to K_P); $\nu(t)$ characterises the rate of the adjustment to the new (actual) environment. The simplification $\nu = \mu$ has some mechanistic background and makes the model more suitable for practical curve fitting procedures, too.

The model described by equations (18.6)–(18.10) has the advantage that it has an algebraic solution if the parameters are constant with time. In dynamic situations, however, it should be solved by numerical methods.

The following reparameterisations of q_0 have biological interpretations and advantageous numerical/statistical properties that are useful when using the model for curve fitting:

$$\alpha_0 = q_0/(1 + q_0) \tag{18.11}$$

$$h_0 = -\ln(\alpha_0) \tag{18.12}$$

18.3 Modelling the effect of environmental factors on the growth parameters

To study the effect of the environmental factors on bacterial growth, four aspects should be considered:

1. What growth parameters are to be modelled as a function of the environment?
2. What are the main environmental factors characterising the environment?
3. What model is to be used to describe that effect?
4. What is the environmental region where predictions can be obtained by interpolation?

In predictive food microbiology, the temperature, pH, and water activity are considered as the main environmental variables determining bacterial growth in food. By doing this, such factors as food structure, the composition of the atmosphere, or possible preservatives, were considered only as factors modifying the basic model but not changing their main structure.

18.3.1 Choosing the growth parameters to be modelled

As has been mentioned, the subject of predictive microbiology is the mapping *Environment* → *Bacterial response*. A variable or parameter of the mathematical model describing this system is intrinsic, if it does not depend on factors outside the system. Extrinsic variables or parameters are those that depend only on factors that are outside the system. Of course, variables and parameters can be defined in such a way, too, that they depend on both intrinsic and extrinsic factors.

Traditionally, the growth of a bacterial population is defined by four parameters: maximum specific growth rate, μ; duration of lag, λ; maximum population concentration, $c_{max} = \exp(y_{max})$; and initial cell concentration, $x_0 = \exp(y_0)$.

It is a principal assumption in food microbiology that, in a constant environment, a homogeneous cell population eventually grows at the same maximum specific growth rate, characterising the species and the environment in question. Hence, μ is a reproducible, intrinsic parameter. So is c_{max}. These two parameters should be modelled without information on the history of the cells. The initial log-concentration, however, depends only on the history of the environment (such as a food treatment), so that is an extrinsic parameter. As for the lag phase, it depends on both the history and the actual environment, so it is neither an intrinsic nor an extrinsic parameter. Baranyi and Roberts (1994) suggested that a parameter called h_0 should be considered instead, from which the lag can be calculated as $\lambda = h_0/\nu$. Robinson *et al.* (1998) pointed out that h_0 quantifies the 'work' required by the cells to prepare for the exponential growth.

From traditional growth data, neither the amount, h_0, nor the rate of 'work to be done', ν, can be measured directly. However, it is a logical assumption that

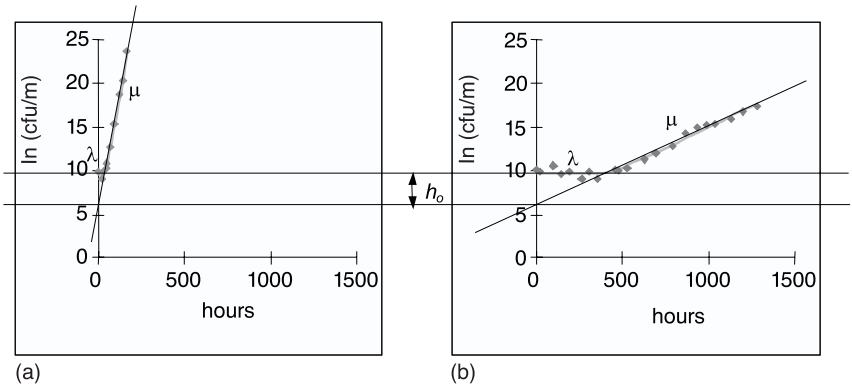


Fig. 18.3 Growth of *L. monocytogenes* at (a) 11ºC, 80% CO_2 and at (b) 1.5ºC, 80% CO_2. The h_0 parameter is fairly independent of the actual environment.

$\mu_1 = 0.109\,(1/h)$ $\lambda_1 = 42.1\,(h)$, therefore $h_0 = \mu_1\lambda_1 = 4.59$ in case (a);
$\mu_2 = 0.010\,(1/h)$ $\lambda_2 = 448\,(h)$, therefore $h_0 = \mu_2\lambda_2 = 4.48$ in case (b).

the subsequent maximum specific growth rate of the population (μ) reflects the rate at which metabolic reactions occur, determined by the environment. Therefore, it can be assumed that $\nu = \mu$. The 'work to be done' before the exponential phase is zero when the ability of the initial cell population to grow is 100% ($h_0 = 0$, so $\alpha_0 = 1$ and there is no lag).

When standardised inoculum is used (inoculation from cultures with identical history) to generate growth curves in different environmental conditions, h_0 reflects the effect of the environmental change on the bacterial strategy to grow. Robinson *et al.* (1998) pointed out that there is a consistent linear relationship between μ_{max} and $1/\lambda$ at different temperatures. This means that, with identical history, h_0 is independent of the actual growth temperature. In Fig. 18.3, it can be seen that the 'work to be done' was the same at two different growth temperatures (1.5ºC and 11ºC) for *Listeria monocytogenes*. The work was carried out at a slower rate at 1.5ºC and, in consequence, the lag phase was longer. Another way of putting this observation in words is that the lag/generation time ratio is constant and characterising the history of the cells.

On the other hand, for a particular environment, h_0 can be used as an indicator of the heterogeneity of the inoculum, related to different physiological states, for example injured or healthy cells, and/or different previous growth conditions. Figure 18.4 demonstrates the different work carried out during the lag phase by *L. monocytogenes* at 5ºC, depending on the previous growth conditions. The cells previously subcultured at 25ºC needed less work to adapt to the new environment (5ºC) than the cells subcultured at 37ºC. These imply that, instead of the effect of the actual environment on the lag, it is the history effect (initial conditions of the actual environment included) on the h_0 parameter that should be modelled. The lag can be derived from the h_0 parameter by $\lambda = h_0/\mu$. If the environment changes with time, it is difficult to interpret a

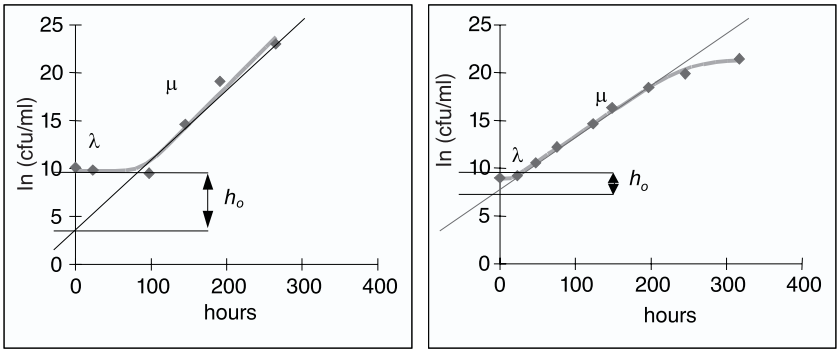


Fig. 18.4 Growth of *L. monocytogenes* at 5ºC, (a) using an inoculum previously subcultured three times at 37ºC; and (b) using an inoculum subcultured three times at 25ºC. The h_0 parameter (work to be done during the lag phase) depends on the history of the cells.

$\mu_1 = 0.074\,(1/h)$ $\lambda_1 = 92.3\,(h)$, therefore $h_0 = \mu_1\lambda_1 = 6.83$ in case (a);
$\mu_2 = 0.063\,(1/h)$ $\lambda_2 = 15.2\,(h)$, therefore $h_0 = \mu_2\lambda_2 = 0.958$ in case (b).

specific lag parameter having only potential meaning; but the h_0 parameter can be interpreted in that case, too.

The effect of the actual environment should be studied only on those growth parameters that are purely intrinsic. Of the intrinsic parameters, the maximum population density is frequently modelled by a constant only, since growth conditions do not affect this parameter very much; besides, accurate models for high bacterial concentrations in food are rarely necessary. Therefore, the maximum specific growth rate (or its variants, such as its logarithm, or the generation time) is usually the modelled parameter.

18.3.2 Models describing the effect of the environment on the maximum specific growth rate

The most frequently used models can be divided into three groups: Arrhenius-based models, square root or Belehrádek type models and polynomial models. The first two have some mechanistic elements inasmuch as the temperature is considered as the only environmental variable. The polynomial approach is purely empirical, but can be used for any number of environmental variables. An example of a nonlinear Arrhenius model is that of Schoolfield *et al*. (1981) which describes the logarithm of the maximum specific growth rate as a function of the reciprocal of the absolute temperature. In the model developed by Ratkowsky *et al.* (1982, 1983), the square root of the maximum specific growth rate is described by a function of the temperature. Zwietering *et al*. (1991) compared these two models by *F*-tests and found that both models are statistically acceptable. Later, attempts were made to extend both models to include other environmental variables, too (Broughall and Brown 1984; Davey 1989; Neumeyer *et al.*, 1997a).

Another possibility to describe the effect of the environment on the maximum specific growth rate is the use of polynomials (see, for example, McClure *et al*., 1993). Although polynomials are empirical models, they have the advantage of being linear in the parameters, therefore lending themselves to easily available linear regression. Second-degree polynomials with two independent variables are called quadratic surface responses. Baranyi *et al*. (1996) have shown that the use of higher degree polynomials (e.g. cubic) improves the fitting but the model loses its robustness (small perturbations in the observations induce significant changes in the fitted model). Moreover, third or higher degree polynomials can show such unrealistic analytical properties that make them unsuitable to describe the relationships between growth rate and environmental factors.

A link function (such as logarithm or square root, as in the other two types of models) is frequently used for the maximum specific growth rate, to stabilise the variance of the observations. The role of link functions was analysed extensively by Ratkowsky *et al.* (1996).

18.3.3 Comparing the effect of environmental factors

A useful tool to study and compare the effect of the environmental factors on the growth parameters is the generalised *z*-value concept as introduced by Pin *et al*. (2000).

Consider a predictive model describing a growth parameter p as a function of a set of n environmental factors ($v_1 \ldots v_i \ldots v_n$):

$$L(p) = f(v_1 \ldots v_i \ldots v_n),$$

where L is a suitable link function (such as the logarithm function).

The first partial derivatives of the f model function, with respect to the environmental variables, are:

$$\frac{df(v_1 \ldots v_i \ldots v_n)}{dv_1} \ldots \frac{df(v_1 \ldots v_i \ldots v_n)}{dv_i} \ldots \frac{df(v_1 \ldots v_i \ldots v_n)}{dv_n}$$

Note that, depending on the model, these derivatives can also be functions of the environmental factors.

The reciprocal of a derivative expresses how much the respective variable should change to induce one unit increase in the modelled parameter, while the other variables remain fixed. Assuming that the model for that parameter is strictly monotone in the studied environmental region, then the partial derivatives differ from zero for any set of values of the variables, and the reciprocals will be always interpretable. They can be expressed as:

$$z_i = \frac{1}{\dfrac{df(v_1 \ldots v_i \ldots v_n)}{dv_i}}$$

If the link function is the $\log_2$ ($L(p) = \log_2(p)$) and the modelled parameter is the maximum specific growth rate ($p = \mu$) then z_i is that change in the ith

environmental factor (increase or decrease) that induces a two-fold increase in the value of the maximum specific growth rate, while the rest of the environmental factors keep their values. For more details, see Pin *et al.* (2000).

If the modelled growth parameter is a nonlinear function of the environmental variables, the generalised *z*-value is not constant but a function of those variables. In this case, to have a general overview on the effect of an environmental factor v_i , it can be useful to calculate a mean generalised *z*-value in the studied environmental region, R (which must be part of the interpolation region of the model):

$$Z_i(R) = \frac{1}{\dfrac{\displaystyle\int_{V(R)} \frac{df(v_1 \ldots v_i \ldots v_n}{dv_i} dv_1 \ldots dv_i \ldots dv_n}{V(R)}}$$

where $V(R)$ denotes the volume of R.

It is not always practical (if possible at all) to calculate Z_i exactly. The boundaries of the interpolation region of models with three or more explanatory variables can be too complex for exact calculations. In these cases, a Monte-Carlo approach can be used. This randomly generates a sufficient number of environmental conditions (or vectors) inside the interpolation region of the model and approximates Z_i by the average of the respective z_i values.

The main use of the Z value is to study whether the effect of one environmental factor can be made equivalent to the effect of another one. The results can be used for optimising storage environments. Another use is to determine which growth parameter is affected most by an environmental factor.

18.3.4 Bacterial growth in a changing environment

It is important to see that if μ is not constant but depends on time, $\mu = \mu(t)$, then the solution of the simple exponential model (equation 18.1), for the logarithm of the cell concentration $y = \ln x$, is NOT

$$y(t) = y_0 + \mu(t)t \tag{18.13}$$

but

$$y(t) = y_0 + \int_0^t \mu(s)ds \tag{18.14}$$

It is easy to check that equation (18.13) is really a special case of equation (18.14), when the specific rate is constant and in which case $y(t)$ against time is a straight line.

Consider the situation when the specific rate is known to be increasing with time. In Fig. 18.5, we applied the model of McClure *et al.* (1993) for the specific growth rate of *Brochothrix thermosphacta* (assuming no lag) for the case when the temperature increases linearly with time from 1 to 18ºC in 48 h. The result is

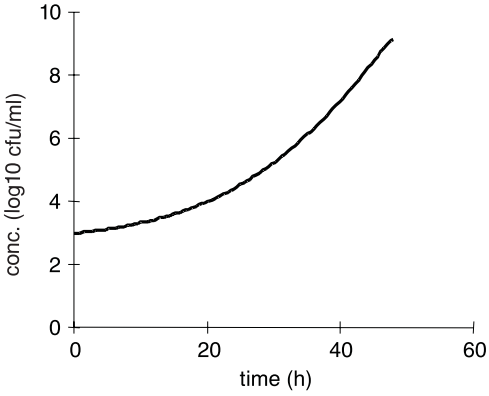


Fig. 18.5 'log$_{10}$conc. *v.* time' growth curve of *Brochothrix thermosphacta* as predicted by the model of McClure *et al.* (1993), during temperature increasing from 1 to 18ºC in 48 h. The curvature is not because of lag, but because of increasing temperature.

a dynamic specific rate, $\mu(t) = \mu(T[t])$, where $\mu(T)$ is from the model, and the temperature can be substituted as

$$T(t) = 1 + (18 - 1)/48 \cdot t$$

Applying equation (18.5), a growth curve can be obtained for the logarithm of the cell concentration (preferred in practice as $\log_{10} x(t) = y(t)/\ln 10$, where $\ln 10 \approx 2.3$). The slope of the growth curve increases with time as shown in Fig. 18.5. It is important to remember that, in this case, there is no lag, but the rising temperature causes the increase of the specific growth rate.

In this case, we described the $\mathbf{E}(t) \rightarrow x(t)$ mapping by the composition of the empirical $\mu(T)$ model of McClure *et al.* (1993) and the mechanistic model (equations 18.5–18.12), with $\mathbf{E}(t) = T(t)$. Note that, by substituting $\mu(T(t))$ in equation (18.5), we assumed that the bacterial specific rate adjusts to the actual temperature instantaneously.

Depending on the $\mu(T)$ model and $T(t)$ environment profile, the predicted curve cannot always be expressed by algebraic expressions. What is more, it is frequently difficult even to find out whether, for a complex $\mu(T)$ function, an algebraic solution exists or not. In practice, it is much more useful to concentrate on numerical (approximate) solutions of differential equation models (see for example, Press *et al.*, 1992), for which there are well-known, commercially available software packages.

18.3.5 Interpolation region

One basic principle of empirical modelling is that it should not be used to predict responses for conditions outside the region in which the observations used to fit the model were made. The nominal region of a model is the Cartesian product of

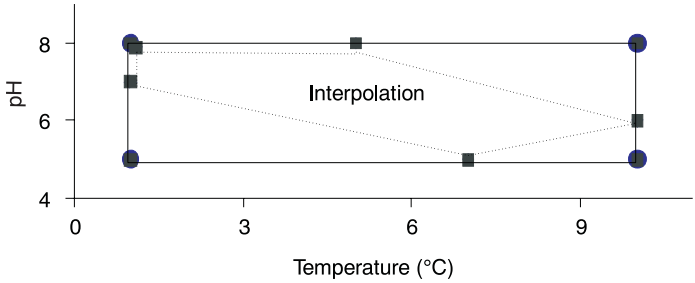


Fig. 18.6 An interpretation of the strict interpolation region by the minimum convex hull containing the observational points. (Continuous lines and circles: boundary and vertices of the nominal experimental region of the environmental factors. Broken line and squares: boundary and vertices of the minimum convex hull spanned by the points that were used to create the model.

the intervals of the different environmental factors for which observations were obtained. The strict interpolation region is often smaller than this nominal region. In multidimensional cases, it can be quite difficult to determine whether a prediction from the nominal region is also inside the interpolation region.

Baranyi *et al.* (1996) defined the strict interpolation region as the minimum convex hull containing all the observations that were used to generate the model. The interpolation region is, in fact, defined by a restrictive rule. A prediction for a combination of environmental conditions is an interpolated value if and only if the conditions can be composed as a convex linear combination of those conditions where observations were made and used to generate the model.

Figure 18.6 shows an example, in two dimensions, for the strict interpolation region inside the nominal region. In more than two dimensions, it is not easy to represent the strict interpolation region. In that case, a simplex algorithm can be applied to decide whether a test point is inside that region or not (see Baranyi *et al.*, 1996).

18.4 Model validation

The conditions under which data used to develop models are generated do not necessarily represent real food conditions exactly. For this reason, one would not expect the predictions of a model to be completely accurate. Estimating the expected differences is as important as giving the predictive value itself. Validation is, in fact, a process of estimating the expected error and comparing model predictions with observations obtained independently of the model.

The data used to develop a model are typically from bacterial cultures grown under controlled conditions. However, this scenario is much simpler than the real situation of a food product (varying composition, storage conditions, etc.).

Several studies have focused on the differences between the predictions of models and observations made in food products. The difference between model prediction and observation made under laboratory conditions is called 'primary error', while the difference between model prediction and observation in food products is called 'overall error' in Pin *et al.* (1999). Because laboratory conditions, such as substrate are optimal (apart from the very modelled factors), the primary error is always smaller than the overall error. The two would be the same only if laboratory conditions mimic food conditions exactly.

In order to quantify these errors, Ross (1996) proposed two indices: the bias factor, which indicates whether the model, on average, gives greater or smaller predictions than the independent observations used in the comparison, and the accuracy factor, an average difference between observations and predictions.

Baranyi *et al.* (1999) suggested a modification of the accuracy factor (A), by introducing the root mean square difference between predictions and observations. Let $\mu(\mathbf{v}_i)$ denote the observed, $f(\mathbf{v}_i)$ the predicted rate at $(\mathbf{v}_i(i = 1 \ldots n)$, the series of environmental vectors where n observations were made. Then, the accuracy factor is defined as:

$$A = \exp\left(\frac{\sqrt{\sum_{i=1}^{n}(\mathrm{Ln}(f(\mathbf{v}_i)) - \mathrm{Ln}(\mu(\mathbf{v}_i)))^2}}{n}\right) \tag{18.15}$$

Similarly, the bias factor is defined as

$$B = \exp\left(\frac{\sum_{i=1}^{n}\mathrm{Ln}(f(\mathbf{v}_i)) - \mathrm{Ln}(\mu(\mathbf{v}_i))}{n}\right) \tag{18.16}$$

One of the advantages of these indices is that they can be used to estimate the precision and the bias of a specific model compared to another model. If, for example, $\mathbf{v}_i = [\mathrm{Temp}_i, \mathrm{pH}_i \ldots]$ and the functions $f(\mathrm{Temp}, \mathrm{pH}\ldots)$ and $g(\mathrm{Temp}, \mathrm{pH}\ldots)$ are two models of the same growth parameter, then they can be compared by

$$A = \exp\left(\sqrt{\frac{\int_V (\mathrm{Ln}(f(\mathrm{Temp}, \mathrm{pH}\ldots)) - \mathrm{Ln}(g(\mathrm{Temp}, \mathrm{pH}\ldots)))^2 d\mathrm{Temp}\, d\mathrm{pH}\ldots}{V}}\right) \tag{18.17}$$

$$B = \exp\left(\sqrt{\frac{\int_V (\mathrm{Ln}(f(\mathrm{Temp}, \mathrm{pH}\ldots)) - \mathrm{Ln}(g(\mathrm{Temp}, \mathrm{pH}\ldots)) d\mathrm{Temp}\, d\mathrm{pH}\ldots}{V}}\right) \tag{18.18}$$

where V represents the intersection of the interpolation regions of these models.

In both above cases (discrete and continuous), the indices can be expressed as percentages. The percentage discrepancy or error (%D) is

$$\%D = (A - 1) \cdot 100$$

and the percentage bias (%B) is

$$\%B = sign(\mathrm{Ln}\,B) \cdot (\exp|\mathrm{Ln}\,B| - 1) \cdot 100$$

where:

$$sign(\mathrm{Ln}\,B) = \begin{cases} +1 & \text{if} \quad \mathrm{Ln}\,B > 0 \\ 0 & \text{if} \quad \mathrm{Ln}\,B = 0 \\ -1 & \text{if} \quad \mathrm{Ln}\,B < 0 \end{cases}$$

When the values predicted by the model are higher than the observations, in the discrete case (or higher than those of other models in comparison, in the continuous case), the bias percentage is positive.

In the following section, two examples are shown for applying these indices. Predictions of the model by Neumeyer *et al.* (1997a), developed for *Pseudomonas* spp., have a mean discrepancy of 30% when compared with data obtained for milk and meat products. The percentage bias indicates that, on average, the predictions were +15% above observed values. The model is biased but fails safe because it predicts faster growth rates than the observed ones (Neumeyer *et al.*, 1997b). The model for *Pseudomonas* spp. developed by Pin and Baranyi (1998) has a mean discrepancy of 46% from observations made in meat spoilt by its natural microbiota. The predicted values of this model were, in almost all cases, higher than the observed values (bias: +35%). The discrepancy of this model (46%) seems to be high but it is important to notice that this was mainly due to the bias of the model. This means that, even if the primary error of the model was small (because the laboratory conditions were more favourable than those in food) the overall error is big. In the main, positive bias of the model indicates safe predictions.

18.5 Available software packages

Probably the most general predictive microbiology software packages are the Pathogen Modelling Program (USDA Eastern Regional Research Centre, Wyndmoor, PA, USA) and the Food MicroModel (Leatherhead Food Research Association, UK). We will refer to them as **PMP** and **FMM**, respectively.

PMP is a nicely written, user-friendly software package. It can be freely downloaded via www.arserrc.gov/mfs/pathogen.htm. The program predicts bacterial growth curves at user-defined sets of values of temperature, pH and NaCl concentration (or water activity). For some organisms, the effect of a fourth factor, such as a certain added preservative or atmospheric composition can also be studied. The bacterial growth is represented by a Gompertz sigmoid

curve, where the Gompertz parameters are modelled by a quadratic response surface fitted to observed values. Advantages are that it is free, easy-to-use, and has nice graphics. Disadvantages are that the user can give only individual predictions of the Gompertz parameters and these cannot be collected in a table for a given set of environmental factors, to copy them to other applications. The lag is interpreted in a static way (consequence of the empirical Gompertz growth curve, as analysed in section 18.2.1). Besides, because the raw data on which the models are based were measured in laboratory conditions, no information is provided on how reliable the model predictions are in real food systems.

FMM is a similar, perhaps more versatile program, based on similar mathematical treatment. However, it is not free and not even cheap. The biggest difference between PMP and FMM is that, at least in the manual for FMM, the user can get an impression of the performance of the models in real food. For further information consult the website www.foodmicromodel.com

A service called **Forecast**, where the predictions are based on models as above, is run via telephone by the Campden and Chorleywood Food Research Association in the UK (+44 (0)1386 842071).

Some other predictive microbiology packages are created for purposes that are more specific. A nicely written program for predicting seafood spoilage at user-defined values of the environmental factors is the Seafood Spoilage Predictor (**SSP**) of the Danish Institute of Fisheries Research. It is freely available from www.dfu.min.dk/micro/ssp. Another example is the Food Spoilage Predictor (see http://www.hdl.com.au/html/body-fsp.htm) which is a time-temperature integration software, predicting the increase in numbers of psychotrophic spoilage pseudomonads in food.

Not predictors, rather packages helping to create predictive models are available from the Institute of Food Research. The program's and **DMFit** are **Microfit** downloadable from the IFR website, www.ifr.bbsrc.ac.uk. The first one fits the model of Baranyi and Roberts (1994) (see section 18.2.2) to measured concentrations of growing bacterial population. The user can carry out a significance-test to compare the specific growth rates of different growth curves. The second one, DMFit, is an Excel add-in, fitting, plotting and analysing many growth curves simultaneously.

18.6 Modelling bacterial growth by a stochastic birth process: *a candidate for future research*

Finally, we mention a relatively new approach (at least new in predictive microbiology) to predict bacterial growth and survival in food. Stochastic birth processes have long been applied in biotechnology (Tsuchiya *et al.*, 1966; Frederickson *et al.*, 1967) and medical studies (Armitage *et al.*, 1965) to model bacterial growth, but it was not until recently that the interests of modellers in food microbiology also turned to this technique. The reason is that predicting the lag and probability of survival of pathogens has become of primary importance

recently and, at low numbers of cells, deterministic models are no longer suitable to describe bacterial kinetics.

The distribution of resistance and adaptability of cells in unfavourable environments is at the centre of this investigation. Below, we show a basic relation between deterministic and stochastic models of bacterial lag.

Let the initial number of a growing cell population be N. The lag defined in the traditional way (see the deterministic models above) will be called the *population lag* and its value will be denoted by $\lambda(N)$. Our investigation focuses on the connection between $\lambda(N)$ and the lag times of the individual cells.

Considering the lag as an adjustment period, a cell does not necessarily divide after the lag, but the first generation time begins, which already belongs to the exponential period. Therefore, the time to the first division is the sum of the lag time and the generation time. The lag for the ith cell of the initial population will be denoted by $\tau_i (i = 1, 2 \ldots N)$. Suppose that the τ_i *individual lag times* $(i = 1, 2 \ldots N)$ are identically distributed independent random variables, and their expected value is $E(\pi_i) = \tau$.

If the cell population grows according to the classical Poisson birth process, the birth intensity parameter is the same as the specific rate (μ) of the population. As before, let $y(t) = \mathrm{Ln}\, x(t)$ denote the logarithm of that population. Suppose that the subsequent subpopulations of each cell of the initial culture grow together, but independently of each other. Then, the expected population lag produced by a single cell is $\lambda(1) = \tau$. Baranyi (1998) proved mathematically that the population lag, generated by N initial cells, is

$$\lambda(N) = -\frac{1}{\mu}\mathrm{Ln}\frac{\sum_{i=1}^{N} e^{-\mu\tau_i}}{N} \tag{18.19}$$

Using the $\alpha_i = \exp(-\mu\tau_i)$ $(i = 1, 2, \ldots N)$ notation,

$$\alpha(N) = e^{-\mu\lambda(N)} = \frac{\sum_{i=1}^{N} \alpha_i}{N} \tag{18.20}$$

The α_i variables are also identically distributed, and the right-hand side of the equation (18.20) is their arithmetic mean. That means that the parameter $\alpha = -\ln(\mu\lambda)$, introduced by Baranyi and Roberts (1994) for deterministic models to quantify the cells' suitability to the actual environment, is a limit value of the $\alpha(N)$ parameters. In other words, the 'suitability parameter' of the population, α, is independent of the population size and it is equal to the expected value of the suitability parameters of the individual cells. This simple relation does not hold between the population lag and the mean individual lag.

Care must be taken therefore, when conclusions are drawn from deterministic models to low cell concentrations. From stochastic birth models, knowing the lag and generation time distributions, the expected population kinetics can be calculated, but this is not true *vice versa*. If one measures the distribution of

population growth parameters from replicate experiments, one cannot conclude the distribution of those parameters among the individual cells, because the variability of the individual cells disappears at the population level, and the observed variability is in fact due to other extra-cellular effects. New techniques are necessary, therefore, to measure individual cell kinetics. Our expectation is that stochastic modelling will be one of the next developments in predictive microbiology.

18.7 References

ARMITAGE P, MEYNELL G G, WILLIAMS T, (1965). Birth-death and other models for microbial infection. *Nature*, **207** 570–2.

BARANYI J, (1998). Comparison of stochastic and deterministic concepts of bacterial lag. *J. Theor. Biol.*, **192** 403–8.

BARANYI J, ROBERTS T A, (1994). A dynamic approach to predicting bacterial growth in food. *Int. J. Food Microbiol.*, **23** 277–94.

BARANYI J, ROBERTS T A, (1995). Mathematics of predictive food microbiology. *Int. J. Food Microbiol.*, **26** 199–218.

BARANYI J, PIN C, ROSS T, (1999). Validating and comparing predictive models. *Int. J. Food Microbiol.* **48** 159–66.

BARANYI J, ROSS T, ROBERTS T A, MCMEEKIN T, (1996). The effects of overparameterisation on the performance of empirical models used in predictive microbiology. *Food Microbiol.*, **13** 83–91.

BROUGHALL J M, BROWN C, (1984). Hazard analysis applied to microbial growth in foods: Development and application of three-dimensional models describing the effect of water activity. *Food Microbiol.*, **1** 13–22.

COLEMAN B D, (1978). Nonautonomous logistic equations as models of the adjustment of populations to environmental change. *Math. Biosci.*, **45** 159–73.

DAVEY K R, (1989). A predictive model for combined temperature and water activity on microbial growth during the growth phase. *Journal of Applied Bacteriology*, **67** 483–8.

FREDERICKSON A G, RAMKRISHNA D, TSUHIYA H M, (1967). Statistics and dynamics of procaryotic cell populations. *Math. Biosci.*, **1** 327–74.

MCCLURE P J, BARANYI J, BOOGARD E, KELLY T M, ROBERTS T A, (1993). A predictive model for the combined effect of pH, sodium chloride and storage temperature on the growth of *Brochothrix thermosphacta*. *Int. J. Food Microbiol.* **19** 161–78.

MCMEEKIN T A, OLLEY J N, ROSS T, RATKOWSKY D A, (1993). *Predictive Microbiology*. John Wiley & Sons Ltd. Chichester, UK.

MALTHUS T R, (1798). An Essay on the Principle of Population, 1st edn. printed for J. Johnson in St Paul's Churchyard, London.

NEUMEYER K, ROSS T, MCMEEKIN T A, (1997a). Development of a predictive model to describe the effects of temperature and water activity on the growth of spoilage pseudomonads. *Int. J. of Food Microbiol.*, **38** 45–54.

NEUMEYER K, ROSS T, THOMSON G, MCMEEKIN T A, (1997b). Validation of a model describing the effects of temperature and water activity on the growth of psychotrophic pseudomonads. *Int. J. Food Microbiol.*, **38** 55–63.

PIN C, BARANYI J, (1998). Predictive models as means to quantify the interactions of spoilage organisms. *Int. J. Food Microbiol.*, **41** 59–72.

PIN C, SUTHERLAND J P, BARANYI J, (1999). Validating predictive models of food spoilage organisms. *Journal of Applied Microbiology.*, **87** 491–9.

PIN C, GARCÍA DE FENANDO G D, ORDÓÑEZ J A, BARANYI J, (2000). Applying a generalized *z*-value concept to quantify and compare the effect of environmental factors on the growth of *Listeria monocytogenes*. Submitted.

PRESS W H, FLANNERY B P, TEUKOLSKY S A, VETTERLING W T, (1992). *Numerical Recipes*. Cambridge University Press, Cambridge.

RATKOWSKY D A, LOWRY R K, MCMEEKIN T A, STOKES A N, CHANDLER R E, (1983). Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *J. Bacteriol.* **154** 1222–6.

RATKOWSKY D A, OLLEY J, MCMEEKIN T A, BALL A, (1982). Relationship between temperature and growth rate of bacterial cultures. *J. Bacteriol.*, **149** 1–5.

RATKOWSKY D A, ROSS T, MACARIO N, DOMMETT T W, KAMPERMAN L, (1996). Choosing probability distributions for modeling generation time variability. *J. Applied Bacteriology* **80** 131–7.

RENSHAW E, (1991). *Modelling Biological Populations in Space and Time.* Cambridge University Press, Cambridge.

RICHARDS F J, (1959). A flexible growth function for empirical use. *J. Experimental Botany*, **10** 290–300.

ROBINSON T P, OCIO M J, KALOTI A, MACKEY B M, (1998). Effect of the growth environment on the lag phase of *Listeria monocytogenes*. *Int. J. Food Microbiol.*, **44** 83–92.

ROSS T, (1996). Indices for performance evaluation of predictive models in food microbiology. *J. Applied Bacteriology*, **81** 501–8.

SCHOOLFIELD R M, SHARPE P J, MAGNUSON C E, (1981). Non-linear regression of biological temperature-dependent rate models based on absolute reaction rate theory. *Journal of Theoretical Biology* **88** 719–31.

TSUCHIYA H M, FREDRICKSON A G, ARIS R, (1966). Dynamics of microbial cell populations. In: *Advances in Chemical Engineering*, **6** 125–206.

TURNER M E, BRADLEY E L, KIRK K A, PRUITT K M, (1976). A theory of growth. *Math. Biosci.* **29** 367–73.

ZWIETERING M H, JONGENBURGER I, ROMBOUTS F M, VAN'T RIET K, (1990). Modelling of the bacterial growth curve. *Appl. Environ. Microbiol.*, **56** 1875–81.

ZWIETERING M H, DE KOOS J T, HASENACK B E, DE WIT J C, VAN'T RIET K, (1991). Modeling of bacterial growth as a function of temperature. *Appl. Environ. Microbiol.*, **57** 1094–1101.

19

Modelling the use of time-temperature indicators in distribution and stock rotation

P. S. Taoukis, National Technical University of Athens

19.1 Introduction

At the dawn of the twenty-first century in an environment of momentous technological progress and evolution of consumer life style the food industry is called to deliver to seemingly contradictory market demands. While in general high technology is rapidly being accepted and absorbed, the position of the consumers with regards to their expectations of food products is relatively ambivalent. They seek upgraded sensory quality, increased functional and nutritional properties combined with a traditional, wholesome image, from food products of guaranteed safety but yet less processing, fewer additives and 'technological' interventions. At the same time they expect extended shelf-life and high convenience in preparation and use.

The attainment of longer shelf-life with minimum processing requires not only intense optimisation and control of all production and preservation parameters but often innovative techniques, to ensure safety and reduce food deterioration. The efforts of producers and regulators concentrate on the development and application of structured quality and safety assurance systems based on prevention through monitoring, recording and controlling of critical parameters through the entire product's life cycle. These systems include the post processing phase and ideally extend to the consumer's table. The ISO 9001 and 9002 quality management systems, widely adopted by the food industry, explicitly refer to documented procedures for storage, handling and distribution.[1] The need for the inclusion of the post processing phase in an efficient dynamically ameliorating system is further emphasised in the newly

released ISO 9000 2000 standard.[2] In practice, one of the weaker links of inadequate control has been the post processing phase. The globally recommended standard safety and hygiene assurance system, Hazard Analysis and Critical Control Point (HACCP) also focuses on this phase.[3–6] Certain stages of the chill chain are being recognised as important critical control points for new technology minimally processed chilled products such as MA packaged, cook and chill and other ready to eat chilled products. Effectively monitoring and controlling these CCPs or other points important to quality is a complicated task.

Practice and industrial studies, mostly unpublished, show that temperature conditions in chilled or frozen distribution and handling very often deviate from recommended ones. Since temperature largely constitutes the determining post-processing parameter for shelf-life under Good Manufacturing and Hygiene Practices, monitoring and controlling it would be of central importance. The complexity of such a proposition is highlighted when the variation in temperature exposure of single products within batches or transportation subunits is considered. Ideally, a cost-effective way to individually monitor the temperature conditions of food products throughout distribution would be required in order to indicate their real safety and quality state. Time-Temperature Indicators (TTI) could potentially fulfil the above requirements. A TTI based system could lead to effective quality control of the chill chain, optimisation of stock rotation and reduction of waste, and provide information on the remaining shelf-life of the product units. Prerequisite for application of this approach is the systematic study and kinetic modelling of the temperature dependence of shelf-life. Based on reliable models of food product shelf-life and the kinetics of TTI response the effect of temperature can be monitored, recorded and translated, from production to the consumer's table.

In the following sections the principles and methodology of modelling, that are being covered throughout this book, will serve as the basis for the description of an optimised food product distribution and handling system. The state of development and the potential of Time Temperature Integrators as tools and fundamental elements of such a system will be explored and demonstrated.

19.2 Definitions and history of TTI

19.2.1 Definitions and classifications

A *time-temperature indicator or integrator* (TTI) can be defined as a simple, inexpensive device that can show an easily measurable, time-temperature dependent change that reflects the full or partial temperature history of a food product to which it is attached.[7]

The principle of TTI operation is a mechanical, chemical, electrochemical, enzymatic or microbiological irreversible change usually expressed as a visible response, in the form of a mechanical deformation, colour development or colour movement. The rate of change is temperature dependent, increasing at

higher temperatures similar to most physicochemical reactions. The visible response thus gives a cumulative indication on the storage conditions that the TTI has been exposed to. The extent to which this response corresponds to a real time-temperature history depends on the type of the indicator and the physicochemical principles of its operation. Indicators can thus be classified according to the kind of their functionality and the information they convey. Different classifications and terminology have been proposed, partly reflecting the evolution of the indicators.

An early classification system introduced by Schoen and Byrne (1972) had indicating devices separated into six categories, including electronic temperature recorders.[8] Byrne (1976) revised the above classification realising that the main functional difference of interest is whether the indicator responds above a preselected temperature or responds continuously giving information on the cumulative time-temperature exposure.[9] He proposed three types: (a) Defrost Indicators, (b) Time-Temperature Integrators, and (c) Time-Temperature Integrators/Indicators. A similar scheme recognised three categories:[10] (a) Abuse indicators, (b) Partial temperature history indicators, and (c) Full temperature history indicators. Full history indicators is an alternative nomenclature for Time-Temperature Integrators.

A three-category classification will be used in this chapter:[11]

1. *Critical temperature indicators (CTI).* CTI show exposure above (or below) a reference temperature. They involve a time element (usually short; a few minutes up to a few hours) but are not intended to show history of exposure above the critical temperature. They merely indicate the fact that the product was exposed to an undesirable temperature for a time sufficient to cause a change critical to the safety or quality of the product. They can serve as appropriate warnings in cases where physicochemical or biological reactions show a discontinuous change in rate. Good examples of such cases are the irreversible textural deterioration that happens when phase changes occur (e.g., upon defrosting of frozen products or freezing of fresh or chilled products). Denaturation of an important protein above the critical temperature or growth of a pathogenic microorganism are other important cases were a CTI would be useful. The 'critical temperature' term is preferred rather than the used alternative 'defrost' that is too limiting. The term 'abuse' might be misleading as oftentimes undesirable changes, warranting warning, can happen at temperatures which are not as extreme or abusive as the term implies and which are presumed within the acceptable range of normal storage for the product in question.
2. *Critical temperature/time integrators (CTTI).* CTTI show a response that reflects the cumulative time-temperature exposure above a reference critical temperature. Their response can be translated into an equivalent exposure time at the critical temperature. They are useful in indicating breakdowns in the distribution chain and for products in which reactions, important to

quality or safety, are initiated or occur at measurable rates above a critical temperature. Examples of such reactions are microbial growth or enzymatic activity that are inhibited below the critical temperature. CTTI combinations can give a discretised approximation of real time-temperature history.

3. *Time-temperature integrators or indicators (TTI).* TTI give a continuous, temperature dependent response throughout the product's history. They integrate, in a single measurement, the full time-temperature history and can be used to indicate an 'average' temperature during distribution and possibly be correlated to continuous, temperature dependent quality loss reactions in foods. In the rest of this chapter, the term TTI will refer to Type 3 indicators, unless otherwise noted.

   A different method of classification sometimes used is based on the principle of the indicators' operation. Thus, they can be categorised as mechanical, chemical, enzymatic, microbiological, polymer, electrochemical, diffusion based, etc.

19.2.2 Requirements and properties of an ideal TTI

The requirements for an effective TTI are that it shows a continuous change, the rate of which increases with temperature and which does not reverse when temperature is lowered. There are a number of other desirable attributes for a successful indicator. An *ideal TTI* would have all the following properties:

- It exhibits a continuous time-temperature dependent change.
- The change causes a response that is easily measurable and irreversible.
- The change mimics or can be correlated to the food's extent of quality deterioration and residual shelf-life.
- It is reliable, giving consistent responses when exposed to the same temperature conditions.
- It has low cost.
- It is flexible, so that different configurations can be adopted for various temperature ranges (e.g., frozen, refrigerated, room temperature) with useful response periods of a few days as well as up to more than a year.
- It is small, easily integrated as part of the food package and compatible with a high speed packaging process.
- It has a long shelf-life before activation and can be easily activated.
- It is unaffected by ambient conditions other than temperature, such as light, RH and air pollutants.
- It is resistant to normal mechanical abuses and its response cannot be altered.
- It is nontoxic, posing no safety threat in the unlikely situation of product contact.
- It is able to convey in a simple and clear way the intended message to its target, be that distribution handlers or inspectors, retail store personnel or consumers.
- Its response is both visually understandable and adaptable to measurement by

electronic equipment for easier and faster information, storage and subsequent use.

19.2.3 History of TTI

The drive for development of an effective and inexpensive indicator dates from the time when the importance of the distribution temperature variations to final food quality became apparent. Initially, the interest was focused on frozen foods. The first application of a 'device' to indicate handling abuse dates from World War II when the US Army Quartermaster Corps used an ice cube placed inside each case of frozen food. Disappearance of the cube indicated mishandling.[8] The first patented indicator goes back to 1933.[12] Over a hundred US and International patents relevant to Time-Temperature Indicators have been issued since. During the last 30 years numerous TTI systems have been proposed of which only few reached the prototype and even less the market stage.[9, 11, 13] Byrne (1976) gives an overview of the early indicators and Taoukis (1989) presents a detailed history of TTI.[9, 13] Taoukis *et al.* (1991) tabulated TTI patents that have been issued since 1976 updating a then published list.[9, 11] In Table 19.1 significant TTI patents of the past decade are listed and classified according to type and principle of operation.

The first commercially available TTI was developed by Honeywell Corp. (Minneapolis, MN).[14] The indicator was tested by the USDA and was judged reliable.[15] The device never found commercial application, possibly because it was costly and relatively bulky, and by the 1970 was not available. In the early seventies, the US government considered mandating the use of indicators on

Table 19.1 List of recent TTI patents and classification according to type and mode of response.

Date	Inventor	Principle of operation	Patent No.
1991	Jalinski, T.J.	Chemical (TTI)	US5,182,212
1991	Jalinski, T.J.	Chemical (TTI)	US5,085,802
1991	Thierry, A.	Chemical (CTI)	US5,085,801
1991	Swartzel, K.R.	Physicochemical (TTI)	US5,159,564
1992	Jalinski, T.	Chemical (CTI)	EP497459A1
1993	Veitch, R.J.	Physicochemical (CTI)	EP563769A1
1993	Loustaunau, A.	Physical (CTI)	EP615614A1
1994	Loustaunau, A.	Physical (CTI)	US5,460,117
1994	Veitch, R.J.	Physicochemical (CTI)	US5,490,476
1995	Prusik, T.	Physicochemical (TTI)	US5,709,472
1996	Cannelongo, J.F.	Physical (CTI)	US5,779,364
1996	Veitch, R.J.	Physical (CTI)	EP835429A1
1997	Arens R. *et al.*	Physicochemical (TTI)	US5,667,303
1997	Schneider, N.	Physical (CTI)	US6,030,118
1999	Simons, M.J.	Physicochemical (CTI)	EP930488A2
2000	Schaten, B.B.	Physical (CTI)	EP1053726A2

certain products.[16] This generated a flurry of research and development. Researchers at the US Army Natick Laboratories developed a TTI that was based on the colour change of an oxidisable chemical system controlled by the temperature dependent permeation of oxygen through a film.[17] Field testing over a two-year period with the TTI attached to rations showed their potential for use.[18] The system was contracted to Artech Corp. (Falls Church, VA) for commercial development. By 1976 six companies were making temperature indicators at least at the prototype stage.[9, 19] The Artech, the Check Spot Co. (Vancouver, WA) (US patent 2,971,852) and the Tempil (S. Plainfield, NJ) indicators could be classified as CTI, whereas I-Point (Malmö, Sweden), the Bio-Medical Sciences (Fairfield, NJ) (US patents 3,946,611 and 4,042,336) and the 3M Co. (St. Paul, MN) indicators were TTI. The Tempil indicator could function as a CTTI. It involved a change to a red colour and subsequent movement when exposed above the critical temperature. The I-Point was an enzymatic TTI, and the 3M, a diffusion based TTI. By the end of the 1970s, very little commercial application of the TTI was achieved. A number of the aforementioned systems were discontinued. Activity in the area of TTI subsided temporarily, noted by a decrease in the relevant publications and in the new TTI models introduced. However, the more sound systems remained available and development continued aiming towards fine tuning of their characteristics and making them more consistent with their claimed performance. In the early 1980s, there were four systems commercially available including the I-Point and the 3M TTI. Andover Labs (Weymouth, MA) marketed up to 1985 the Ambitemp and Tempchron devices. Both were for use in frozen food distribution and could be classified as CTTI. Their operation was based on the displacement of a fluid along a capillary.

19.2.4 Current TTI systems

In the last fifteen years three types of TTI have been the focus of both scientific and industrial trials. They claim to satisfy the requirements of a successful TTI and have evolved as the major commercial types in the market. They are described in detail in the following sections, coded in this chapter as TTIs A, B and C.

TTI A is a diffusion based indicator, the 3M Monitor Mark® (3M Co., St. Paul, Minnesota) (US Patent, 3,954,011, 1976). One of the first significant applications of TTI was the use of this indicator by the World Health Organization (WHO) to monitor refrigerated vaccine shipments. The response of the indicator is the advance of a blue dyed ester diffusing along a wick. The useful range of temperatures and the response life of the TTI are determined by the type of ester and the concentration at the origin. Thus the indicators can be used either as CTTI with the critical temperature equal to the melting temperature of the ester or as TTI if the melting temperature is lower than the range of temperatures the food is stored at, e.g., below 0ºC for chilled storage. The same company has marketed recently the successor to this TTI the Monitor

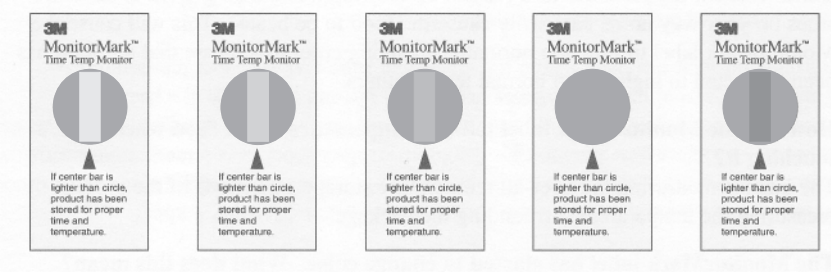


Fig. 19.1 Diffusion based TTI Type A.

Mark® Temperature Monitor (Fig. 19.1) and Freshness Check, based on diffusion of proprietary polymer materials (US patent 5,667,303).

A viscoelastic material migrates into a diffusely light-reflective porous matrix at a temperature dependent rate. This causes a progressive change of the light transmissivity of the porous matrix and provides a visual response. The response rate and temperature dependence is controlled by the tag configuration, the diffusing polymer's concentration and its glass transition temperature and can be set at the desirable range. The TTI is activated by adhesion of the two materials that before use can be stored separately for a long period at ambient temperature.

TTI B is an enzymatic indicator, VITSAB Time Temperature Indicator, successor of the I-Point Time Temperature Monitor (VITSAB A.B., Malmö, Sweden). The indicator is based on a colour change caused by a pH decrease which is the result of a controlled enzymatic hydrolysis of a lipid substrate (US Patents 4,043,871 and 4,284,719). Before activation the indicator consists of two separate compartments, in the form of plastic mini-pouches. One compartment contains an aqueous solution of a lipolytic enzyme, such as pancreatic lipase. The other contains the lipid substrate absorbed in a pulverised PVC carrier and suspended in an aqueous phase and a pH indicator mix. As substrates, glycerine tricapronate (tricaproin), tripelargonin, tributyrin, and mixed esters of polyvalent alcohols and organic acids are mentioned. Different combinations of enzyme-substrate types and concentrations can be used to give a variety of response lives and temperature dependencies. At activation, enzyme and substrate are mixed by mechanically breaking the barrier that separates the two compartments. Hydrolysis of the substrate (e.g., tricaproin) causes acid release (e.g., caproic acid) and the pH drop is translated in a colour change of the pH indicator from deep green to bright yellow. Reference starting and end point colours are printed around the reaction window to allow easier visual recognition and evaluation of the colour change (Fig. 19.2). The continuous colour change can also be measured instrumentally.[7] The TTI Type B are claimed to have a long shelf-life if kept chilled before activation.

Fig. 19.2 Enzymatic TTI Type B.

TTI C, Lifelines Freshness Monitor® and Fresh-Check® indicators (Lifelines Inc., Morris Plains, NJ) are based on a solid state polymerisation reaction (US Patent, 3,999,946 and 4,228,126).[22] The TTI function is based on the property of disubstituted diacetylene crystals ($R-C=C-C=C-R$) to polymerise through a lattice-controlled solid-state reaction proceeding via 1,4-addition polymerisation and resulting in a highly coloured polymer. During polymerisation, the crystal structure of the monomer is retained and the polymer crystals remain chain aligned and are effectively one dimensional in their optical properties.[23] The response of the TTI is the colour change measured as a decrease in reflectance. Freshness Monitor consists of an orthogonal piece of laminated paper the front face of which includes a strip with a thin coat of the colourless diacetylenic monomer and two barcodes, one about the product and the other identifying the model of the indicator. The Fresh-Check® version, for consumers, is round, and the colour of the 'active' centre of the TTI is compared to the reference colour of a surrounding ring (Fig. 19.3). The laminate has a red or yellow colour so that the change is perceived as a change from transparent to black. The reflectance of the Freshness Monitor can be measured by scanning with a laser optic wand and stored in a hand-held device supplied by the TTI producer. The response of Fresh Scan can be visually evaluated in comparison to the reference ring or continuously measured by a portable colorimeter or an optical densitometer. Before use the indicators, active from the time of production, have to be stored deep frozen where change is very slow.

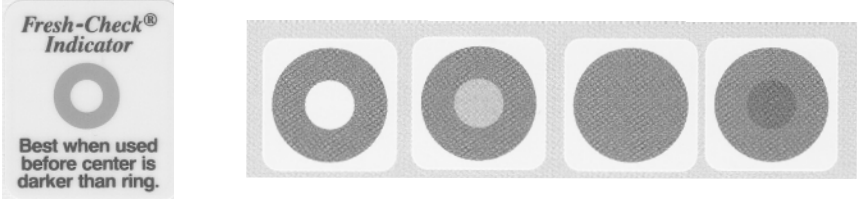


Fig. 19.3 Polymer based TTI Type C.

Despite the potential of TTI to substantially contribute in improved food distribution, reduce food waste and benefit the consumer with more meaningful shelf-life labelling their application up to now has not lived up to the initial expectations. The main reasons for the reluctance of food producers to adopt the TTI have been cost, reliability, and applicability. The cost is volume dependent, ranging from 2 to 20 cents per unit, and if the other questions are resolved, the cost-benefit analysis should well be in favour of the indicators. The reliability question has its roots in the history of indicators, due partly to exaggerated claims by manufacturers of some early models and partly on lack of sufficient data, both from studies and from the suppliers. Initial attempts in using TTI as quality monitors were not well designed and hence unsuccessful. Re-emerging discussions by regulatory agencies to make the TTI use mandatory, before the underlying concepts were understood and their reliability demonstrated, resulted in resistance by the industry and may have hurt TTI application up to the present time. Current TTI systems have achieved high standards of production quality assurance and provide reliable and reproducible responses according to the specifications stated. Testing standards have been issued by the BSI and can be used by the TTI manufacturers as well as the TTI users.[24] Testing and modelling of the TTIs will be covered in depth in the next sections.

The question of applicability, however, has been the most substantial hurdle to TTI use. Suppliers and earlier studies have been ineffective in establishing a clear methodology on how the TTI response can be used as a measure of food quality. The initial approach was to assume an overall temperature dependence curve (or zone) for the shelf-life of a general class of foods, e.g., frozen foods, and aim for an indicator that has a similar temperature dependence curve for the time to reach a specific point on its scale. Such a generalisation proved insufficient as even foods of the same type differed significantly in the temperature dependence of the deterioration of their quality. What is needed is a thorough knowledge of the shelf-life loss behaviour of the food system to be monitored, expressed quantitatively in, as accurately as possible, kinetic models. It is not reasonable to expect the TTI monitoring ability to improve on the ability to predict quality and shelf-life of a food exposed to fully known temperature conditions, such as one would record e.g. with an electronic data logger attached to the food. Such unfounded expectations have been one of the main reasons for the slow progress and often observed reversals in the history of TTI application.

It has been widely assumed and proposed that the behaviour of the TTI should strictly match that of the particular food to be monitored at all temperatures. This approach, even if feasible, is impractical, and requires an unlimited number of TTI models. Instead of a TTI exactly mimicking quality deterioration behaviour of the food product, a meaningful, general scheme of translating TTI response to food status is needed. This should be based on systematic modelling of both the TTI and the food. This kinetic modelling approach and the methodology for applying TTI for food quality monitoring and chill chain optimisation will be detailed in the next sections.

19.3 Food quality modelling

Kinetic modelling of food quality has also been addressed in previous chapters of this book. Effective and quantitative knowledge of food deterioration and determination and modelling of the shelf-life or keeping quality of the food products is the most important prerequisite for the application of a TTI based monitoring system. For consistency of approach and terminology, the principles and equations of food quality modelling used in the development of the TTI scheme will be outlined in the present section.

Food is a physicochemical system of high complexity involving numerous physical and chemical variables. It is imposible or impractical to quantitatively determine all these variables. Food quality change in general may be expressed as a function of composition and environmental factors:

$$\frac{dQ}{dt} = F(C_i, E_j) \tag{19.1}$$

where C_i are composition factors, such as concentration of reactive compounds, inorganic catalysts, enzymes, reaction inhibitors, pH, water activity, as well as microbial populations and E_j are environmental factors, such as temperature, relative humidity, total pressure and partial pressure of different gases, light and mechanical stresses. Even if this system could be explicitly expressed in terms of measurable parameters, no analytical solution is attainable and possible numerical solutions are too elaborate for any practical purpose. The established methodology consists of first identifying the chemical and biological reactions that influence the quality and the safety of the food. Then, through a careful study of the food components and the process, the reactions judged to have the most critical impact on the deterioration rate, are deternined.[25] Based on this analysis and without underestimating the underlying complexity of food systems, food degradation and shelf-life loss is in practice represented by the loss of desirable quality factors A (such as nutrients, characteristic flavours) or the formation of undesirable factors B (such as off flavours, microbial load, discoloration). The rate of loss of A (correspondingly of formation of B) is expressed as:

$$r_A = \frac{-d[A]}{dt} = k[A]^m \tag{19.2}$$

The quality factors [A] are usually quantifiable chemical, physical, microbiological or sensory parameters characteristic of the particular food system, k the apparent reaction rate constants and m the apparent reaction order. The apparent reaction order and constant are determined by fitting the change with time of the experimentally measured values of [A] to equation (19.2). In the integral method, variables are separated and integration is carried out. Thus for equation (19.2), we have:

$$-\int_{A_0}^{A} \frac{d[A]}{[A]^m} = kt \tag{19.3}$$

Table 19.2 Quality function and half life time for deterioration of quality index A

Apparent reaction order	Quality function $f(A)_t$	Half life $t_{1/2}$
0	$A_0 - A_t$	$A_0/(2k_0)$
1	$\ln(A_0 - A_t)$	$\ln 2/k_1$
2	$1/A_0 - 1/A_t$	$1/(k_2 A_0)$
$m\,(m \neq 1)$	$\frac{1}{m-1}(A_t^{1-m} - A_0^{1-m})$	$\frac{2^{m-1}-1}{k_m(m-1)} - A_0^{1-m}$

Regardless of the value of m (the order of the reaction) equation (19.3) can be expressed in the form:

$$f(A) = kt \tag{19.4}$$

where the expression $f(A)$ is defined as the *quality function* of the food.

The form of the quality function of the food for an apparent zero, 1st, 2nd and mth order reaction can be derived from equation (19.4) and is shown in Table 19.2 along with the half life time of the reaction i.e. the time for the concentration of the quality index A to reduce to half its initial value.

A more detailed discussion on the definition and methodology of determination of the quality function and the involved caveats can be found in Taoukis *et al.* (1997).[26]

In order to include in the quality function the effect of the environmental factors the commonly used approach is to model it into the apparent reaction rate constant, i.e. expressing k of equation (19.4) as a function of E_j: $k = k(E_j)$. The factor most often considered and studied is temperature. This is justifiable because temperature not only strongly affects reaction rates but is also directly imposed to the food externally (direct effect of the environment), the other factors being at least to some extent controlled by the food packaging. The Arrhenius relation developed theoretically for chemical and biochemical reactions, has been experimentally shown to hold empirically for a number of more complex chemical and physical phenomena (e.g., viscosity, diffusion, sorption). Food quality loss reactions described by the aforementioned kinetic models have also been reported widely in the literature to follow an Arrhenius behaviour for temperatures within the particular temperature range of practical interest for each food product category. Deviations from the Arrhenius equation and alternative approaches for modelling the temperature dependence such as the Eyring, the Belehradek and the WLF equations have been addressed in other chapters of this book and in the literature.[26, 27]

In summary, the shelf-life of a food product evaluated by the measurement of a characteristic quality index, A, can be expressed as:

$$f(A) = kt = k_{\text{ref}} \exp\left[\frac{-E_A}{R}\left(\frac{1}{T} - \frac{1}{T_{\text{ref}}}\right)\right] t \tag{19.5}$$

where $f(A)$ is the *quality function* of the food and k the reaction rate constant. The rate constant is an exponential function of inverse absolute temperature, T, given by the shown Arrhenius expression, where k_{ref} is the reaction rate constant at a reference temperature T_{ref}, E_A is the activation energy of the reaction that controls quality loss and R the universal gas constant. The activation energy of food related chemical reactions and spoilage or pathogenic microbial growth usually falls within 30–120 kJ/mol. The reference temperature used is characteristic of the storage range of the food, e.g. for chilled foods $T_{\text{ref}} = 273$ K can be used.

The value of the quality function, $f(A)_t$, at time t, after exposure of the food at a known variable temperature exposure, $T(t)$, can be found based on equation (19.5) by calculating the integral of $k[T(t)]\,dt$, from 0 to time t.

$$f(A)_t = k_{\text{ref}} \int_0^t \exp\left[\frac{-E_A}{R}\left(\frac{1}{T(t)} - \frac{1}{T_{\text{ref}}}\right)\right] dt \tag{19.6}$$

The integral can be calculated analytically for simple $T(t)$ functions or numerically for more complex ones.[26]

We can define the effective temperature, T_{eff}, as the constant temperature that results in the same quality change as the variable temperature distribution, $T(t)$. Thus, $f(A)_t$ can be expressed as

$$f(A) = kt = k_{\text{ref}} \exp\left[\frac{-E_A}{R}\left(\frac{1}{T_{\text{eff}}} - \frac{1}{T_{\text{ref}}}\right)\right] t \tag{19.7}$$

19.4 TTI response modelling – application scheme

The same kinetic approach as described for food quality can be used to model the measurable change X of the TTI. If a *response function* $F(X)$ can be defined such that $F(X) = k_I t$, with k_I an Arrhenius function of T, the response of the TTI can be expressed by the following equation:

$$F(X) = k_I t = k_{I_{\text{ref}}} \exp\left[\frac{-E_{A_I}}{R}\left(\frac{1}{T} - \frac{1}{T_{\text{ref}}}\right)\right] t \tag{19.8}$$

where $k_{I_{\text{ref}}}$ is the TTI response rate constant at the reference temperature T_{ref} and E_{A_I} is the activation energy of the response.

The effective temperature concept as described above can also be used for the TTI. For an indicator exposed to the same temperature profile, $T(t)$, as the food product, the response function can be expressed by equations similar to (19.6) and (19.7) which for the TTI take the form:

$$F(X)_t = k_{I_{\text{ref}}} \int_0^t \exp\left[\frac{-E_{A_I}}{R}\left(\frac{1}{T(t)} - \frac{1}{T_{\text{ref}}}\right)\right] dt \tag{19.9}$$

Using the effective temperature, $T_{eff\,(TTI)}$, i.e. the constant temperature that results in the same TTI response as the variable temperature profile, $T(t)$, $F(X)_t$ can be expressed as

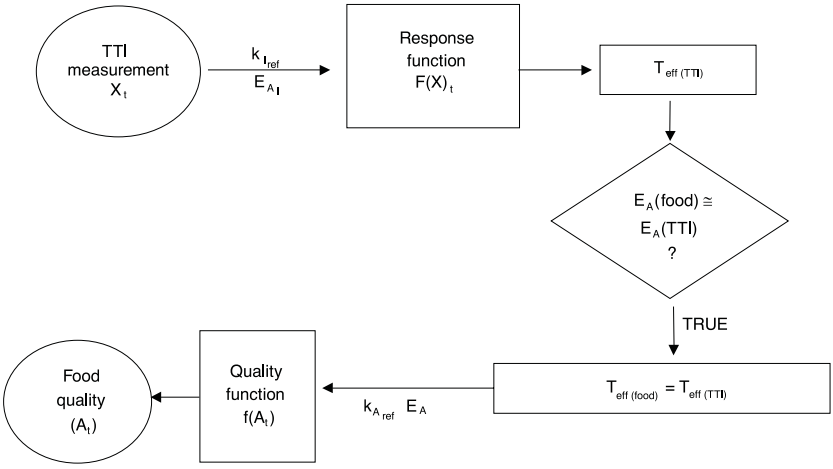


Fig. 19.4 Schematic representation of the systematic approach for applying TTI as quality monitors.

$$F(X)_t = k_I t = k_{I_{\text{ref}}} \exp\left[\frac{-E_{A_I}}{R}\left(\frac{1}{T_{\text{eff(TTI)}}} - \frac{1}{T_{\text{ref}}}\right)\right] t \qquad (19.10)$$

Based on the above kinetic equations an application scheme was developed that allowed the calculation of the value of the quality factor value A, at any time t, from the measured change X of the indicator, at that time.[7] The application scheme is illustrated in Fig. 19.4. From the measured value X of the TTI at time t the value of the response function is calculated, from which by solving equation (19.10), T_{eff} is derived. If the TTI and the food product quality deterioration reactions have similar temperature dependence, translated into activation energies differing by less than 25 kJ/mol, the same T_{eff} can be used for the food. With the T_{eff} and the kinetic parameters of the quality loss reaction known, the quality function $f(A)$ value is calculated from equation (19.7) and from it the value of the quality index A_t is found. Knowledge of the value of A gives the extent of the quality deterioration of the food. It also allows the calculation of the remaining shelf-life at any assumed average temperature.

To apply the developed systematic approach, the response characteristics of the different types of TTI *i.e.*, the $F(X)$ expression and the values of k_I and E_A (if they show Arrhenius behaviour) must be determined from kinetic experiments.[7, 28, 29] The author and co-workers have conducted extensive testing and modelling work on all available types of TTIs. The response and Arrhenius plots of the response rates of selected TTIs of Types A, B, and C are shown in Figures 19.5, 19.6 and 19.7. Continuous objective instrumental CIELab measurements of the colour change with a Minolta 200 and/or a Hunter Miniscan colorimeter were employed. Different indices that quantified better the response of each TTI type were used. Then the response function, such that $F(X) = k_I t$, was determined. From the $F(X)$ v. time plots, the value of the rate of the TTI

response k was determined at each temperature by linear regression analysis (Fig. 19.5a, 19.6a and 19.7a). The correlation coefficients of the fit were 0.9 or larger. The temperature dependence of the response rates, k, was modelled by the Arrhenius equation. For TTI Type A the chromaticity change $\Delta E = (\Delta L^2 + \Delta a^2 + \Delta b^2)$, from the Lab values corresponding to the end point colour, was used as the response X of the TTI. For TTI Type B the normalized chroma $X_c = (C - C_{\min})/(C_{\max} - C_{\min})$, where $C = (a^{*2} + b^{*2})^{1/2}$, $C_{\min}$ is the minimum measured chroma value at the time of TTI activation and $C_{\max}$ the chroma value that corresponds to the colour reached long after what is considered as the end point colour of the TTI, was used as the response X of the TTI. The index that quantified better the change of colour of TTI Type C with time was the b value of the CIELab system i.e. $X = b$.

Table 19.3 summarises the kinetic parameters of representative models of the three types of TTI. The E_A values of the three indicators cover the range of the most important deteriorative reactions in foods. The kinetic expressions of the

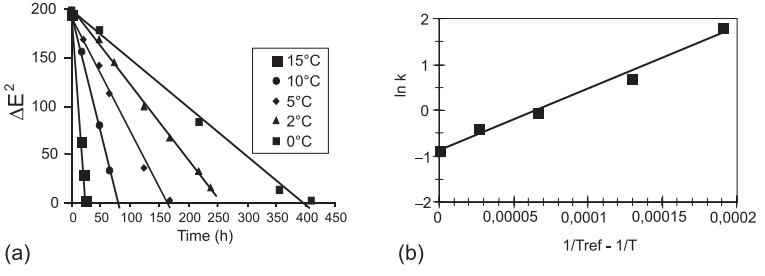


Fig. 19.5 (a) Response of a TTI Type A at different isothermal storage conditions. $F(X) = \Delta E^2$ is the response function of the indicator. Points are average of measured response of ten TTI units at each temperature. Lines are the regression fit lines for $F(X)$ v. time. (b) Arrhenius plot of the response rate of TTI Type A. $T_{ref} = 273$ K.

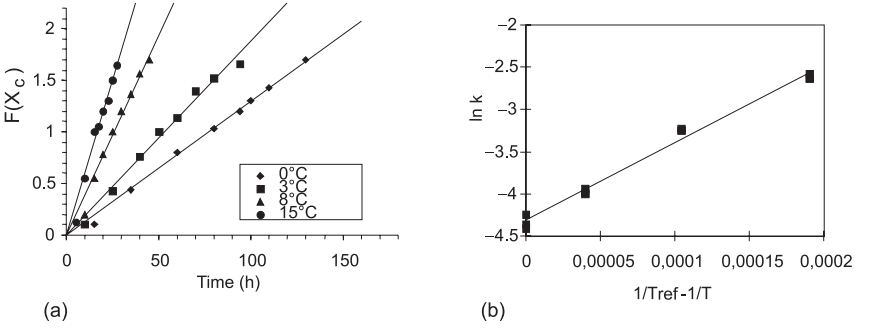


Fig. 19.6 (a) Response of a TTI Type B – Model M at different isothermal storage conditions. $F(X_c)$ is the response function of the colour changing 'window' of the indicator. Points are average of measured response of twelve TTI units at each temperature. Lines are the regression fit lines for $F(X_c)$ v. time. (b) Arrhenius plot of the response rate of TTI Type B – Model M. $T_{ref} = 273$ K.

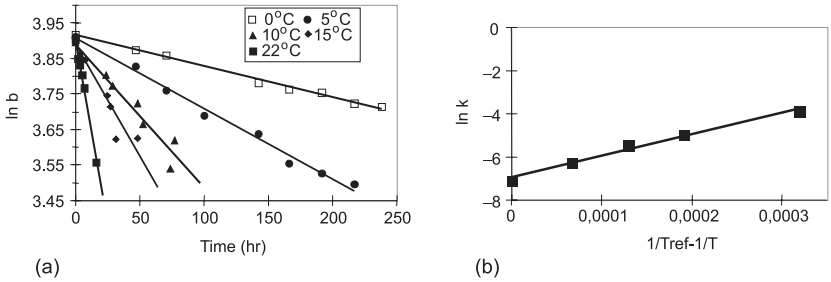


Fig. 19.7 (a) Response of a TTI Type C at different isothermal storage conditions. $F(X) = \ln(b_o/b)$ is the response function of the indicator. Points are average of measured response of ten TTI units at each temperature. Lines are the regression fit lines for $F(X)$ v. time. (b) Arrhenius plot of the response rate of TTI Type C. $T_{ref} = 273$ K.

behaviour of the TTI have to be validated under dynamic, i.e. variable temperature conditions. Such validation for selected TTI was conducted and was reported satisfactory.[29, 30]

The developed principles give a potential user the ability to develop an application scheme specific to a product and to select the most appropriate TTI without the need of extensive side by side testing of the product and the indicator. The storage effective temperature (T_{eff}) is the ultimate information obtained from the TTI response and the reliability of a TTI is directly linked to the error in T_{eff}. The error in the T_{eff} calculation is due to three sources:

1. The variability in the value of the response function, $F(X)$, between indicators of the same model, at same time and temperature. It can be expressed as an average coefficient of variation for a large number of different conditions. For previously studied TTI models it ranged from 1 to 10%.
2. The uncertainty in the Arrhenius equation, modelling the TTI temperature behaviour, statistically expressed by the confidence limits of the regression values of E_A and k_I. Values of ±5% to 14% for E_A from 41 to 141 kJ/mol were obtained. These first two sources of error reflect the quality of design and manufacturing of the TTI and the reliability of the kinetic modelling of the TTI temperature behaviour.
3. The third error in T_{eff} is due to the difference in activation energies between the food and the TTI. This error is systematic and not random. In the developed application scheme there is an important underlying assumption;

Table 19.3 Kinetic parameters for the three TTI types

	TTI A	TTI B -C	TTI B -M	TTI B -S	TTI B - L	TTI C
$F(X)$	X^2	$[\ln(1/1-X)]^{1/2}$	$[\ln(1/1-X)]^{1/2}$	$[\ln(1/1-X)]^{1/2}$	$[\ln(1/1-X)]^{1/2}$	$\ln(X_0/X)$
E_{A_i}(kJ/mol)	112	48	69	102	160	83

that the effective temperature of the food is equal to the effective temperature of the TTI for a given temperature distribution. This is only true when the activation energies of the food and the TTI are equal or in the trivial case when the temperature is constant throughout the cycle. Calculations with assumed variable temperatures showed that if the activation energies of the food and the TTI differ by less than 40 kJ/mol, i.e. $|E_{\mathrm{ATTI}} - E_{\mathrm{Afood}}| < 40$ kJ/mol, then in general the two T_{eff} differ by 0.4 to 1.8ºC. An error of 1ºC in the T_{eff} results in an error of the quality estimation in the range of 10−15%, which in many cases is acceptable. For example it was computed that $\Delta T_{\mathrm{eff}} = 1$ºC corresponds to 12 and 8.3% error in the value of quality for a zero- and 1st-order reaction with an activation energy E_A of 85 kJ/mol respectively. The recommended approach is to select a TTI with an E_A close to the foods, possibly within ±20 kJ/mol. Effect of E_A difference on the accuracy of quality estimation will be demonstrated in the next section. Furthermore, in section 19.6 it is discussed that even TTIs with larger E_A differences from the food could be useful for stock rotation applications.

The importance of the prerequisite systematic knowledge of the shelf-life loss behaviour of the food, expressed quantitatively in, as accurate as possible kinetic terms, has been pointed out. Quality loss of chilled foods is often due, directly or indirectly, to microbial spoilage. In this context the significant progress that is being achieved in the area of predictive microbiology can lead to effective shelf-life modelling of chilled foods and contribute to increased reliability in the use of TTIs as monitors of the chill chain.[31] What might be seen as a hindrance to a direct application of the accumulated kinetic information, to the above developed application scheme for TTIs, is the often used different kinetic terminology and quantitation approach of the temperature effect. Belehradek or square root type expressions for modelling the temperature dependence of the microbial growth rate has most often been used:[27]

$$\sqrt{k} = b(T - T_0) \tag{19.11}$$

The shown square root function, where b the slope and T_0 is a nominal minimum temperature, is the simplest form, applicable to the suboptimal conditions of microbial growth of the chill chain. Most available TTIs show an Arrhenius behaviour and are expressed in E_A terms. The developed application scheme based on the T_{eff} concept could equally well be used with the temperature dependence of both the food and the TTI expressed with an equation other than the Arrhenius. If however the two systems are expressed by different equations, the T_{eff} calculated from the TTI response based on the Arrhenius characteristics will introduce an additional error when used to estimate the quality of the food from the square root equation. A practical solution in these cases is to calculate, based on the Belehradek kinetics, the corresponding activation energy of the food spoilage phenomenon at the temperature range of interest, select the available TTI with the closest E_A and use the described application scheme based on Arrhenius kinetics.

19.5 Shelf-life monitoring in distribution

TTI can be used to monitor the temperature exposure of food products during distribution, from production up to the time they are displayed at the supermarket. Attached to individual cases or pallets they give a measure of the preceding temperature conditions at selected control points. The obtained information is used for continuous, overall monitoring of the distribution system, leading to recognition and correction of the consistently problematic links. Furthermore, it allows targeting of responsibility and serves as a proof of compliance to contractual requirements by the producer and distributor. It can guarantee that a properly handled product was delivered to the retailer, thus eliminating the possibility of unsubstantiated rejection claims by the latter. The presence of the TTI itself would probably improve handling, serving as an incentive and reminder to the distribution employees throughout the chain, of the importance of proper temperature storage.

The same TTI can be used as end point indicators readable by the consumer and attached to individual products. Tests using continuous instrumental readings to define the end point under constant and variable temperatures showed that the end points could be reliably and accurately recognised visually by panellists.[32] However, for a successful application, there is a much stricter requirement that TTI response matches the behaviour of the food, since there is no correlation algorithm used but a single visual end point that should indicate closely the end of shelf-life at any temperature. To achieve this the TTI end point should coincide with the end of shelf-life at one reference temperature and the activation energy should differ by less than 10kJ/mol from that of the food. The TTI attached to individually packaged products, can serve as dynamic or active shelf-life labelling instead of, or in conjunction with, open date labelling. The TTI assure the consumers that the products were properly handled and indicate remaining shelf-life. A consumer survey showed that consumers were very receptive to the idea of using these TTI on dairy products along with the date code.[33] Thus, use of TTI can additionally be an effective marketing tool. As such TTI Type A has been used by the Cub Foods Supermarkets in USA and TTI Type C by the Monoprix chain in France and the Continent stores in Spain.

A number of experimental studies that aimed to establish correlations between the response of specific TTI and quality characteristics of specific products have been reported. They employed side by side storage of indicators attached to the tested foods at different temperatures, plotting the response of the TTI v. time and the values of selected quality parameters of the foods v. time and testing the statistical significance of the TTI response correlation to the quality parameters. Foods correlated to TTI include pasteurised whole milk,[34–36] ice cream,[37] frozen hamburger,[38] chilled cod fillets,[39] refrigerated ready to eat salads,[40] frozen bologna,[41] UHT milk,[42] refrigerated orange juice,[43] pasteurised cream,[36] cottage cheese,[36, 44] frozen strawberries,[45] chilled lettuce and tomatoes,[46] and chilled fresh salmon.[47] This type of studies offers useful information, but do not involve any modelling of the TTI response as a function

of time and temperature and thus are applicable only for the specific foods and the conditions that were used. Extrapolation to other similar foods or quality loss reactions, or even use of the correlation equations for the same foods at other temperatures or for fluctuating conditions is not accurate.

The kinetic approach outlined allows the potential user to develop an application scheme specific to a product and to select the most appropriate TTI without the need of extensive side by side testing of the product and the indicator. This approach emphasises the importance of reliable shelf-life modelling of the food to be monitored. Shelf-life models must be obtained with an appropriate selection and measurement of effective quality indices and based on efficient experimental design at isothermal conditions covering the range of interest. The applicability of these models should be further validated at fluctuating, non isothermal conditions representative of the real conditions in the distribution chain. Similar kinetic models must be developed and validated for the response of the suitable TTI. Such a TTI should have a response rate with a temperature dependence, i.e. activation energy E_{A_I}, in the range of the E_A of the quality deterioration rate of the food. Also the total response time of the TTI should be at least as long as the shelf-life of the food at a chosen reference temperature. TTI response kinetics should be provided and guaranteed by the TTI manufacturer as specifications of each TTI model they supply.

The above concepts were thoroughly applied in studying the suitability of TTI in monitoring the seafood chill chain within the FAIR-CT96-1090 research project funded by the European Commission entitled 'Development, Modelling and Application of Time-Temperature Integrators to monitor Chilled Fish Quality'. Shelf-life of different fresh and minimally processed fish products was systematically studied and modelled. The fish chill chain, noted for substantial losses by spoilage is due to benefit significantly from effective monitoring and controlling of storage conditions. Temperature largely determines the rate of microbial activity, the main cause of spoilage of fresh and minimally processed fish products, thus being the determining parameter of shelf-life under Good Hygiene Practices. Shelf-life study requires establishing a time correlation between measured chemical/biochemical changes, microbiological activity and sensory value for the conditions of interest. Since each type of fish product, depending on intrinsic and extrinsic factors, has its own specific spoilage microflora investigation of the spoilage domain provides the fundamentals for understanding the spoilage phenomenon and for reliable shelf-life predictive modelling.[48, 49] Models of sensory quality and growth of spoilage microflora were developed and validated in dynamic temperature conditions for a variety of different fish. Kinetic study of alternative TTIs at isothermal and non isothermal conditions was conducted and their use as fish quality monitors was assessed. In this context the natural microflora of different Mediterranean fish of commercial interest such as boque, seabass, seabream, red mullet was studied and growth of the specific spoilage bacteria *Pseudomonas* spp. and *Shewanella putrefaciens* was modelled and correlated to organoleptic shelf-life.[29, 50, 51] Arrhenius and square root functions were used to model temperature dependence of maximum

growth rates. For example experimental data for growth of the different measured constituents of the boque natural microflora showed that at all temperatures, growth of Pseudomonads and *Shewanella putrefaciens*, followed closely the decrease of average sensory score of the cooked fish. End of shelf-life coincided with an average level of 10^7 for these two bacteria from 0 to 15ºC. At 0ºC it was determined at 174 hr. The Arrhenius temperature dependence of the rate of sensory degradation and Pseudomonads and *Shewanella putrefaciens* exponential growth rate was determined in terms of activation energy (E_A) as 86.6, 81.6 and 82.7 kJ/mol respectively.

Based on the above kinetic data the effect of the difference in the activation energies of TTI response (from Table 19.3) and spoilage rate of the monitored fish on the shelf-life predictive ability of the TTI can be assessed. The actual effective temperature (based on growth kinetics) of variable temperature profiles is compared to the one calculated from the response of the TTI. This was also translated based on the application scheme of Fig. 19.4 into remaining shelf-life at chilled conditions of 0ºC.

The total shelf-life at 0ºC is 174 hr based on the Pseudomonads growth with $N_0 = 1000$ and $N_{\max} = 10^7$. This coincides with the sensorial shelf-life. Setting these limits allows the estimation of remaining shelf-life at 0ºC after the 'abusive' storage conditions of the first 24 hr. In Table 19.4, T_{eff} for the fish, after exposure for 24 hours at the variable temperature profiles (shown in Fig. 19.8), is given.

It can be seen that for the first temperature profile, TD_1, indicator TTI B-Model C that has an activation energy more than 40 kJ/mol different to the fish spoilage gives a T_{eff} error of more than 1ºC. This results in a prediction of remaining shelf-life of 74 hr respectively, compared to the 90 hr of actual remaining shelf-life. T_{eff} and remaining shelf-life from TTI B-Model M and TTI C-Model A6, 92 and 91 hr respectively, are very close to the actual. It should be noted, however, that even the erroneous estimations from the TTIs with different activation energies are in practice much better than the 150 hr that would be presumed for shelf-life if no indication of improper storage was available. For the second temperature profile, TD_2, predictions from all TTIs are practically sufficient. This illustrates the fact that the error depends on the actual

Table 19.4 Effective temperature and remaining shelf-life (t_r) of boque at chilled conditions of 0ºC, for variable storage temperatures (TD1, TD2) during the initial 24 hr estimated by different TTI

	TD1		TD2	
	T_{eff} (ºC)	t_r (hr)	T_{eff} (ºC)	t_r (hr)
'ACTUAL'	8.93	90	9.8	81
TTI B-Model C	10.50	74	10.0	79
TTI B-Model M	8.74	92	9.8	82
TTI C-Model A6	8.95	91	9.8	81

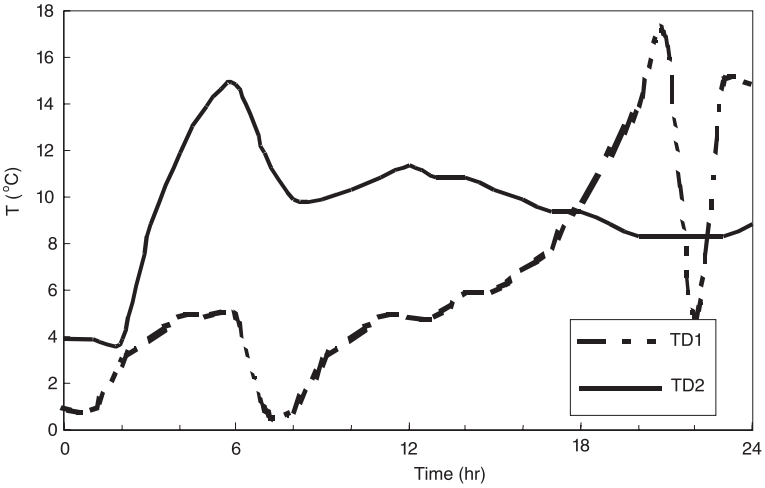


Fig. 19.8 Variable temperature scenarios of storage of boque during 24 hours.

temperature distribution. TD_1 and TD_2 are qualitatively different in that the first represents a profile with more abrupt changes than the second. The problem is that the temperature profile in a real situation is unknown. It is therefore advisable to select the TTI that, in addition to other requirements, has an activation energy close to the one of the quality loss rate of the food. Alternatively, response of two or three TTIs (i.e. a multiple TTI) with different E_{A_i}s could provide a corrected estimate of T_{eff} giving a reliable estimate of the food quality even when these E_{A_i}s differ substantially from the E_A of the food.[52]

The example shows the potential and demonstrates the methodology of monitoring the chill chain based on a continuous scale TTI response, translatable to the effective temperature history as described above. This methodology can be applied to chilled products other than fish if appropriate quality loss models are available. Especially long shelf-life chilled foods can benefit from the ability to monitor their temperature history by the introduction of a TTI based distribution control and stock rotation system. Such a system will be described and evaluated in the next section.

Frozen foods can also be monitored based on the same approach. TTI Type B and TTI Type A have been tested and modelled at temperatures in the range of -1 to -30ºC.[53] TTI Type A can respond above a temperature at which diffusion commences. This temperature can also be set, based on the type of polymer materials used and their glass transitions. Certain caveats should however be taken into account when TTIs are applied to frozen foods. These are related to the applicability of Arrhenius models developed under isothermal conditions to the prediction of the effect of dynamic, fluctuating temperature storage. The

response of tested TTIs does show Arrhenius behaviour also at the subfreezing range. Furthermore, the response models can reliably be used at non isothermal conditions. However, foods can seriously deviate from such behaviour. Deviations from Arrhenius type temperature dependence can be due to freeze concentration effects, recrystallisation dependent quality deterioration and glass transition phenomena.[26] For products like frozen vegetables, assuming that thawing is avoided (which can be verified with the aid of a CTI or a CTTI), Arrhenius behaviour in the range of -1 to -30ºC has been modelled and validated for fluctuating conditions.[53] TTI monitoring can effectively be applied in these cases. In cases such as ice cream or other frozen desserts where recrystallisation phenomena seriously affect, if not determine, the product quality, a cumulative effective temperature (obtained by a single response TTI) might not be sufficient to accurately predict the quality loss.

19.6 Optimised distribution and stock rotation system

The information provided by the TTI, translated to remaining shelf-life at any point of the chill chain can be used to optimise distribution control and apply a stock rotation system. Such an inventory management and stock rotation tool at the retail level was initially proposed by Labuza and Taoukis (1990).[54] The approach currently used is the First In First Out (FIFO) system according to which, products received first and/or with the closest expiration date on the label are shipped, displayed and sold first. This approach aims in establishing a 'steady state' with all products being sold at the same quality level. The assumption is that all products have gone through uniform handling, thus quality is basically a function of time. The use of the indicators can help establish a system that does not depend on this unrealistic assumption. The objective will again be the reaching of a 'steady state' situation with the least remaining shelf-life products being sold first. This approach is coded LSFO (Least Shelf-life First Out). The LSFO system would reduce the number of rejected products and largely eliminate consumer dissatisfaction since the fraction of product with unacceptable quality at consumption time can be minimised. The development of LSFO system is based on validated shelf-life modelling of the controlled food product, specification of the initial value of the quality index, A_0, and the value A_s at the limit of acceptability (end of shelf-life), and temperature monitoring in the chill chain with TTI. The above elements form the program core of an integrated software that allows the calculation of the actual remaining shelf-life of individual product units (e.g. small pallets, $5-10$ kg boxes or even single product units) at strategic control points of the chill chain. Based on the distribution of the remaining shelf-life, decisions can be made for optimal handling, shipping destination and stock rotation, aimed at obtaining a narrow distribution of quality at the point of consumption. The diagram of the decision making routine is illustrated in Fig. 19.9. Also, the mechanism of decision making at a hypothetical control point of the chill chain is shown. For example, at a certain point, e.g. the

supermarket main storage, product half of a shipment is forwarded to the retail display, immediately, the other half the next day. The split could be at random according to conventional FIFO practice or it can be based on the actual individual product quality and LSFO. For all units the response of the TTI, cumulatively expressing the temperature exposure of the product, is put in either electronically as a signal of a suitable optical reader or keyed in manually based on visual readings. This information directly fed into a portable unit with the LSFO progamme, is translated to quality status, A_t, based on the kinetics of the used TTI, which integrates the time-temperature history of each product into an effective temperature value, T_{eff}, and the shelf-life model of the product. Having estimated A_t for all the n product units, the actual quality distribution for the products at the decision point is constructed. Based on the quality of each product unit relative to this distribution, decisions about its further handling are made.

For the scenario illustrated in Fig. 19.9, products B with less remaining shelf-life, i.e. higher A_t, will be displayed first at the retail display cabinets of the supermarket and will therefore be consumed sooner whereas products with longer remaining shelf-life (lower A_t) will be displayed later. The decision process can involve more options with regard to e.g. handling methods, shipping means or destinations, stock rotation timing and planning. Points of the chill chain where actions are taken with regard to handling, transportation, distribution and stocking of products can be designated and used as decision points of the LSFO system.

In order to evaluate the results of the application of the LSFO system and quantitatively prove its effectiveness a Monte Carlo simulation can be applied, with data and information provided by surveys on the conditions of the distribution chain. It is based on the generation of hypothetical 'scenarios'. Values of the controlling parameter, temperature, are treated as probability distributions, which represent uncertainty (lack of sufficient knowledge, see also Chapter 7) or the commonly encountered variation in the parameter. The procedure, repeated many times, requires the random selection of a value from each of the probability distributions assigned for the input parameters, in order to calculate a mathematical solution, defined by the shelf-life model used. At each iteration, a value is drawn from the defined distribution, calculations are performed and the results are stored. Eventually, the analysis provides a frequency distribution for the output of interest (quality status and remaining shelf-life), that has taken into account the probability distribution of temperature conditions, instead of using a single-point estimate (see also Chapter 5).

The results for the simulated application of LSFO in the cases of two long shelf-life chilled products are presented. These products have both a shelf-life of three months at 4ºC and their quality loss rate shows a low and high temperature dependence (low and high E_A respectively).

Russian salad, a chilled product widely consumed in Greece, microbiologically stable through application of hurdle technology principles with a shelf-life of three months was used as a case study.[55] For this microbiologically stable, complex food, modelling of shelf-life was based on overall organoleptic

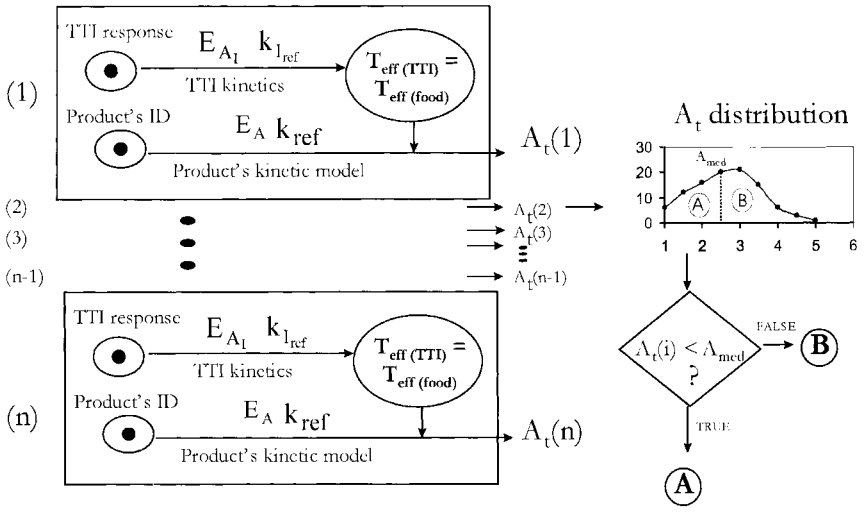


Fig. 19.9 Logical diagram of the decision making routine of LSFO system at important control points of the distribution chain. Quality at time t (A_t) is computed for all n product units. The computation is based on the response of the TTI, translated to the effective storage temperature (T_{eff}) of the product. The distribution function of quality is constructed and decision for the further handling of each unit is taken based on its value within this function.

deterioration and development of rancidity. Use of Weibull Hazard Analysis facilitated shelf-life determination and modelling of sensory evaluation data. The activation energy of shelf-life loss was estimated at 31.5 kJ/mol. For a realistic estimate of the storage temperature conditions at the different stages, data of chilled product temperatures previously collected at the commercial level and from a survey of home refrigerators was used. The temperature condition distributions are illustrated in Fig. 19.10.

To demonstrate the effectiveness of the LSFO approach as compared to FIFO, a 60-day cycle, from production to consumption, was used. This consisted of three stages: Stage 1, 30 days at local distribution centres; stage 2, 15 days at the supermarket storage; stage 3, 15 days at the domestic refrigerator. Based on a Monte Carlo simulation approach, 2000 temperature scenarios were run, using a program code written in FORTRAN 77. The temperatures used were obtained at random from the distributions of Fig. 19.10 (distribution 10a for stages 1 and 2 and distribution 10b for stage 3). The results of this simulation are illustrated in Fig. 19.11 which shows the probability for the product to be consumed at a certain quality level, expressed as Remaining Shelf-Life (SLR). The FIFO approach shows a significant portion of products (8%) consumed with quality

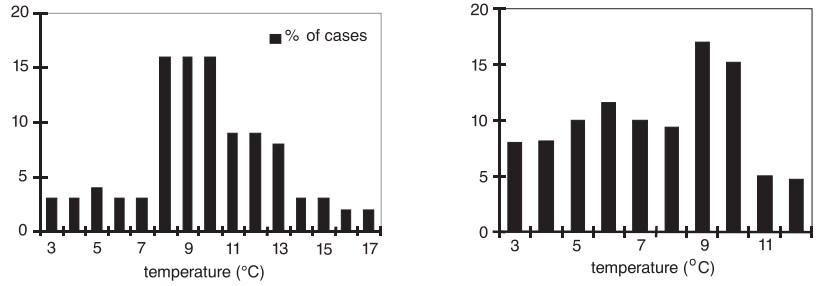


Fig. 19.10 Left: Temperature distribution in commercial chilled storage. (Measurements in 150 supermarkets in the metropolitan area of Athens). Right: Temperature distribution in domestic refrigerators. (Based on measurements in 40 households). (Adopted from Taoukis *et al.*, 1998.[55])

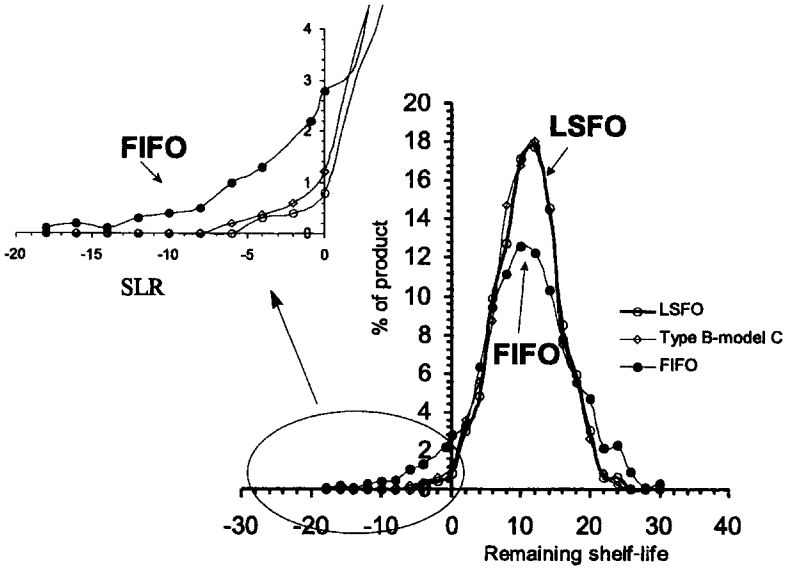


Fig. 19.11 Distribution of quality of Russian salad products after 60 days distribution, retail and domestic storage. For each point the percentage of the products that have a remaining shelf-life in the range of ± 2.5 days of the abscissa value can be read on the vertical axis. The line with solid circles corresponds to the FIFO and with open circles to the LSFO system based on actual temperature monitoring or a TTI with $E_{A_I} = E_A$. Open diamonds line is the LSFO line based on the TTI, Type B (practically coincides with the actual LSFO line).

lower than the one used to set end of shelf-life (expressed as a negative remaining shelf-life). Using the LSFO approach, products in the 2nd stage are advanced to retail cabinets for sale every five days based not on FIFO but on the response of the attached TTI (Type B- Models C and M) showing which products should be

advanced first. This system leads to a narrower range of quality at consumption time (less than 1% unacceptable products) and can practically eliminate the 'tails', i.e. the portions of products consumed at extreme qualities. Thus a situation where products are consumed at a uniform quality, with no 'below standard' products can be obtained. As tools for a comparative selection of least shelf-life products at control points both TTIs can be used effectively.

The same system was applied to distribution and stock rotation of shrink-packed marinated seafood products (marinated fish fillets, shrimp, squid, octopus) with a target shelf-life of 3 months at 4ºC. Shelf-life temperature dependence of such products, based on sensory evaluation, varied but was in the high range of activation energies. An E_A value of 110 kJ/mol, a distribution cycle of 35 days, consisting of the same three stages and temperature distributions as above (10 days at stage 1, 15 days at stage 2 and 10 days at stage 3) and TTIs Type B (Models C, S and L with activation energies 48.3, 102 and 160 kJ/mol respectively) were used in the Monte Carlo simulation to assess application of LSFO. Results are shown in Fig. 19.12. It can be seen that in products with high activation energies the distribution of quality at consumption time is much wider as temperature variation affects more intensely the rates of quality loss. Application of the LSFO system reduces the percentage of unacceptable products to less than 5% compared to 22% with the FIFO approach. It can also be seen that even TTIs that differ from the food in terms of E_A approximately 50 kJ/mol can serve as tools for the relative comparison of the shelf-life of the products at the control points of the LSFO system.

A further development to LSFO was an intelligent system proposed and evaluated coded as Shelf Life Decision System (SLDS).[56, 57] SLDS integrates predictive kinetic models of food spoilage, data on initial quality from rapid techniques and the capacity to continuously monitor temperature history of the food product with Time Temperature Integrators (TTI), into an effective chill chain management tool that leads to an improved narrow distribution of quality at consumption time, effectively reducing the probability of products consumed past shelf-life end. For most processed food products, 'zero time' post processing parameters, including a target range of initial microbial load, can be fixed and achieved by proper design and control of the processing conditions. This is the working assumption of LSFO. However, initial microflora in fresh foods such as fish or meat can fluctuate significantly, depending on a number of extrinsic factors at slaughter or catch, and the following handling and processing.[58–60] SLDS takes not only the history of the product in the distribution chain into account but also this variability of initial contamination. Rapid methods of microbial enumeration can be employed to provide such information as input. The Shelf Life Decision System can incorporate in the calculation of the quality distribution at each control point, other parameters of variability of the product's quality. Such parameters can be variation of initial pH, water activity, packaging gases composition, provided the used shelf-life predictive models can account for the effect of these parameters on the microbiological and chemical reactions responsible for the loss of quality.

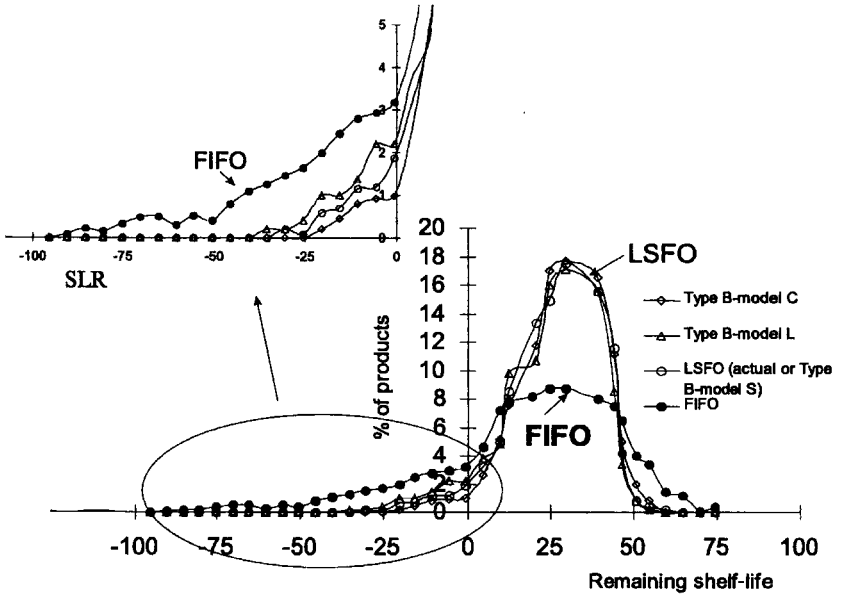


Fig. 19.12 Distribution of quality of shrink-packed products after 30 days distribution, retail and domestic storage. For each point the percentage of the products that have a remaining shelf-life in the range of ±5 days of the abscissa value can be read on the vertical axis. The line with solid circles corresponds to the FIFO and with open circles to the LSFO system based on actual temperature monitoring or the TTI B- Model S with $E_{A_i} \approx E_A$. Open diamonds and open triangles lines are the LSFO lines based on the TTI, Type B–Models L and C.

19.7 Future developments

The described systematic approach points out the future prospects of Time Temperature Integrators. TTIs will inevitably find wider application as tools to monitor and control distribution as their concept, their potential and their limitations are thoroughly understood by the food science and industry. Progress on both ends of the equation, i.e. on the variety, reliability and flexibility of TTI and stricter, quantitative shelf-life characterisation of food products will allow successful application of the described chill chain optimisation tools such as the LSFO and the intelligent Shelf Life Decision System. Research progress in the area of quality kinetic modelling and predictive microbiology, will show how the TTI concept can be meaningfully and safely expanded to contribute in the quality assurance of more foods. User friendly softwares will integrate support systems designed to predict effects of processing parameters and product design to food product quality.[61] Such systems could provide the data input on initial product quality distribution, based on processing and raw material parameters, that is needed for the SLDS calculations at the control points of the chill chain on which the TTI based management of the products occurs.

19.8 References

1. ISO 9001: 1994, 'Quality systems – Model for quality assurance in design, development, production, installation and servicing', *International Standard*, 1994.
2. ISO/FDIS 9001: 2000, 'Quality management systems – Requirements', *International Standard*, 2000.
3. 93/43/EEC, 'Hygiene of foodstuffs', *General Directive*, 1993.
4. Commission Paper III/5383/95-EN, 'Implementation of the hygiene of foodstuffs Directive 93/43/EEC', 1995.
5. Codex Committee on Food Hygiene, 'Revised International Code of Practice – General Principles of Food Hygiene', *WHO/FNU/97*, 1997.
6. US Federal Register, 'Pathogen Reduction; Hazard analysis and Critical Control Point (HACCP) Systems; Final Rule',1996 Vol. **61**, No. 144, July 25, 1996.
7. TAOUKIS P S, LABUZA T P, 'Applicability of Time Temperature Indicators as shelf-life monitors of food products', *J. Food Sci.*, 1989 **54** 783–8.
8. SCHOEN H M, BYRNE C H, 'Defrost indicators: many designs have been patented yet there is no ideal indicator', *Food Technol.*, 1972 **26** (10) 46–50.
9. BYRNE C H, 'Temperature indicators – the state of the art', *Food Technology* , 1976 **30** (6) 66–8.
10. SINGH R P, WELLS J H, 'Keeping track of time and temperature: new improved indicators may become industry standards', *Meat Processing*, 1986 **25** (5) 41–2.
11. TAOUKIS P S, FU B, LABUZA T B, 'Time–Temperature Indicators.' *Food Technology*, 1991 **45** (10) 70–82.
12. MIDGLEY T, 'Telltale means' *U.S. Patent 1917048*, 1933.
13. TAOUKIS P S, *Time-temperature indicators as shelf life monitors of food products*. Ph.D. Thesis, University of Minnesota, St. Paul, 1989.
14. RENIER J J, MORIN W T, 'Time-temperature indicators' *Intl. Inst. Refrig. Bull. Annex*, 1962 **1** 425–35.
15. GUADAGNI D G, 'Time-Temperature Indicator: a laboratory evaluation', *Frosted Food Field*, 1963 **36** (4) 42–4.
16. OTA *Open Shelf Life Dating of Food*, Office of Technology Assessment, Congress of the U.S., Library of Congress Cat. No. 79-600128, 1979.
17. HU K H, 'Time-Temperature indicating system ''writes'' status of product shelf-life', *Food Technology*, 1972 **26** (8) 56–62.
18. KILLORAN J, 'A time-temperature indicating system for foods stored in the nonfrozen state' *1976 Activities Report, U.S. Army Natick Labs.*, 1976 137–42.
19. KRAMER A, FARQUHAR J W, 'Testing of time-temperature indicating and defrost devices', *Food Technology*, 1976 **32** (2) 50–6.
20. MANSKE W J, 'The application of controlled fluid migration to temperature limit and time temperature integrators', *Intl. Inst. Refrig. Commission C2 Preprints*, 16th Intl. Cong. Refrig., 1983 632–5.

21. TAOUKIS PS, LABUZA TP, FRANCIS RC, 'Time Temperature Indicators as Food Quality Monitors', *Food Packaging Technology – American Society of Testing and Materials*, 1991 ASTM STP 1113 51–63.
22. FIELDS SC, PRUSIK T, 'Time-Temperature monitoring using solid-state chemical indicators', *Intl. Inst. Refrig. Commission C2 Preprints*, 16th Intl. Cong. Refrig., 1983 636–40.
23. PATEL GN, YANG N, 'Polydiacetylenes: An ideal system for teaching polymer science', *J. Chem. Ed.*, 1983 **60** (3) 181–5.
24. BS 7908: 1999 'Packaging – Temperature and time-temperature indicators – Performance specification and reference testing.' British Standard, 1999.
25. TAOUKIS PS, LABUZA TP, 'Summary: Integrative Concepts (Shelf-life testing and modelling).' In: *Food Chemistry, 3rd Edition*, O. Fennema (ed.), Marcel Dekker, New York, Ch. 17 1013–42, 1996.
26. TAOUKIS PS, LABUZA TP, SAGUY Y, 'Kinetics of Food Deterioration and Shelf-Life prediction.' In: *The Handbook of Food Engineering Practice*, K. J. Valentas, E. Rotstein, R. P. Singh (eds.), CRC Press, Boca Raton, FL, Ch. 10 361–403, 1997.
27. MCMEEKIN TA, OLLEY JN, ROSS T, RATKOWSKY DA, *Predictive Microbiology: Theory and Application*, John Wiley & Sons Inc., London, 1993.
28. TAOUKIS PS, LABUZA TP, 'Chemical Time – Temperature Integrators as quality monitors in the chill chain'. In: *Predictive Microbiology Applied to Chilled Food Preservation*. C.M. Bourgeois, and T.A. Roberts (eds.) Refrigeration Science and Technology Proceedings Series, International Institute of Refrigeration (IIR), Paris, France, 291–300, 1999.
29. TAOUKIS PS, KOUTSOUMANIS K, NYCHAS G-JE, 'Use of Time Temperature Integrators and Predictive Modeling for Shelf Life Control of Chilled Fish under Dynamic Storage Conditions.' *Int. J. Food Microbiol.*, 1999, **53** 21–31.
30. TAOUKIS PS, LABUZA TP, 'Reliability of Time Temperature Indicators as food quality monitors under non isothermal conditions.' *J. Food Sci.*, 1989 **54** 789–92.
31. LABUZA TP, TAOUKIS PS, 'Use of Time Temperature Integrators and Predictive Microbiology', 2nd International Conference on Predictive Microbiology, Hobart, Tasmania, Australia, 1996.
32. SHERLOCK M, FU B, TAOUKIS PS, LABUZA TP, 'A systematic evaluation of time temperature indicators for use as consumer tags', *J. Food Protection*, 1991 **54** (11) 885–9.
33. SHERLOCK M, LABUZA TP, 'Consumers perception of consumer type time-temperature indicators' *J. Food Distrib. Res.*, 1991.
34. MISTRY VV, KOSIKOWSKI FV, 'Use of time-temperature indicators as quality control devices for market milk.' *J. Food Protection*, 1983 **46** (1) 52–7.
35. GRISIUS R, WELLS JH, BARRET EL, SINGH RP, 'Correlation of time-temperature indicator response with microbial growth in pasteurized milk.' *J. Food Proc. Preserv.*, 1987 **11** 309–24.

36. CHEN J H, ZALL R R, 'Packaged milk, cream and cottage cheese can be monitored for freshness using polymer indicator labels.' *Dairy Food Sanitation*, 1987 **7** (8) 402–5.
37. DOLAN K D, SINGH R P, WELLS J H, 'Evaluation of time-temperature related quality changes in ice cream during storage.' *J. Food Proc. Preserv.*, 1985 **9** 253–71.
38. SINGH R P, WELLS J H, 'Use of time-temperature indicators to monitor quality of frozen hamburger.' *Food Technology*, 1985 **39** (12) 42–5.
39. TINKER J H, SLAVIN J W, LEARSON R J, EMPOLA V G, 'Evaluation of automated time-temperature monitoring system in measuring the freshness of chilled fish' *IIF-IIR Commissions C2, D3 4* 1985 286–90.
40. CAMBELL L A, *Use of a time-temperature indicator in monitoring quality of refrigerated salads.* M. S. Thesis. Michigan State University, 1986.
41. SINGH R P, WELLS J H, 'Keeping track of time and temperature: new improved indicators may become industry standards.' 1986 *Meat Processing*, **25** (5), 41–2, 46–7.
42. ZALL R R, CHEN J, FIELDS S C, 'Evaluation of automated time-temperature monitoring systems in measuring freshness of UHT milk.' *Dairy and Food Sanitation*, 1986 **6** (7) 285–90.
43. CHEN J H, ZALL R R, 'Refrigerated Orange Juice can be monitored for freshness using a polymer indicator label.' *Dairy Food Sanitation*, 1987 **7** 280–3.
44. SHELLHAMMER T H, SINGH R P, 'Monitoring chemical and microbial changes of cottage cheese using a full history time-temperature indicator.' *J. Food Sci.*, 1991 **56** 402–5.
45. SINGH R P, WELLS J H, 'Monitoring quality changes in stored frozen strawberries with time-temperature indicators' *Int. J. Refrig.* 1987 **10** 296–9.
46. WELLS J H, SINGH R P, 'Application of time-temperature indicators in monitoring changes in quality attributes of perishable and semi-perishable foods.' *J. Food Sci.*, 1988 **53** 148–52.
47. RONNOW P, SIMPSON R, OTWELL S, 'The use of an enzymatic TTI to monitor time-temperature exposure in distribution of chilled seafood' In: *Predictive Microbiology Applied to Chilled Food Preservation*. C.M. Bourgeois, and T.A. Roberts (eds.) Refrigeration Science and Technology Proceedings Series, International Institute of Refrigeration (IIR), Paris, France, 308–15, 1999.
48. DALGAARD P, 'Modelling of microbial activity and prediction of shelf life for packed fresh fish', *Int. J. of Food Microbiol*. 1995 **26** 305–17.
49. DALGAARD P, HUSS H H, 'Mathematical modelling used for evaluation and prediction of microbial fish spoilage'. In: *Seafood safety, processing and biotechnology*, Kramer, D., Shahidi, F. and Jones, Y. (eds.) Technomic Pub. Co. Inc, Lancaster, Canada, 73–89, 1995.
50. KOUTSOUMANIS K, NYCHAS G-J E, 'Application of a systematic experimental procedure to develop a microbial model for rapid fish shelf-life

procedure.' 2000 *Int. J. Food Microbiol.*, 2000 **60** 171–84.

51. KOUTSOUMANIS K, TAOUKIS P S, DROSINOS E, NYCHAS G-J E, 'Applicability of an Arrhenius model for the combined effect of temperature and CO_2 packaging on the spoilage microflora of fish.' *Appl. Env. Microbiol.*, 2000 **66** 3528–34.
52. STOFOROS N G, TAOUKIS P S, 'A theoretical procedure for using multiple response time-temperature integrators for the design and evaluation of thermal processes.' *Food Control*. 1998 **9** (5) 279–87.
53. GIANNAKOUROU M C, SKIADOPOULOS A, POLYDERA A, TAOUKIS P S, 'Shelf-life modelling of frozen vegetables for quality optimization with Time-Temperature indicators.' *Eighth International Congress on Engineering and Food*, Puebla, Mexico, 2000.
54. LABUZA T P, TAOUKIS P S, 'The relationship between processing and shelf life', In: *Foods for the 90's*, Birch, G. G., Campbell-Platt, G., and Lindley, M. G. (eds), Developments Series, Elsevier Applied Science Publishers, London and New York, Ch. 6, 73–105, 2000.
55. TAOUKIS P S, BILI M, GIANNAKOUROU M, 'Application of shelf-life modeling of chilled salad products to a TTI based distribution and stock rotation system.' In: *Proceedings of the International Symposium on Applications of Modeling as an Innovative Technology in the Agri-Food-Chain*, Tijskens L.M.M.and Hertog M.L.A.T.M. (eds.), *Acta Horticulturae*, 1998 **476** 131–40.
56. GIANNAKOUROU M, KOUTSOUMANIS K, NYCHAS G-J E, TAOUKIS P S, 'Development and assessment of an intelligent Shelf Life Decision System for quality optimization of the food chill chain' *J. Food Protection*. In press.
57. KOUTSOUMANIS K, GIANNAKOUROU M C, TAOUKIS P S, NYCHAS G-J E, 'Application of Shelf Life Decision System (SLDS) to marine cultured fish quality' *Int. J. Food Micro.*. In press.
58. EISEL W G, LINTON R H, MURIANA P M, 'A survey of microbial levels for incoming raw beef and ground beef in a red meat processing plant.' *Food Micro.* 1997 **14** 273–82.
59. GILL C O, 'Microbiological contamination of meat during slaughter and butchering of cattle, sheep and pigs' In: *The microbiology of meat and poultry*, A. Davies and R. G. Board (eds.), Blackie Academic & Professional, London, 1998 118–57.
60. HUSS H H, 'Quality and quality changes in fresh fish.' FAO Fisheries Technical Paper. No. 348. FAO. Rome 1995.
61. WIJTZES T, VAN'T RIET K, HUIS IN'T VELDAND J H J, ZWIETERING M H, 'A decision support system for the prediction of microbial food safety and food quality.' *Int. J. Food Microbiol*. 1998 **42**: 79–90.

20

Modelling the management of distribution centres

R. A. C. M. Broekmeulen, Eindhoven University of Technology

20.1 Introduction

The physical distribution of products such as dairy, meat, vegetables and fruits is a complex process due to the perishable nature of these agricultural products. The mainly soil-bound production of the raw materials takes place at relatively small companies in a pattern that for the produced quantity and/or quality is often seasonal. The demand for perishables fluctuates from day-to-day and has in recent years diversified into products that distinguish themselves in quality and appearance. Besides fresh produce, this chapter also considers prepared and chilled products such as ready-to-eat meals and desserts, but not dried or deep frozen products. The variation and decay in quality and the relatively short keeping quality makes it difficult to balance supply and demand. In the supply chain of perishables, the distribution centre plays an important role as decoupling point, as defined by Hoekstra and Romme.[1] Attention in the distribution centre to maintain optimal conditions in relation to product characteristics, combined with a short throughput time, increases the keeping quality as well as the customer service.

The storage of perishables occurs at manufacturers, wholesalers, and at third-party logistic service providers. Examples are the storage facilities at the horticultural auctions and the fresh produce departments at retail chains. An air-conditioned storage accommodation and several cold storage rooms with insulation distinguish the layout of a distribution centre for perishables from a standard distribution centre, according to Sims.[2] The usual activities of a distribution centre, such as groupage, storage and repackaging are carried out at specific storage conditions. Value-adding activities such as conditioning or ripening of products often require special installations.

The specific layout and operations management of a distribution centre for perishables influence the amount of keeping quality loss and the efficiency of the operations. The management of a distribution centre for perishables has to deal with questions concerning operations management such as:

- What kind of storage and handling equipment is needed?
- How much storage space is required for the different products?
- Where should the products be located in the distribution centre and does that location depend on the time of year?
- What kind of storage conditions are needed in each location to reduce the keeping quality loss?
- What kind of storage policy should be applied?

The problem associated with answering these questions is relatively unstructured, since it is practically impossible to formulate a model that can answer all these questions at the same time. A hierarchical structure for the different, functionally-related decisions that must be made to solve the problem is proposed. The structure divides the decisions into nested or hierarchical levels. Each level answers different questions for management. The problem-solving task of the management of the distribution centre is facilitated by introducing a decision support system with which the management can iterate between the different decision levels. The models and methods that are introduced in this chapter have been implemented in a decision support system by Broekmeulen[3] and will help the user of the decision support system to generate and evaluate alternative decisions.

20.2 Characteristics of perishables

20.2.1 Definitions and properties

Fresh foodstuffs are usually from an agricultural origin, and consist of biological material. In the agribusiness, a product is for instance a cultivar of a vegetable with well-defined properties like colour, taste and overall appearance. The diversity of marketed fresh foodstuffs in type, quality, and lifetime is enormous and is still increasing due to changing consumer preferences, globalisation of markets and technological innovations. Only a limited number of foodstuffs can be treated as normal commodities because of their relatively uniform quality and long lifetime. Examples are cereals and coffee. Meffert[4] stresses that all other fresh foodstuffs have to be given special attention in the distribution chain from producer to consumer in order to deal with their specific properties such as ageing and breakdown. The assortment of a distribution centre is a list of the products that are handled or in stock during a specific period of time. A year can be divided in one or more planning periods. The supply of many perishables depends on the seasons. The numerous changes in the assortment during the year make the activities in a distribution centre non-repetitive.

Keeping quality

Fu & Labuza[5] define keeping quality or shelf-life as the period of time until which a product becomes unacceptable. According to Shewfelt,[6] the attribute that limits the product acceptance is the first attribute that becomes unacceptable, given the circumstances. The limiting attribute can be predefined. An example is firmness of tomatoes in the study of Polderdijk *et al.*[7] The keeping quality loss of food is based on biological processes which remain active after the harvest. Floros[8] and Labuza[9] mention a range of processes that influence keeping quality of agricultural products. The static keeping quality is the average number of days that a product is 'fit for use' if it is kept under the same storage conditions. Storage conditions which maximise the keeping quality result in minimal keeping quality loss.

The following items are examples of factors that influence the keeping quality; see also Ryall and Lipton.[10]

- *Initial keeping quality.* The initial quality of an agricultural product varies because of differences in the cultivars used, the production method, the season and the production region.
- *Storage time.* Short storage times reduce the keeping quality loss.
- *Temperature.* Each product has its own optimal temperature. Many agricultural products that now or in the past originated from tropical or subtropical climates are chilling-sensitive.
- *Relative humidity.* A low relative humidity accelerates the dry out and therefore the ageing of a product. On the other hand, a high relative humidity makes the product an ideal medium for all kinds of microorganisms.
- *Handling.* Some perishables are very susceptible to mechanical damage like shock and vibration. Suitable packaging can reduce the effects of shock and vibration during handling. Handling has, in general, a negative effect on keeping quality.
- *Ethylene.* The gas ethylene, a plant ripening hormone, plays a role in the interaction between products. High concentrations of ethylene accelerate the ripening and decay of a large number of products. According to Abeles *et al.*,[11] products that produce ethylene are susceptible to ethylene at the same time. The production of ethylene as well as the influence of this product interaction on the keeping quality are dependent on the temperature.
- *Odours.* Products like onions and garlic produce odours that are adsorbed by fruits such as melons.

A quality change group consists of products with the same quality change characteristics, such as optimal temperature and production of or susceptibility to ethylene. The quality change models used in this research describe for each quality change group the static keeping quality under specific storage conditions. Some of the factors mentioned can be controlled in the storage rooms of a distribution centre. Each storage room has a specific temperature that is a compromise between the optimal temperatures of the quality change groups that are assigned to that storage room.

The keeping quality loss in the distribution centre depends on the initial keeping quality when the product enters the distribution centre. The initial keeping quality just after the harvest or production changes in the distribution chain with the storage conditions and the duration of stay in the different stages in the distribution chain. Tijskens and Polderdijk[12] discuss these dynamic aspects of quality change models for vegetables and fruits and their importance for the simulation of the distribution chain of perishables. The effect of changes in the distribution on the keeping quality of the products can be predicted with simulation models of the distribution chain that incorporate dynamic quality change models.

The effects of odours and hormones on keeping quality can be avoided by spatially separating the products, by using an adequate packaging for the susceptible products or can be reduced by ventilation of the storage room. Milk is normally very susceptible to the odour of garlic, but milk in cartons does not pose any problems. High ventilation rates reduce the effects of odours and hormones by increasing the effective volume in a storage room but increase the energy requirements.

Packaging

Storage conditions such as relative humidity and product interactions are difficult to control in the storage rooms of a distribution centre. A suitable (transport) packaging has several advantages for maintaining keeping quality:

- regulation of the relative humidity
- shielding of the product against odours and hormones
- protection against contamination by microorganisms
- reduction of the effect of shocks and vibrations during handling.

The complete sealing of the packaging for a living product can lead to too low a concentration of oxygen due to the still active metabolism of the product. Low oxygen concentrations can lead to unwanted or even dangerous secondary metabolites during fermentation such as alcohol. In the absence of oxygen, toxin-producing microorganisms such as *Clostridium botulinum* can start to grow. More and more products are therefore distributed in sealed (transport) packaging inside which the atmospheric conditions are adjusted to the product (see Chapter 14 for modified atmosphere packaging or MAP). The relatively low value of the majority of fresh products and the reduced perception of freshness of pre-packed products still prevent all perishables from being distributed in sealed packaging. Reduced product loss makes pre-packing often beneficial for the environment compared with unpacked products.

Seasons

The effect of growing seasons is reflected in the assortment and the stock levels of perishables. On the demand side, this can be explained by the consumption pattern, such as Christmas for luxury vegetables and in summer for salads. Depending on the agricultural origin of perishables or their raw materials,

different harvest periods in the world and different production methods, such as in the field or in greenhouses, result in peaks in the supply of products. The seasonal pattern has a profound effect on the layout and management of distribution centres for perishables.

20.2.2 Perishables in a distribution centre

A distribution centre performs activities such as storage, conditioning, accumulation and consigning. These activities need resources such as buildings, equipment and personnel. Specific operating policies applied to a set of resources completes a warehousing system.

A distribution centre has two main functions: warehousing and distribution. Warehousing includes all activities concerned with storage and retrieval of products. The distribution function concentrates on the centre's groupage and shipment of customer orders. Groupage is the combination of one or more customer orders in a single consignment. With direct delivery from supplier to customer, the products skip the distribution centre. This option is interesting for large shipments of only a few products. In all other cases, the products enter the distribution centre. All products that are received in the distribution centre are subject to inspection and quality control.

The storage accommodation is the part of the distribution centre where products can be stored. The storage equipment such as racks and shelves creates in the storage accommodation locations and storage space where products can be stored. An important activity in a distribution centre for fruits and vegetables is the conditioning of products. Adequate storage conditions such as relative humidity and temperature, minimal handling and avoidance of product interactions such as odour and hormone transmission can reduce keeping quality loss. An area in the storage accommodation with specific storage conditions is called a zone. By order picking the articles are collected from the storage accommodation and moved to the staging area. In the staging area, customer orders for more than one item of the assortment or for a product that has to be collected from several locations are accumulated before shipment to the customer.

Picking personnel or pickers move the products in the distribution centre. Personnel need training to know what a specific article looks like, and whether the article picked for shipment is the same as is ordered by the customer. An administrative system such as a warehouse management system (WMS) can support the day-to-day operations by giving storage and retrieval advice to the pickers. A storage policy determines on the operational level where a product can be stored and a retrieval policy which product can be retrieved.

A correct and fast information flow between the actors is a requirement in the coordination of the distribution chain of perishables. Uniform article coding and barcodes are essential to achieve the necessary quality of the information and speed of the information flow. The large variation in product types, packaging and quality have made uniform article coding impractical for agricultural

products. The short lifetime and the fast handling of the articles hinder the separate introduction of barcodes in the distribution centre. The storage time is too short to give all incoming articles a detailed internal code or sticker. Due to the absence of a uniform article coding and the special conditions needed by perishables, all articles that are kept in stock need a fixed slot or location in a distribution centre for perishables. Personnel still have to check by visual inspection that the correct article is handled. Wrong quality or incorrect packaging frequently result in mistakes.

The type of handling equipment and the degree of mechanisation determine the differences in handling methods. The dimensions in the layout such as the maximum height of a rack and the minimum width of an aisle define the possible handling equipment and vice versa. The chosen handling equipment and the degree of mechanisation are two of the factors that influence the maximum handling capacity or throughput of the distribution centre and the productivity of the personnel. Modern equipment such as automatic carousels and automated sorters are often dedicated to the type of transport packaging and require specific storage equipment and high investment. The combination of racks, carts, and forklift trucks remains a flexible, low cost handling system with the disadvantage of relatively low throughput per picker. Because of the relatively low added value of perishables, the application of carts and forklift trucks will probably continue in the near future.

20.3 Maintaining keeping quality with a slot plan

A slot plan describes for each planning period where in the distribution centre and how much storage space must be reserved for each product or quality change group. Usually a planning period in a slot plan coincides with a season. With a slot plan that is optimised for the characteristics of perishables, the keeping quality of these perishables during storage is maintained, compared to conventional storage policies.

20.3.1 The value of a slot plan during operations

Slot planning is primarily used to optimise the efficiency of the order picking operations. The class-based storage policy, introduced by Hausman *et al.*,[13] concerns the assignment of incoming products to classes or sets of slots in order to reduce the mean travel time for storage and retrieval. The storage accommodation has a fixed number of classes with a specific storage space requirement and the products are assigned to a class based on the turnover rate. An incoming load is stored at an arbitrary open slot in the assigned class. When all the slots in the assigned class are occupied, the class with open slots closest to the assigned class is taken. In a distribution centre for perishables, a class often coincides with a zone. A random storage policy, where incoming products are assigned to the first open slot, has a better space utilisation compared to the

class-based storage policy, but a lower order picking productivity. The random storage policy does not take account of the specific storage conditions needed for perishables.

The influence of the daily requests for storage and retrieval during a whole year can be studied with a simulation model that assigns incoming loads of products to zones. The daily storage and retrievals follow the seasonal production and demand of the products. All requests are handled on a first-come-first-served basis. The following three storage policies are studied with the simulation model.

1. *Free zone policy*. The product is stored in the first zone with enough unoccupied storage space to satisfy the request.
2. *Temperature zone policy*. The product is stored in a zone where the storage temperature is closest to the optimal storage temperature of the product.
3. *Preferred zone policy*. The preferred zone to store the product is determined by the slot plan. If the preferred zone has not enough unoccupied storage space, the product is stored in a zone with available storage space where the expected relative keeping quality loss is minimal.

The first two storage policies are based on existing practices in distribution centres for perishables. The free zone policy only focuses on storage capacity utilisation. The temperature zone policy requires only a little amount of additional information about the keeping quality characteristics to reduce the keeping quality loss compared to the last policy. The preferred zone policy is derived from the class-based storage policy. In the case of a retrieval, the first-in-first-out rule is only followed when no alternative zones are occupied with the product. Stock in alternative zones with harmful storage conditions for the product has priority over stock in the assigned or default zone and is therefore used first for a retrieval.

In the thesis of Broekmeulen[3] it is shown that the preferred zone policy outperformed the other two policies in all tested problem instances. The use of a slot plan compared to a simple quality-oriented assignment rule in the temperature zone policy reduces the keeping quality loss by at least 50% and avoids exceeding the maximum allowed keeping quality loss that is often imposed in distribution chains for perishables.

20.3.2 Slot planning models

The following constraints have to be met when slotting the products for storage in a distribution centre:

- Each product must be assigned to a slot.
- The storage assignment cannot exceed the capacity of a slot.
- The loss in keeping quality of the products may not exceed more than a predefined level agreed in the distribution chain.

The objective of a slot plan is, given an instance of the problem, that the management of the distribution centre finds a slot for all products in the assortment of the distribution centre such that the sum of the keeping quality loss is minimal, and that the constraints are met. The sum of the keeping quality loss is called the object value. An instance of the problem can, among others, be defined by data about the layout of the building, the expected customer orders, and the properties of the products in the assortment. The customer orders determine the flow of products through the distribution centre. The minimal required information about the expected customer orders should include the expected stock levels for each individual product in each planning period. This information can be obtained by analysing the historical transaction data of the distribution centre. A distribution centre may use the same slot plan for several years if the yearly turnover remains constant over the years.

The assignment of storage conditions to the zones and the selection of a storage policy are considered long-term decisions. The choices made in the layout and in the selection of the equipment determine the available slots and the characteristics of these slots in the distribution centre.

20.3.3 Integration of product characteristics in a slot plan

This section will describe general aspects of models for ethylene production and keeping quality that are used for constructing a slot plan. The development and validation of these quality change models are open research topics. For a limited number of quality change groups, these models are founded on a solid theoretical basis and are verified by experiments.

The following assumptions were made with respect to the keeping quality of the products in the distribution centre.

- *Stock level.* The average volume of a product that is stored overnight in a slot is equal to the allocated storage space of that slot. Therefore, the ethylene production of a product is based on the allocated storage space of that product instead of the actual stock level.
- *Quality change group.* All articles are assigned to a quality change group, based on the quality change properties of the product. The factors initial keeping quality and storage time are used to assign a product to a quality change group.

Temperature

Kopec[14] states that temperature is the most important effect with respect to the keeping quality of perishables. The effect of temperature on the keeping quality in the absence of all other effects is called the normalised keeping quality. The normalised keeping quality is influenced by two quality change processes: ageing and chilling injury. According to Tijskens *et al.*,[15] ageing speeds up with higher storage temperatures, and chilling injury gets more severe at lower storage temperatures.

The current models assume that the keeping quality at a specific temperature is always reduced by the effects of product interactions and handling. This is not the case for controlled atmosphere storage and modified atmosphere packaging that can also improve the normalised keeping quality. The reference temperature is the optimal storage temperature in the absence of product interactions and handling effects.

Product interaction

In general, the interaction between various products is modelled based on agents such as odour compounds and ethylene. For example, the odour of onions is transferred by a different agent compared to the odour transfer of melons. Furthermore, some products produce the agent and some are susceptible to the agent. Moreover, the product that produces the agent may also be susceptible to it at the same time. The levels of production of and sensitivity to those agents are specific for the quality change group and dependent on the storage temperature. For example, mushrooms produce no ethylene and bananas and tomatoes are examples of products with high ethylene production.

The ethylene concentration in a zone depends on the number of quality change groups assigned to and present in the zone and the total volume of the zone. The volume of the zone has to be adjusted with respect to the volume of the quality change groups that are assigned to it and that are present in the zone, and with respect to the rate of ventilation. A ventilation rate of 2 means that the effective volume of the zone also increases by a factor of 2. High utilisation and low ventilation of a zone may result in a high ethylene concentration, even when the ethylene production is relatively low. The effect of ethylene on keeping quality increases dramatically with concentrations between 1.5 and 3.5 ppm.

Handling

Handling causes shock and vibration and thus stress to fresh produce. The effect of handling on the keeping quality depends primarily on the quality change group. Strawberries are an example of products that are highly susceptible with respect to handling. Processed foodstuffs are seldom susceptible to handling.

Relative keeping quality

Since the distribution centre acts as a decoupling point in the distribution of perishables, the past and future storage times and conditioning of the products that pass the distribution centre are not known in sufficient detail. Therefore, the slot planning model has to use a measure of keeping quality loss in a distribution centre that is independent of the initial keeping quality of the products, the duration of stay in the distribution centre, and the remaining time of the products in the distribution chain until consumption by the end consumers. The proposed measure relates the keeping quality loss in a specific slot plan to the keeping quality loss at ideal conditions. The relative keeping quality loss is the decrease in keeping quality compared with the keeping quality at the reference temperature without any effects of handling or ethylene.

In the proposed measure, the relative keeping quality loss is equal to the fraction of the initial keeping quality that is lost every day compared to storage at the ideal conditions for that quality change group. A relative keeping quality loss of 50% in a distribution centre implies that during each day the product is stored in the distribution centre, the shelf-life is reduced by one and a half days instead of just one day and the subsequent stages in the distribution chain have less time to deliver the product to the consumer. When the normalised keeping quality corresponds with a shelf-life of less than one day, the storage of that quality change group is impractical, the total shelf-life expires during storage, even at ideal conditions.

20.3.4 Influence of climate zones and planning periods

The possibility of assigning ideal conditions to each product is determined by the number of climate zones in the distribution centre. Therefore, the number of available zones and the specified storage conditions in these zones have a large impact on the keeping quality loss of the stored products. A large number of zones makes it possible to store interacting products separately. But a large number of zones lowers the potential total utilisation rate of the distribution centre. The use of ventilation and/or sealed (transport) packagings can avoid product interactions to a large extent. Ventilation of cold stores removes the precious cold air and therefore increases the energy bill. Since ventilation does not reduce the risk of contamination, products such as fresh meat and vegetables still have to be stored separately. For most assortments of vegetables and fruits, a distribution centre needs a minimum of two and preferably three zones, with storage temperatures 0ºC, 8ºC and 12ºC as an indication. Greens, such as lettuce and endive, require a storage temperature between 0ºC and 2ºC in the absence of ethylene-producing products. Ethylene-producing exotic fruits prefer storage temperatures above 12ºC. Without ample ventilation, an additional third zone with a storage temperature between 8ºC and 10ºC is needed for products of the common *Solenacea* family such as potato, peppers and egg plant.

For seasonal products the number of planning periods determines the storage capacity utilisation, since the peak stock level of a seasonal product depends on the time of year and the length of the planning period considered. A longer planning period increases the chance that a certain peak occurs in the planning period considered. The sum of the daily stock levels for their demand period complementary products such as chicory and strawberry results in a lower combined peak stock level than the sum of the individual peak stock levels. During the summer, the low in the stock level of chicory is compensated by the peak in the stock level of strawberry. The size of this effect is illustrated in Fig. 20.1, where the peak stock levels of all products of a major wholesaler of vegetables and fruits are summed. The summed peak stock level goes down when the number of planning periods increases. Even with twenty planning periods, the summed peak stock level is almost three times as high as the summed average stock level of all products.

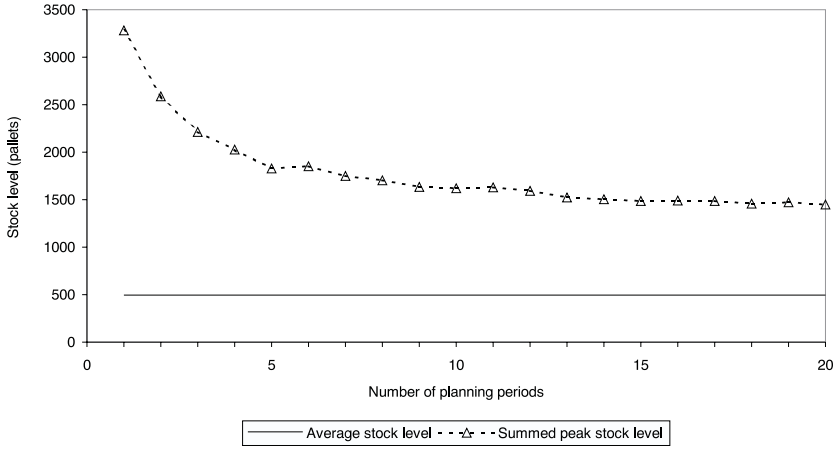


Fig. 20.1 The summed peak stock levels of all products as a function of the number of planning periods in the dataset of a wholesaler of vegetables and fruits.

20.4 Construction of a slot plan

A solution for the slot planning problem can be constructed by a greedy heuristic. A heuristic is called greedy when it assigns values to the decision variables one by one without changing the previous assignments afterwards. The proposed construction algorithm assigns products with high peak stock levels before products with lower peak stock levels to the zones until the capacity is exhausted.

20.4.1 Construction algorithm for a slot plan

The following algorithm can construct for each planning period a slot plan, given the number of zones and the storage conditions in these zones.

1. Determine for each product the expected stock level in the planning period. Determine for each zone the available storage space.
2. Sort the products by stock level in descending order. Start the assignment with the largest product.
3. Calculate for the current product the relative keeping quality loss in each zone with enough available storage space to store the product. To minimise possible product interactions, take already assigned products in the zone into consideration. In the case of interactions induced by ethylene, one can do this by calculating the expected ethylene concentration after the assignment is calculated and from there the effect on the keeping quality of the assigned products in the zone.

4. Assign the product to the zone with the lowest expected relative keeping quality loss for the current product as well as the already assigned products. Subtract the available storage space in the selected zone with the expected stock level of the current product. When the product does not fit in any of the zones, return to Step 1 to modify the expected stock levels of the products or the storage capacities of the zones.
5. When all products are assigned, then the algorithm is finished. Otherwise, take the next product from the sorted list and continue with Step 3.

When a slot plan is required where each product must be assigned to the same zone in each planning period, then the algorithm has to assign the product to the same zone concurrently for all planning periods. The main modification of the algorithm occurs in Step 2 where the products are sorted by the maximum of the expected stock levels of the planning periods. A so-called fixed slot plan makes it easier for personnel to find the products all year round.

20.4.2 Example of the construction of a slot plan

The following detailed example illustrates how to make a slot plan with the construction algorithm for a distribution centre for vegetables and fruits. Assume a distribution centre with three zones A, B and C. Zone A has a temperature of 2ºC, zone B of 8ºC and zone C of 12ºC. Zones A and B have no mechanical ventilation and zone C is well ventilated. Each zone has storage space for 100 units. Table 20.1 lists the products that have to be stored in the distribution centre with the expected keeping quality losses according to the keeping quality change models. When all the products are assigned, the utilisation will be 80%.

Each zone initially has 100 units storage space available and the ethylene concentration is low. The sequence in which the products are assigned is based on the stock level (from large to small): potato, banana, Elstar (apple),

Table 20.1 The assortment of the distribution centre with the expected keeping quality loss at three temperatures and two levels of ethylene (H = high, L = low)

Product name	Stock level (units)	Relative keeping quality loss (%)				
		2L	2H	8L	8H	12
Banana	45	82	84	41	81	14
Cucumber	14	76	78	35	79	7
Elstar	35	0	0	35	35	54
Garlic	5	15	15	48	48	62
Grapefruit	25	71	72	14	37	53
Haricotvert	10	56	57	0	27	39
Lettuce	17	26	30	70	90	83
Melon	9	72	71	18	40	26
Potato	65	72	73	10	34	38
Satsuma	15	36	38	33	51	55

grapefruit, lettuce, satsuma (mandarin), cucumber, haricotvert (green bean), melon and garlic.

In Table 20.2, one can follow the assignment process. After assignment step 3, when Elstar is assigned to zone A, the ethylene level in zone A changes from low to high. For the assignment of lettuce in step 5, zone A is selected despite the high ethylene level because the alternatives are worse. Since zone B has not enough space left for satsuma, this product is assigned to zone A in step 6. The additional keeping quality loss of this assignment is only 5%. Since zone C is well ventilated, cucumber is not influenced by the ethylene producing banana after assignment step 7. Melon is assigned to zone C because the first choice, zone B, is already full. Fruits such as Elstar and satsuma are slightly damaged by the odour of garlic, but the keeping quality loss for garlic in the alternative location, zone C, is too large.

All products are now assigned, but not always to the most suitable zone for the individual product. The additional keeping quality loss is only limited, as is shown in Table 20.3. The relative keeping quality loss is never above 50%.

20.4.3 Improvement of a slot plan

The slot plan found with a construction algorithm can often be improved with local search techniques such as tabu search and genetic algorithms. In these local search techniques, the following cycle is carried out several times. In a first step, candidate solutions are derived from the current solution based on the following types of moves: a product is assigned to a different zone or two products swap their zone assignments. All candidate solutions that can be reached with these two types of moves are called the neighbourhood. In the second step, an acceptance criterion selects the new current solution for the next cycle. In the neighbourhood, the first candidate solution is accepted that has a better object value than the current solution. A local optimum has been reached if the object value of last accepted solution is better than all its neighbours. If such a local optimum is reached, a new initial solution is generated by assigning random values to the decision variables. After a limited amount of computation time the search is stopped.

Note that in the example different stock levels lead to another sequence of assignments and therefore may lead to large differences in keeping quality losses. The exchange of the stock levels of grapefruit and melon results in a completely different assignment. Therefore, improvement of the solutions found with the greedy heuristic remains necessary. This improvement can be achieved with the proposed local search heuristics.

20.5 Summary

This chapter showed a practical application of quality change models in the planning of the layout and the operations of a distribution centre for perishables.

Table 20.2 The assignment process with the proposed construction algorithm for the products in Table 20.1

Assignment step	Product	Zone A			Zone B			Zone C		
		Content	Space available	Ethylene level	Content	Space available	Ethylene level	Content	Space available	Ethylene level
0	–	–	100	L	–	100	L	–	100	L
1	Potato	–	100	L	Potato	35	L	–	100	L
2	Banana	–	100	L		35	L	Banana	55	L
3	Elstar	Elstar	65	H		35	L		55	L
4	Grapefruit		65	H	Grapefruit	10	L		55	L
5	Lettuce	Lettuce	48	H		10	L		55	L
6	Satsuma	Satsuma	33	H		10	L		55	L
7	Cucumber		33	H		10	L	Cucumber	41	L
8	Haricotvert		33	H	Haricotvert	0	L		41	L
9	Melon		33	H		0	L	Melon	32	L
10	Garlic	Garlic	28	H		0	L		32	L

Table 20.3 The effect of the assignment on the expected keeping quality loss compared with the lowest possible keeping quality loss at the available storage conditions

Product name	Relative keeping quality loss (%)		
	lower bound	slot plan	additional loss
Banana	14	14	0
Cucumber	7	7	0
Elstar	0	0	0
Garlic	15	15	0
Grapefruit	14	14	0
Haricotvert	0	0	0
Lettuce	26	30	+4
Melon	26	26	0
Potato	10	10	0
Satsuma	33	38	+5

The problem of constructing a slot plan has been implemented in a decision support system with a hierarchical decomposition. The chosen model formulation and solution strategy focus on the seasonal production and demand of perishables, and the need for special storage conditions during the distribution process, described in the applied quality change models.

The use of an assignment plan is useful for the operation of a distribution centre for vegetables and fruits. The additional effort to make such a plan is relatively small compared with the benefits for keeping quality. The assignment rules used in the currently used storage policies are insufficient to ensure specific conditions for the products.

It is recommended that the layout of a distribution centre for perishables is based on the assortment and on the position of the distribution centre in the supply chain for the products in the assortment.

20.6 References

1. HOEKSTRA SJ, ROMME J H J M, *Towards Integral Logistics Structures* (in Dutch, second edn.), Kluwer, Deventer, 1987.
2. SIMS JR. E R, 'Facilities design and maintenance'. In: Tompkins J A and Harmelink D A (eds), *The Distribution Management Handbook*, McGraw-Hill, London, 1994.
3. BROEKMEULEN R A C M, *Layout and operations management of distribution centers for perishables*, PhD thesis, Eindhoven University of Technology, Eindhoven, 1998.
4. MEFFERT H F TH, 'Economic developments pertinent to chilled foods'. In: Gormley T R (ed.), *Chilled Foods: The State of the Art*, 337–61, Elsevier Science Publishers, Barking, 1990.

5. FU B, LABUZA TP, 'Shelf-life prediction: theory and application', *Food Control*, 1993 **4** 125–33.
6. SHEWFELT RL, 'Quality of fruits and vegetables', *Food Technology*, 1990 **44** 99–106.
7. POLDERDIJK JJ, TIJSKENS LMM, ROBBERS JE, VAN DER VALK HCPM, 'Predictive model on keeping quality of tomatoes', *Postharvest Biology and Technology*, 1994 **2** 179–85.
8. FLOROS JD, 'The shelf life of fruits and vegetables'. In: Charalambous G (ed.), *Shelf Life Studies of Foods and Beverages*, 195–216, Elsevier Science Publishers, Amsterdam, 1993.
9. LABUZA TP, *Shelf-life Dating of Foods*, Food and Nutrition Press, Westport, 1982.
10. RYALL LA, LIPTON WJ, *Handling, Transportation and Storage of Fruits and Vegetables* (second edn.), AVI Publishing Co., Westport, 1979.
11. ABELES FB, MORGAN PW, SALVEIT JR ME, *Ethylene in Plant Biology* (second edn.)., 182–221, 294–6. Academic Press, San Diego, 1992.
12. TIJSKENS LMM, POLDERDIJK JJ, 'A generic model for keeping quality of vegetable produce during storage and distribution', *Agricultural Systems*, 1996 **51** 431–52.
13. HAUSMAN WH, SCHWARZ LB, GRAVES SC, 'Optimal storage assignment in automatic warehousing systems', *Management Science*, 1976 **22** 629–38.
14. KOPEC K, 'Effect of storage condition on harvested vegetables and their mathematical simulation', *Acta Horticulturae*, 1983 **138** 343–54.
15. TIJSKENS LMM, VAN KOOTEN O, OTMA EC, 'Photosystem 2 quantum yield as a measure of radical scavengers in chilling injury in cucumber fruits and bell peppers', *Planta*, 1994 **194** 478–86.

21

Concepts of chain management and chain optimisation

L. M. M. Tijskens, A. C. Koster, J. M. E. Jonker, ATO BV, Wageningen

> 'Let Pharaoh appoint commissioners over the land to take a fifth of the harvest of Egypt during the seven years of abundance. They should collect all the food of these good years that are coming and store up the grain under the authority of Pharaoh, to be kept in the cities for food.
>
> 'This food should be held in reserve for the country, to be used during the seven years of famine that will come upon Egypt, so that the country may not be ruined by the famine.'
>
> Genesis 41: 34–36

21.1 Introduction

Since man started to grow his own food, instead of searching and hunting for it, the sites where we consume our food are gradually taken away from the production sites. In recent decades this process has sped up enormously. Industry has almost completely taken over control of the production and distribution of our foods from local production and consumption. The food chain becomes more and more a global issue.

Within the context of this book, several techniques of modelling have been described and examples are given to elucidate the building and application of models. All these examples are somehow related to the behaviour, quality and safety of a commodity. All these models, however, have to be applied within a larger framework that includes the responses of sellers, buyers and consumers of that commodity. This introduces all consumer research and economics into the game. Commercial companies will only be interested in modelling when a clear

increase of sales and profit will be the result of this application. This framework in modern sales technology is the supply chain.

The modelling techniques used and the models described in the earlier chapters of this book primarily deal with issues directly or indirectly connected to food properties and food behaviour. The main difference of food supply chains as compared to non-food supply chains, is the time critical and ever-changing quality and properties of the perishable food during their journey from growing site to consumption site. Food commodities have to be handled, transported, distributed and eventually sold to and consumed by the consumers with as high a quality as possible for as little cost as achievable. This whole sequence of actions and the omnipresent decay of quality require proper management to achieve the high standards as set by the modern consumers and retailers. Due to the alienation of consumers with growers and retailers, the reliability of continuous supply with high and acceptable quality inevitably leads to a risk of acceptance both for the buyer as for the seller of commodities. As a consequence, chain supply management not only comprises the purely physical aspects of delivering commodities but also the reduction of risks involving sales, quality, safety and consumer acceptance.

Apart from the perishable nature of food, another property of food is psychologically very important to the consumer. Unlike other commodities bought by consumers, food is used as the raw material for building up and maintaining our bodies. When the commodity does not prove safe and reliable, consumers can no longer put it aside, as it is already built-in in our bodies. This places a very major emphasis on safety and health aspects of our food, both in the short term and in the long term.

The necessity for the food industry to set high standards for food safety and overall food quality has led to an enormous increase in the development of conceptual models for the optimisation of the throughput of products in the chain of production.[1] The modelling of food processes and secondary product characteristics has to be complemented by development and implementation of conceptual models that deal with more managerial type of problems that arise when managing and optimising food supply chains. This chapter will focus on some important issues in chain optimisation and chain management, within the framework of consumer driven supply chains.

21.2 Key principles and methods

Chains in general and food supply chains in particular can be considered on a number of aggregation levels. For most product-oriented experts and modellers, the food chain consists of the successive actions performed to that particular batch

[1] Within the area of economics and chain management, production signifies more than just the physical production. It also comprises the auxiliary infrastructure for making decisions and all the actions to support and maintain the flow of information.

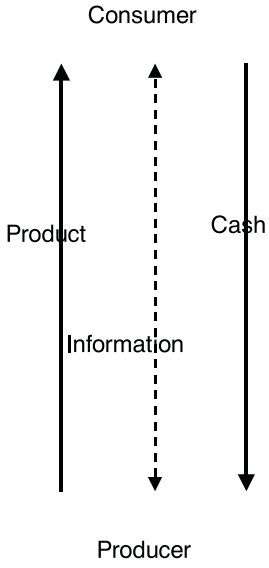


Fig. 21.1 Different types of flows within a chain.

of product ranging from growing, transportation, storage, and distribution over retail up to the final consumer. As such it consists, from the products point of view, merely of the actual scenario, expressed as time-temperature combinations during its journey from the producing area to the consumption area. From an economic and marketing point of view, however, a chain is a much more complex structure of product flow, cash flow and information flow (see Fig. 21.1) that needs to be understood and described for proper management and optimisation. Several views and definitions have developed during the last decades.

21.2.1 Chains and networks

A chain is a network of autonomous and specifically named organisations, systematically cooperating in the production of a commodity. The cooperative relations are more than incidental and can vary from direct linear to complex network forms (van Dalen 1994).

The notion 'chain' has been defined in different ways (Meulenberg and Broens, 1997). Three definitions will be considered in more detail:

1. The *value chain* as defined by Porter (1980): *a functional definition*
2. *Vertical coordination*: A cooperative relation of two or more organisations in a production column involving the coordination of decision making towards an increased performance for a common third party (Porter 1985): *an institutional definition*.
3. The *network chain*: the attitude or viewpoint that enterprises participating in a production chain, have to coordinate their actions as well as possible

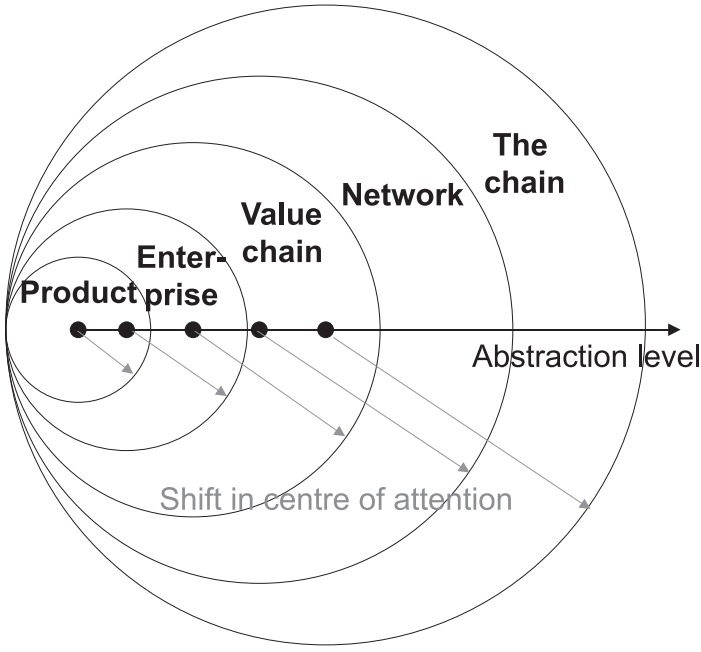


Fig. 21.2 Increasing levels of conceptual abstraction in chains: from the single produce going through a distribution history, over the single enterprise dispatching produce, to the three types of chain definitions. With increasing conceptual abstraction level the centre of attention shifts from the product over the enterprise to a centre outside the actual actors in the chain.

based on consumer defined requirements of risk, quality and acceptability: *a conceptual definition* (Ford 1990, Haakansson and Snehota 1989, Jarillo 1988, Nohria and Eccles 1992, Beers *et al.* 1998, Thorelli 1986).

These three viewpoints form a conceptual development with an increasing complex and abstraction level. A visualisation of this development is shown in Fig. 21.2.

A functional definition: value chain

One of the oldest interpretations of food chains is provided by Porter (1980). It considers five operations, conducted within one company that *collectively* prepare a commodity for a specific customer: inbound or production logistics, production, outbound or distribution logistics, marketing and sales. These five operations are among others supported by the information structure within one company. A disadvantage of the description/definition of Porter is that it only concerns activities within one single company.

Whereas Porter (1980) starts off with the concept of strategy of the individual enterprise, he also combines the 'forces' that affect the competitive position of the enterprise within the concept of the value chain, in which the enterprise

enhances its competitive position by improving the links between the value adding processes in its value chain.

Extrapolation of Porter's views to a multi-company effort provides an interpretation of the chain as a consecutive sequence of logistic and marketing *activities* in preparing a commodity or goods for delivery to a specific end-user or consumer. In this definition, ICT (Information and Communication Technology) is not an inherent part of the chain as a whole but remains a supporting activity that exists separately.

An institutional definition: decision making in vertical coordination

Vertical coordination is the central issue in the *institutional approach* (Porter 1985, Zuurbier *et al.* 1996). The enterprises participating in a chain do not act independently but are coordinated, like retailers in a franchise organisation with a centralised purchasing facility. This coordinated action is not necessarily formalised in contracts. Most of the time, due to the emphasis on cooperation, not every enterprise, acting in that chain, is a member of this coordinated action framework.

Enterprises form a kind of network of mutually cooperating enterprises. The enterprises are not only interdependent for their physical products and raw materials but also in terms of service, information interchange and cash flow as influenced by third parties in the supply chain. This transforms the supply chain into a network of actors (Fig. 21.3), whose operations are mainly related to getting high quality products and information on to the market. Within the institutional definition, the cooperation is primarily driven by the necessity for improvement.

Vertical coordination has both strategic and operational aspects. Strategic coordination comprises, e.g., coordination of product design, process design and structuring information infrastructure. Operational coordination can occur on the level of production means and information flow, parallel to flow of goods, cash and property[2] (Mallen 1977). This explicit cooperation regarding information flow and infrastructure will inevitably lead to needs and demands for specific information. Hence, within this definition, ICT plays a central role. Not only the infrastructure of information exchange, but also the actual usage of information from this structure and its interpretation are essential common resources.

A conceptual definition: the network chain

In the first two approaches on chain structure and meaning, the viewpoint was obtained from one enterprise or from a cluster of enterprises looking at the necessary actions to get a commodity to the final user. In the conceptual definition, only one chain, 'the' chain exists, as an abstract entity outside the enterprise(s) involved. The enterprise(s) can cooperate voluntarily with this

[2] This multidimensional approach in information flow can nowadays also be perceived at the level of flow of goods. Not only the traditional flow from producer to consumer is being considered in this approach, but also the return flow of auxiliary material like packaging and waste material.

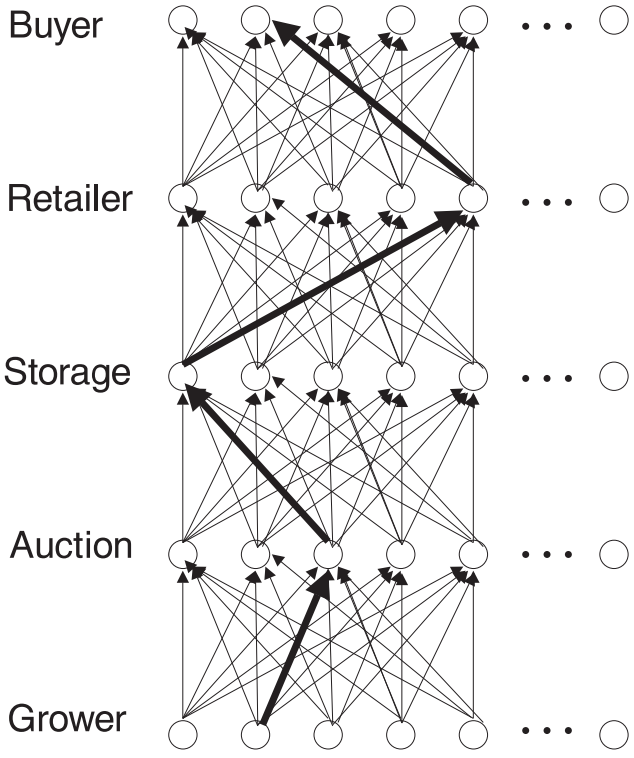


Fig. 21.3 Network chain and vertical coordination, with one possible pathway highlighted in bold.

chain, but the existence of the chain has to be taken into consideration in the commercial policy. Chain policy can then be considered as a policy for the whole branch to achieve the requirements of the customer by cooperation. This third definition of chains puts the main emphasis on the awareness for the need of cooperation, rather than the cooperation itself. Note that this definition of *the chain* comprises all issues defined by the *institutional* definition.

The type of cooperation between actors in a chain can reach from almost non-existing (e.g. sellers at a local market place) over contractual cooperation to a complete merger (see Fig. 21.4). In this figure a number of features of cooperation are visualised for non-cooperative actors (local markets), a network chain and a fully merged enterprise. Of course the coefficient of vertical coordination increases from local markets over network chains to the merged enterprise. The flexibility of local markets as well as for fully merged enterprises is high: they can decide on the spur of the moment to change suppliers or target other buyers. However, the cost of changing over is for local markets quite low,

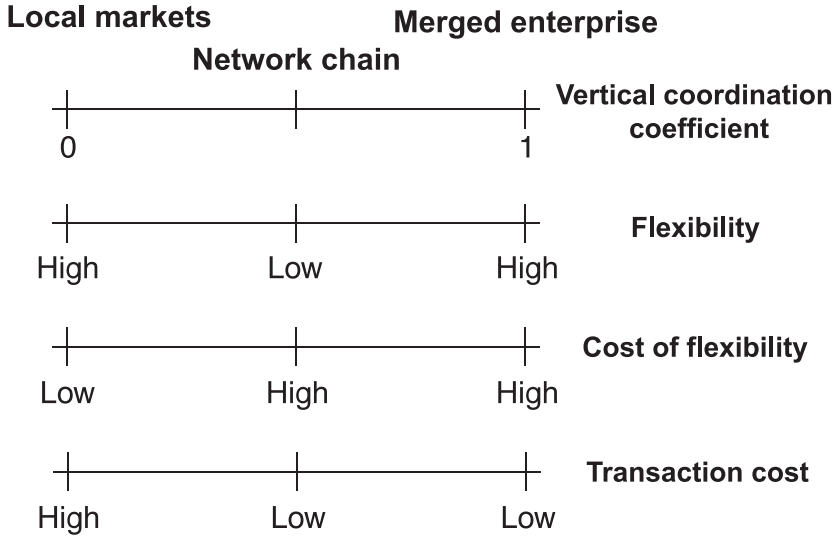


Fig. 21.4 Levels of vertical coordination, adaptability and its cost indicated for different levels of networking.

while for network chains and merged enterprises the costs are relatively high due to the massive changes throughout the organisation. The costs of transactions are, on the other hand, relatively high for local markets, since they cannot rely on mutual built-up trust, as the network chains and the merged enterprises can. Within the conceptual definition and the connected efficient consumer response (ECR) concept, the enterprises in a network chain cooperate non-contractually but on a long-term basis, due to the recognition of mutual advantage.

21.2.2 Incentives for supply chain management

The great changes in market structure, consumer behaviour and consumer needs and the massive changes in technology, especially the information technology, have been major reasons for the need of supply chain management. Retailers have gained more and more power on our food supply (supermarkets), and enterprises are merging all over the world to grow into large multinationals. Consumers act more and more as individuals in an increasing population (mass individualisation). The innovations in computer technology and information made it possible to start considering to manage the whole chain.

Supply chain management deals with total business process excellence and represents a new way of managing business within each link and the relationship with other members of the supply chain (Lambert *et al.* 1998). Since the 1980s, literature on supply chain management stresses the need for collaboration among successive actors, from primary producer to final consumer, to better satisfy consumer demands at lower costs (Ellram 1991, Towill 1996, Bechtel and

Jayaram 1997, Cooper *et al.* 1997). Stevens (1989) refers to the interdependency of activities in the supply chain: 'if one activity fails, the chain is disrupted, creating poor performance and destabilising the workload in other areas, thereby jeopardising the effectiveness of the supply chain'. This holds true especially in food supply chains because of the shelf life constraints of food products and because of increased consumer attention for safe and environmental/animal friendly production methods (Boehlje *et al.* 1995). Recent events like e.g. the BSE crisis in Western Europe and the classical swine fever in the Netherlands and Germany made producers aware of the necessity of supply chain control and intensified supply chain cooperation.

The increasing distance between production site and consumption site from local production to production more distant from the consumer, inherently calls for an equal shift in reliability of information, from pure confidence of consumers in the retailer at the street corner, to more objective and measurable quality indicators. In the early days, consumers relied almost completely upon previous experience, buying at a specific retailer, building up gradually the confidence in that retailer. With the more impersonal supermarkets growing more and more important, the consumers need increasing product information regarding quality, safety, content and ultimate selling date to reach a decision whether or not to accept that particular product. But not only the final consumer is faced with this increasing need for information. For each actor in the chain, that product and history information becomes more and more important in deciding which products to buy to be certain he or she is going to be able to sell (van Trijp and Meulenberg 1996).

The increasing importance of information for the successive actors in the chain of events has a number of severe practical and managerial consequences (economics of scale and scope). Coordinated cooperation of all actors in the chain can:

- Increase the efficiency of material and information resources
- Reduce risk of safety and quality
- Decrease the production life cycle
- Decrease the cost of obtaining and retaining that information throughout the chain (transaction costs).
- Decrease the cost and number of repackaging a commodity by different actors in the chain (packaging cost and environmental benefit).
- Reduce the stock level by one actor, based on production information by the successor actor.
- Decide the optimal location in the chain for quality increasing investments.
- Increase profits.

Also strategic issues can be reasons to search for closer cooperation and sharing of information. A good example of sharing information to reduce costs and increase profits is the system applied by the world's largest retailer Wal Mart (USA). This enterprise allows free and total access to the purchase and consumer database without restriction or prescription, as long as added value of its usage can be substantiated by increased sales results.

Motives for cooperation within the supply chain, asking for *complex innovations*, not only come from inside the food supply chain (e.g. commercial, organisational and technological influences), but also from outside the chain (e.g. social, legal, political and community influences).

21.3 Food supply chains

For consumers of a product of a food production and supply chain, and hence for the chain as a whole, the quality of a purchase comprises more than the pure intrinsic quality of the purchased product. When deciding to buy a certain product, the buyer finds a balance between, among others, his own preferences, the preferences of its final consumer, the assumed properties of the product, the availability of the product on the market and the costs involved. The overall picture of this balance evaluated at the moment of every purchase decision can be found in Fig. 17.3 of Chapter 17. Combining the information from that chapter, and the viewpoints behind it, with the actual motives and considerations of purchase, the issue commonly called quality, is more the acceptance of a product by the consumer/buyer (Meiselman and MacFie 1996). In this acceptance the intrinsic quality of a product is only one of the many issues affecting the consumer's decision. Most of these issues are subjective in nature, including the risk assessment, safety factors, and product quality itself. For a useful application in the integrated supply chain, all these factors have to be recognised, described and modelled, and techniques have to be developed to measure all these factors affecting the consumer's acceptance in an objective manner. All these factors are dealt with in previous chapters of this book.

21.3.1 Modelling the chain

The art of modelling is not yet very advanced in the realm of chain description and chain management, especially on the level of individual commodities and batches of commodities. The subject is too young and too complex to have already achieved a comprehensive mathematical description. The models concerning the chain as an integral entity are still limited to conceptual models as described in previous sections. However, increasingly models generated in other parts and disciplines, concerning product behaviour, product quality, keeping quality (or shelf-life) and safety are used in description and optimisation of parts of the integral chain. A number of these models and viewpoints are dealt with in previous chapters of this book (Chapters 5, 19 and 20). Specific conferences are being organised that cover the development in this area (Ziggers *et al.* 1998, Triekenens and Zuurbier 2000).

21.3.2 Acceptance versus quality

Acceptance of a product by the consumer is the ultimate goal and the ultimate test of the supply chain. Managing that chain and optimising it, signifies that one

has to make sure that the consumer in the (near) future goes on purchasing his foods from that particular chain. Within the decision of purchase, a number of issues play an important role. These issues are of very different nature and impact.

- Product-related issues:
  - The intrinsic quality of the product purchased, however ill defined, and the connected keeping quality. This issue brings in the whole physical pathway the product has to go through, including all aspects of actual scenarios with respect to e.g. temperature, and the actual physical handling performed to the product like, e.g. cleaning, packaging and processing.
  - Safety of the commodity e.g. microbial infection, radiation, health.
- Market-related issues:
  - Availability of comparable products on the local market
  - Confidence in the enterprise with respect to, e.g. quality, applied technology and information
- Economic-related issues:
  - Costs of purchase of the commodity
  - Price reductions
  - Sales promotions
- Social issues within the boundaries of the social community:
  - Is the commodity ethically acceptable, e.g., child labour?
  - Is the technology applied acceptably e.g. genetic modifications, organic production, etc.
- Psychological issues:
  - Does this commodity provide the buyer some social status?
  - Does this commodity fulfil consumer's expectations? (MacFie and Thomson 1994).
  - Is the information about the product, its contents and its production and processing adequate and reliable? (advertising and labelling).

Given the ongoing mass-individualisation of the contemporary consumer, the understanding of and the compliance to all these items obtain more and more weight for commercial companies. Some effects in the decision making of a consumer are already clear and described while many effects are not clear at all and still remain very vague. All efforts and actions within a particular chain are directed towards that ultimate goal: keeping the consumer happy so he will come back. Managing and optimising a supply chain aims at maintaining the structure and functioning of the complete supply chain in such a way that the buyers/consumers do come back and keep on purchasing commodities from that chain, while still making some profit in the process. This signifies that efforts have to be directed to the many unresolved relations between decision making and supply, while maintaining the paramount properties of quality and safety. Managing a chain is finding the balance between purely economic issues (profits, transaction costs), flexibility in pathway and commodity throughput (see e.g. Fig. 21.3) and maintaining food safety and food quality.

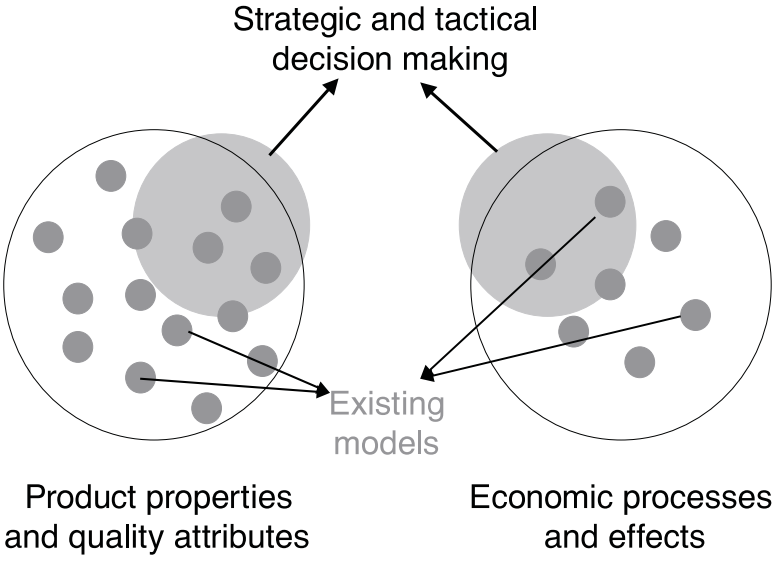


Fig. 21.5 Availability and application of dedicated product, quality and economic models in strategic and tactical decision making.

In the realm of product behaviour and product properties, many items are already clear. Relations and effects have been described and modelled, as can be taken from many previous chapters in this book. All these models on food quality, safety and handling can be and are being used in optimising the product going through supply chains. In the realm of specific economic issues, the science of modelling in the chain and network area, is still in its infant stage. When making strategic and tactical decisions, a fair part of information on product properties and behaviour is already modelled and implemented. Structured information on economic effects and processes are, however, still quite sparsely available (see Fig. 21.5). Some progress has, however, been made (ECR or Efficient Consumer Response (see section 21.5), sales promotions (Wierenga and Soethoudt 2000). In the realm of psychological and of social issues, consumer research has come up with a number of empirical relations and rules of thumb, but has hardly achieved any fundamental breakthrough in really understanding the ongoing processes. Some examples of psychological relations and unstructured facts are, e.g., a more emotional involvement when buying food products compared to non-food products: buying of food occurs much more at the spur of a moment or affected by a particular space; food is bought on a repetitive basis; consumers walk preferentially at the outer perimeter of supermarkets and avoid more the inner space.

Combining all these effects for managing and optimising supply chains, inherently leads to very diverse levels of understanding of the effects within and throughout a supply chain.

What has been developed during the past decades is a decomposition of supply chains into the participating actors and into the relations between those actors: an understanding of the structure of a supply chain.

21.3.3 Technological issues

Technology of food production, processing and handling, has shown some drastic changes over the years. All of these changes were initiated by incentives and motives of increasing the quality of the delivered commodities combined with an increasing profit or decreasing production costs (*technology push*). The results of some technological changes were accepted by the customers without any protest or objection, e.g. sterilisation, MAP packaging, microwave heating, minimal processing. Other technologies studied never came to appreciable commercial application by lack of economic results, e.g. ohmic heating, infra red heating, large-scale maceration. However, in some cases, objections of ethical, safety and risk hazards nature were so strong, customers did not accept product treated with that particular technology (negative *market pull*). An example of a rejected technology is, e.g., application of ionising radiation. Recently quite some fierce discussions are ongoing on the risks and ethical acceptability of genetically modified substances in our food products and the occurrence and avoidance of BSE in bovine products.

As a consequence, innovations applied to our food production and supply system, have always been and will always be subject to acceptance by the customer/consumer. Accepting or rejecting a particular commodity is in fact the only action possible for consumers. It is in these terms that consumer driven innovations have to be considered (*market pull*).

The food supply chain is ultimately the structure where this type of information from producer to customer and back from the customer to the producer takes place. Managing a food supply chain means therefore more than tactical and operational planning, but also has to include the long term strategic views on consumer response, together with providing adequate means of communication and information interchange between all actors in that chain.

21.4 Problems and issues of global sourcing

The main goal of enterprises whether or not as partner in a chain, is not to bring as high quality products as possible to the final consumers, but to bring products onto the market at as low cost and as high profit as possible, while keeping the consumers satisfied.

One of the major trends in supply chains as a consequence of increased technological abilities, is globalisation of resources and consumption or internationalisation of trade. More and more a substantial portion of our daily food no longer originates from local or national resources, but the origin spreads

out all over the world e.g. Europe and North America import more and more kiwis from New Zealand, mangoes from the southern hemisphere, beef from Australia and Argentina, etc. Customers in North America and Europe are more and more used to obtaining all products the year round, without limitation of growing season. This aspect of consumer demands also puts considerably more emphasis on understanding and modelling product quality and the interactions between successive enterprises in the chain.

To optimise the whole process of buying raw material and selling final products on a global market, several aspects come into play that for itself have no bearing on the product and the product behaviour itself. A common quality language is absolutely necessary throughout the chain, not only to communicate but also to define which aspects of quality and product properties should be measured and in what way. That is why consistent theories on quality as e.g. described in Chapter 17 and the related and emerging common language (Shewfelt and Tijskens 2000) are so important. Also essential is a transparency and understanding of the complete process of a product going through the chain. A major factor in this transparency results in a mutual trust between the successive actors in the chain. With sufficient and reliable exchange of information between the partners in a chain, transaction cost can be decreased considerably.

To ascertain the quality and safety of commodities and to minimise risks (of all kinds), several difficulties arise in this so-called global sourcing. All difficulties involved are already mentioned in previous sections of this chapter. By the sheer distance between original producer and final consumer, the difficulties require, however, special attention and precautionary measures.

Food safety is an issue of such paramount importance, consumers in the western world are becoming more and more aware of the risks involved in the food production chains. Whether it is a case of BSE in the animal production chain or a case of pesticide residues in the fruit and vegetable chains, consumers demand that food companies and retail outlets do everything that is required to avoid any health risks.

In the second place the paramount quality has to be guaranteed and maintained. Solutions for this problem are sought for by technological means like modified atmosphere packaging (MAP, see Chapter 14), transport in controlled atmosphere containers (CA) and developing breeds of produce with long shelf-life and temperature-time indicators (see Chapter 19). However, only part of the problem of maintaining food safety and quality throughout the chain, can be tackled with high tech end-of-pipe solutions. It is essential to overcome the gap between local (third world country) production systems and western consumption. The first step is taken by investing and developing the local infrastructure.

In order for private companies to be competitive in both local and international markets, chains of production have to be built from scratch. Most of the existing infrastructure is inefficient and usually not geared for change. These changes will not be implemented within some months. Periods of transmission may even mount up from five to ten years, depending on the

possibilities of implementation and awareness of the necessity of new concepts of supply chain management.

Farmers have to be educated to handle seed materials properly and apply suitable growing and harvesting techniques. Resources have to be made available to upgrade the primary chain in the network. These funds can only be made available if the products, coming out of the system, meet local and global quality standards, including food safety.

21.5 Practical application of supply chain management: efficient consumer response

Related to the concept of supply chain management, the concept of efficient consumer response (ECR) emerged in food supply chains, in which distributors and suppliers work closely together to offer better value to the consumer. By jointly focusing on the efficiency of the total supply system, rather than the efficiency of individual components, they reduce total system costs, inventories, and physical assets while improving the consumer's choice of high-quality, fresh products. In accordance with supply chain management the ultimate goal of ECR is to achieve a responsive, consumer-driven system in which distributors and suppliers work together as business allies to maximise consumer satisfaction and minimise system costs. Accurate information and high-quality products flow through a paperless system between manufacturing line and check-out counter with minimum degradation or interruption both within and between trading partners (Kurt Salmon Associates 1993).

The ECR working group comprising a group of industry leaders in the United States developed five guiding principles that concisely articulate the ECR strategy (Kurt Salmon Associates 1993): providing better value, committed business leaders, accurate and timely information, ensure the right product is available at the right time and a common and consistent performance measurement and reward system. Coopers and Lybrand (Anon. 1996) translated these guiding principles into a family of 14 ECR improvement practices, categorised into three clusters concerning marketing or category management, logistics or operations improvement and information technology or enabling technologies (see Fig. 21.6).

Three different approaches exist to explain and describe cooperative relations as found in efficient consumer response, on theoretical premises (Overboom 2000, Boehlje and Schrader 1998, Verhaegen *et al.* 1999, Bash and Davies 1998, Zylbersztajn *et al.* 1996, Thomas *et al.* 1995).

1. *Transaction cost theory*
   Costs of transaction are those costs directly caused by the simple transaction of buying and selling from one partner to another. They depend among other things on incomplete information, uncertainty about the market, about the behaviour of business partners, and about the draft and check-up of contracts. Without sufficient confidence in the quality of the bought

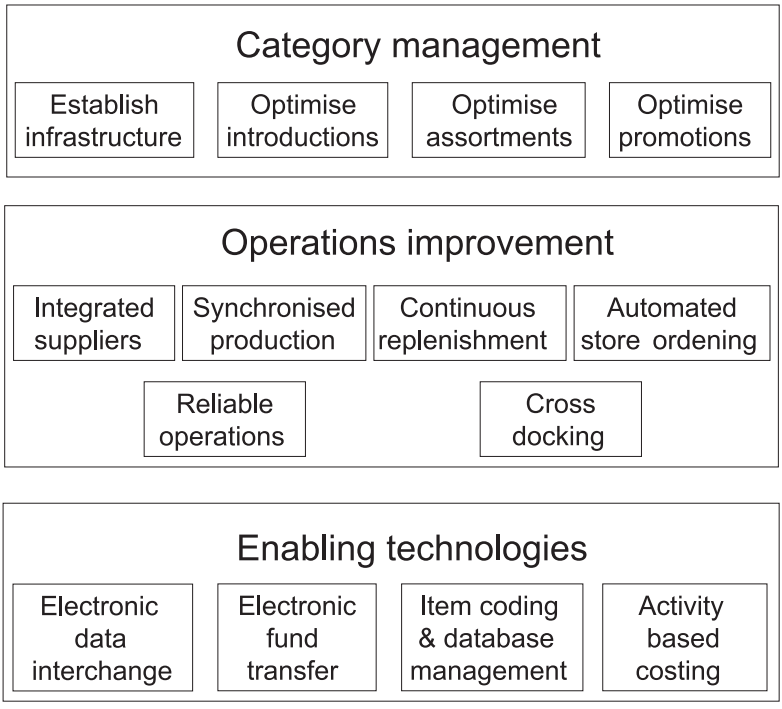


Fig. 21.6 Aspects of efficient consumer response.

product, every enterprise, every actor in the chain has to check the quality of the newly acquired goods, no matter what information is provided by the seller. With sufficient confidence, whether or not by contractual agreement, these costs can be reduced considerably. It then becomes a matter of mutual agreement how and when commodities have to be and will be tested for quality and content. The concept of transaction cost theory is useful for explaining and understanding the behaviour of enterprises and for improving their efficiency. Especially logistic issues of the ECR concept (operations improvement) can be understood using the principles of the transaction cost theory since the goals of logistics are always formulated in terms of efficiency.

2. *Strategic management theory*
   Competing strategies are the subject of strategic management. Michael Porter (1985) is the most prominent author in this school. The theory starts from the traditional industrial organisation, explaining the *performance* and the behaviour of the enterprise (*conduct*). The emerging cooperation between enterprises is in this theory a strategic response to uncertainties and interdependencies. The higher the uncertainties the more intense the cooperation will be (Pfeffer and Salancik, 1978). Using strategic management, concerted relations can be treated and considered as category

management. Especially the uncertainties with respect to consumer (buying) behaviour generate an interdependency on information exchange, and increase the willingness to introduce the category management-concept.

3. *Resource-based view*
   In the search for explaining the *conduct* of enterprises in the agri-business it was also proposed to use the theory of key competences (e.g. Prahalad and Hamel 1990). Key competences are the specific skills ensuring an enterprise performs better than another enterprise and ensuring a sustainable competitive performance. To build up and maintain these key competences, *resources* and *capabilities* are needed. Within the ECR concept, the pooling of complementary resources and capabilities of all partners involved can explain the added value of close cooperation between individual enterprises.

The concepts upon which the ECR approach exist, are partly fully developed, other parts still need further extension and study. The added value of ECR, however, is to be found in the cooperation of all supply chain actors as equal partners in the development, implementation and evaluation of these concepts.

21.5.1 Category management

Category management focuses on product categories for the optimisation of assortments, product introductions and promotions. Food industries and retailers jointly work on providing consumers with the right product that has the right specifications, resulting in the purchase of the products.

In literature many definitions for category management are used. The ECR category management subcommittee defines category management as (ECR Board 1995):

> The process between parts in the chain, where categories are being managed as strategic business units, producing enhanced business results by focusing on delivering consumer value.

A category is defined as *a distinct, manageable group of products/services that consumers perceive to be interrelated and/or substitutable in meeting consumer needs*.

One of the most important goals of category management is the optimisation of the assortment as perceived by the consumer. The central issue for applying category management is the knowledge of and the information on how the consumer perceives the available brands of product in their mutual relation. Within the issues of ECR is category management focused on increasing the efficiency of logistics and on increasing the efficacy to satisfy the consumer.

For category management, a number of tools are available, one of them is the cost approach direct product profitability or DPP. Direct product profitability is basically a quantitative analysis of the assortment, based on profitability calculations. DPP is assumed to be a better indication of profitability than the gross margin (Floor 1996). Indications based on gross margin sometimes give a

completely wrong impression. For example, two different products with the same gross margin can show in the end severely different assets to profitability.

The costs of handling and sale, associated with different products can be very different for each product. In direct product profitability, the contribution to the overall profit is the difference between gross margin and direct product costs. Direct product costs (DPC) are composed of three aspects of costs: the costs made by the distribution centres, costs of transport to and costs at the retailers. Some of the items constituting these costs are invariable while other costs are variable costs. Unlike the variable costs, invariable costs do not increase parallel with the number of commodities sold.

DPP at a retailer can be represented in different ways:

DPP per week	DPP_{week}
DPP per sold consumer unit	DPP_{cu}
DPP per week per m^3	DPP_{vol}

The basic relation for DPP calculations is shown in equation (21.1). This equation expresses that the profitability increases linearly with weekly sales.

$$\begin{aligned} DPP_{week} = &(\text{Gross\_margin} - \text{variable\_costs}) \cdot \text{weekly\_sales} \\ &- \text{invariable\_costs} \end{aligned} \tag{21.1}$$

The three representations of DPP can be converted into one another and are consequently proportional to each other:

$$\begin{aligned} DPP_{cu} &= \frac{DPP_{week}}{\text{weekly\_sales}} \\ DPP_{vol} &= \frac{DPP_{cu}}{\text{shelf\_space}} \end{aligned} \tag{21.2}$$

Costs of transport, costs of distribution centre operation and costs at the retailers through the entire chain are all contained in the variable and invariable costs. DPP can be regarded for one actor in the chain, but can in principle also be used to estimate the DPP of the entire chain. In Table 21.1 an example of the calculation is given.

So, the DPP of a product can be increased by, e.g., a higher rate of sales, a more efficient use of shelf space, an increase in gross margin, a more efficient handling routine or a different way of supplying commodities. Based on DPP analysis, advice can be formulated on a different form or kind of packaging, package content or repackaging.

21.5.2 How are DPPs affected

For calculating DPP values, the most sensitive and important input variable is the weekly sales. The relation between DPP_{cu} and the mean weekly sales is an orthogonal hyperbola as can be taken by combining equations (21.1) and (21.2):

Table 21.1 Shelf efficiency of fresh milk and meat products

	Fresh milk		Meat products	
Price per consumer unity	0.582	€	1.291	€
Purchase price	0.545	€	0.882	€
Gross margin per consumer unity	0.037	€	0.409	€
DPC per consumer unity	0.023	€	0.091	€
DPP per consumer unity	0.014	€	0.318	€
Sales per week	180	#	15	#
DPP per week	2.52	€	4.770	€
Shelf space	0.050	m^3	0.016	m^3
DPP per week per m^3	50.400	€	298.125	€

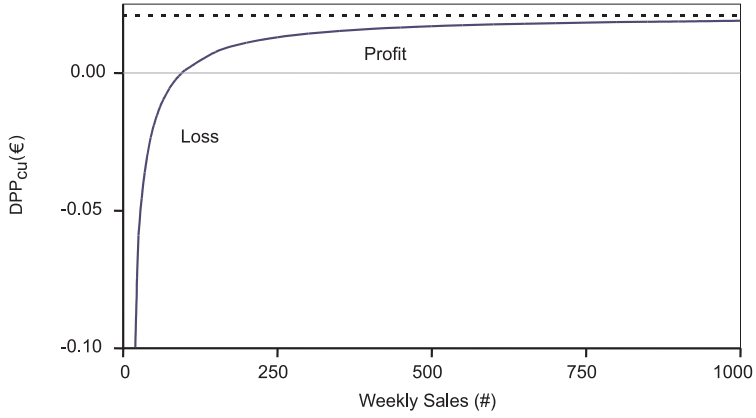


Fig. 21.7 Direct product costs per sold consumer unit versus weekly sales, based on data of Table 21.1, fresh milk with variable costs set arbitrarily to 0.015 € and the invariable costs to 2.0 € per unit.

$$\mathrm{DPP_{cu}} = \mathrm{Gross\_margin} - \mathrm{variable\_costs} - \frac{\mathrm{invariable\_costs}}{\mathrm{weekly\_sales}} \qquad (21.3)$$

It approaches an upper limit at ever-increasing sales, consisting of gross margins minus the variable costs. This upper limit hence represents the maximum $\mathrm{DPP_{cu}}$ that can be obtained at unlimited sales. The intercept at sales = 0 represents the (negative) invariable costs of operation (see Fig. 21.7).

Within the distribution operation the degree of distribution, that is the number of retailers serviced, is of major importance, quite similar to the retailer sales in determining the costs of the distribution centre.

The parcel magnitude also has a major effect on DPP. With increasing parcel magnitude the direct product costs (DPC) and exploitation costs decrease considerably.

DPP analysis constitutes a valuable tool for category management and composition of assortment, but can also assist in the planning of retail shelves,

the optimisation of distribution and the development of price and advertising policies. As such DPP is a valuable tool for managing and optimising the entire supply chain and to support and promote the cooperation between producer and retailer and to increase overall the efficiency in that chain (Stichting Ketenmoduul, 1995).

21.6 Conclusions

Although the number of models that describe the changes of properties and quality attributes and safety in our food steadily increases, market research and economic applications of these models, and especially the development of dedicated economic models is still in its infant stage. A vast effort applying good and sound principles, based on problem decomposition and borrowing knowledge and techniques from other disciplines, has to be undertaken in the near future to achieve structured information on economic effects and processes. Also in the area of consumer research and psychological effects, large gaps in modelled knowledge still exist. All these voids in structured knowledge inherently lead to very diverse levels of understanding within and throughout a supply chain.

The necessity for the global food producing and food retailing companies to achieve high standards for food safety and overall food quality strongly directs the efforts in both the private domain as well as in the public domain.

With 'situations' such as the ongoing BSE crisis in Western Europe or the case of strong consumers' 'dislike' of genetic modified organisms, transparency in the food chain is of utmost importance. In terms of the three schools of thought mentioned above, uncertainty has to be reduced, transaction costs have to be reduced, and new levels of trust have to be achieved.

These changes in network structures are not accomplished by mere input of financial resources from the major food companies. The capability of the (local) actors themselves have to be made available. Awareness, why certain standards on food aspects and food quality have to be implemented, has to grow gradually over time. The public domain – the knowledge infrastructure – has to play its part as well.

The translation of developed standards on food safety and food quality is a major responsibility for research in the food domain. It is important to understand the mechanisms in the sector of food processing, but it is even more important to understand how and with which stimuli an overall higher level of food safety and quality can be achieved. After all, garbage in = garbage out.

Economic theory on chains and networks is useful in trying to understand why economic actors act as they do and where the best opportunities for improvement are. It is also indicated that cooperation can lead to results that are on a higher level than those achieved on the basis of individual performance. With the increasing importance of linking the final consumers directly to the basic production itself, e.g., with the help of Internet, transparency and uniform standards are the only recipe for success.

21.7 References

ANON. (1996): *Efficient Consumer Response Europe*, Value Chain Analysis project overview. Coopers and Lybrand. Utrecht, NL.

BASH K, DAVIES S P, (1998): Vertical integration in agribusiness foreign investment. In : Royer J.S., Rogers R.T. (eds.): *The industrialization of agriculture: vertical coordination in the US food system*. pp. 295–317, Ashgate Publishing Ltd., Aldershot, UK.

BECHTEL C, JAYARAM J, (1997): Supply Chain Management: a strategic perspective, *International Journal of Logistics Management*, 8, 1, 15–33.

BEERS G, BELENS A J M, VAN DALEN J CHR, (1998): Chain science as an emerging discipline. Proceedings 3rd International Conference on Chain Management in Agribusiness and the Food Industry. Department of Management Studies, Wageningen Agricultural University, Wageningen, NL.

BOEHLJE M, SCHRADER L F, (1998): The industrialization of agriculture: questions of coordination. In: Royer J.S., Rogers R.T. (eds.): *The industrialization of agriculture: vertical coordination in the US food system*. pp. 3–26, Ashgate Publishing Ltd., Aldershot, UK.

BOEHLJE M, AKRIDGE J, DOWNEY D, (1995): Restructuring Agribusiness for the 21st century, *Agribusiness*, 11, 6.

COOPER M C, LAMBERT D M, PAGH J D, (1997): Supply Chain Management: more than a new name for logistics, *International Journal of Logistics Management*, 8, 1, 1–13.

ECR BOARD, (1995): ECR Europe Executive Board Vision Statement 1995. http://www.cs.vu.nl/~sbhulai/ecr/cm.html#what.

ELLRAM L M, (1991): Supply Chain Management: the industrial organisation perspective, *International Journal of Physical Distribution and Logistics Management*, 21, 1, 13–22.

FLOOR K, (1996): *Trade Marketing; de strijd om de schaarse schapruimte*. (In Dutch: Trade Marketing; The battle for scarce shelf space). Academic Service, Schoonhoven, NL.

FORD D I, (ed.) (1990): Understanding business markets: Interaction, relationships, networks. Academic Press, London, UK.

HAAKANSSON H, SNEHOTA L, (1989): No Business is an island: The network concept of business strategy. *Scandinavian Journal of Management* **2**, 187–200.

JARILLO J C, (1988): On strategic networks. *Strategic Management Journal* **9**, 31–41.

KURT SALMON ASSOCIATES, (1993): *Efficient Consumer Response; Enhancing Consumer Value In The Grocery Industry*, Food Marketing Institute, Washington D.C., USA.

LAMBERT D M, COOPER M C, PAGH J D, (1998): Supply Chain Management: implementation issues and research opportunities, *International Journal of Logistics Management*, 9, 2, 1–19.

MACFIE H J H, THOMSON D M H, (1994): *Measurement of food preferences*. Blackie Academic & Professional, London, UK.

MALLEN B, (1977): *Principles of marketing channel management: interorganizational distribution design and relations*. Lexington Books, D.C. Heath and Company; Lexington, Massachusetts, USA.

MEISELMAN H L, MACFIE H J H, (1996): *Food choice, acceptance and consumption*. Blackie Academic & Professional, London, UK.

MEULENBERG M T G, BROENS D F, (1997): Trends in de productie van voedingsmiddelen: Inzicht in ketens helpt bij ketenvorming. *Voedingsmidelen Technologie (VMT)* **29**, 19–25 (in Dutch: Trends in food production: understanding chains assists in chain formation).

NOHRIA N, ECCLES R G, (eds.) (1992): Networks and organizations: Structure, form and action. Harvard Business School, Boston, USA.

OVERBOOM M A, (2000): Analysing governance structure of international supply chains. In: Triekenens J. H., Zuurbier P. J. P. (eds.) Proceedings of the *Fourth International Conference on Chain Management in Agribusiness and the Food Industry*. Management Study Group, Wageningen Agricultural University, Wageningen Press, NL.

PFEFFER J, SALANCIK G R, (1978): *The external control of organisations*. Harper & Row Publishers, New York.

PORTER M E, (1980). *Competitive Strategy: Techniques for analysing industries and competitors*. The Free Press, New York.

PORTER M E, (1985). *Competitive Advantage*. The Free Press, New York.

PORTER M E, MILLAR V E, (1985). How information gives you competitive advantage. *Harvard Business Review*, **63**, July–August.

PRAHALAD C K, HAMEL G, (1990): *The core competence of the corporation*. Harvard Business School, USA.

SHEWFELT R L, TIJSKENS L M M, (2000): A more integrated view. In Shewfelt R. L., Brückner B. (eds.): *Fruit and vegetable quality: an integrated view.* pp. 296–306, Technomic Press, USA.

STEVENS G C, (1989): Integrating the supply chain. *International Journal of Physical Distribution and Materials Management*, 19, 8, 3–8.

STICHTING KETENMODUUL, (1995): Uniforme Nederlandse DPP-basismodel. Manual Koppeling Spaceman – DPP versie 2.0 (in Dutch). August 1995, A.C. Nielsen, Amsterdam, NL.

THOMAS J M, STAATZ J M, PIERSON T R, (1995): Analysis of grocery buying and selling practices among manufacturers and distributors: implications for industry structure and performance. *Agribusiness*, New York. **11: 6**, 537–51.

THORELLI H B, (1986): Networks: between markets and hierarchies. *Strategic Management Journal* **7**, 37–51.

TOWILL D R, (1996): Time compression and Supply Chain Management – A Guided Tour, *Supply Chain Management*, 1, 1, 15–27.

TRIEKENENS J H, ZUURBIER P J P, (eds.) (2000): Proceedings of the Fourth International Conference on *Chain Management in Agribusiness and the Food Industry*. Management Study Group, Wageningen Agricultural University, Wageningen Press, NL.

VAN TRIJP H C M, MEULENBERG M T G, (1996): Marketing and consumer behaviour with respect to foods. In: Meiselman H. L., MacFie H. J. H. (eds): *Food choice, acceptance and consumption*. Blackie Academic & Professional, London, UK.

VAN DALEN J C, (1994). Geboeid door ketens (Fascinated by chains). Inaugural speech, Wageningen University, Wageningen NL.

VERHAEGEN I, HUYLENBROECK G VAN, VAN HUYLENBROECK G, (1999): A transaction cost analysis of a small innovative marketing channel for beef in Belgium. *Journal of International Food and Agribusiness Marketing*. **10: 3**, 1–17.

WIERENGA B, SOETHOUDT H, (2000): How to Create a Win-Win Situation for Retailer and Producer in Supermarket Sales Promotions. Proceedings Marketing Science Conference, *Marketing for the New Millennium*, The Anderson School at UCLA, Los Angeles, California, 67–68.

ZIGGERS G W, TRIEKENENS J H, ZUURBIER P J P, (eds.) (1998): Proceedings of the Third International Conference on *Chain Management in Agribusiness and the Food Industry*. Management Study Group, Wageningen Agricultural University, NL.

ZUURBIER P J P, TRIEKENENS J H, ZIGGERS G W, (1996): *Verticale samenwerking* (in Dutch: Vertical cooperation). Kluwer Bedrijfswetenschappen, Deventer, NL.

ZYLBERSZTAJN D, GOLDBERG R A, GOLDBERG R A, (1996): Governance structures and agribusiness coordination: a transaction cost economics based approach. *Research in domestic and international agribusiness management*, **12**, 245–310.

Appendix: Notation

Chapter 1

Symbol	*Dimension*	*Definition*
τ	s	lag time
[x]	mol/l	concentration of compound x
$<>$	–	mean
A	–	probability of growth
a_w	–	water activity
c	–	dimensionless concentration of carbon resources
C_p	$J\,kg^{-1}$	heat of vaporisation
D	$m^2\,s^{-1}$	diffusion coefficient
D	s	decimal reduction time for death
F_i, f_i	–	unnamed functions
h	–	dimensionless concentration of hydrogen ions
K	$m^2\,s^{-1}$	thermal diffusion coefficient
k	s^{-1}	growth rate constant $= 1 < \tau >$
m	–	dimensionless biomass
n	–	number
N	–	number
o	–	dimensionless concentration of oxygen
p	–	probability
Q	–	intensity of quality factor
r	m	dimension in x, y, z plane
t	s	time
T	K or ºC	temperature
W		water content

Index	*Definition*
exp	exponential growth
eff	effective
d	doubling
inj	injury
res	resuscitation
lag	lag phase
0	Initial

Chapter 2

Symbol	*Dimension*	*Definition*
AC	mol/l	active complex
E	mol/l	concentration of enzyme
E_a	J/mol	energy of activation
k	*	reaction rate constant
K_m	mol/l	Michaelis Menten equilibrium constant
P	mol/l	concentration of product formed
S	mol/l	concentration of substrate
t	s	time
T	ºC or K	temperature
V_{max}		maximal rate
	*	depending on order of reaction: 1st order 1/s 2nd order mol/l/s

Index	*Definition*
0	initial
abs	absolute
p	of product formation
ref	at reference temperature
*s*1	of forward reaction
*s*2	of backward reaction

Chapter 3

Symbol	*Dimension*	*Definition*
ν	–	frequency
ρ	–	correlation coeffcient
σ_I^2	–	variance of component I
ΔG	J	Gibbs energy
ΔH	J/mol	enthalpy
σ_{I}	–	standard deviation of component I

σ_{ij}	–	covariance of components i and j
σ^{ij}	–	elements of the inverse of the matrix $\mathbf{Cov_{ee}}$
ΔS	J/mol/K	entropy
[A]	mol/l	concentration reagent
[AB]	mol/l	concentration reaction product
[B]	mol/l	concentration reagent
[C]	mol/l	concentration reagent
[E]	mol/l	enzyme concentration in an enzymatic reaction
[ES]	mol/l	concentration of the enzyme-substrate complex
$[H^+]$	mol/l	concentration hydrogen ions
[S]	mol/l	substrate concentration in an enzymatic reaction
[S]	mol/l	concentration substrate
A	*	pre-exponential factor
c	mol/l	concentration of a component
c_{ij}	–	elements of the dispersion matrix **C** (multiresponse modelling)
$\mathbf{Cov_{ee}}$	–	covariance matrix of experimental errors
c^θ	mol/l	concentration in standard state
E_a	J/mol	activation energy
h	Js	Planck's constant (6.626×10^{-34} J s)
k	*	reaction rate constant
K	*	equilibrium constant
k'	*	pseudo first order reaction rate constant
k_B	$1.3807E^{-23}$ J/K	Boltzmann's constant
m	–	molecularity of a reaction
n	–	order of reaction
P	mol/l	concentration reaction product
pH	–	$-\log_{10}([H^+])$
R	8.314 J/mol/K	gas constant
S		substrate
t	s	time
T	K or ºC	absolute temperature
u	–	number of experimental runs (multiresponse)
v	1/s	maximum velocity of an enzymatic reaction
y_I	–	experimental datum points for component i
$\hat{y}$	–	model prediction for component i
	*	depending on order of reaction: 1^{st} order 1/s 2^{nd} order mol/l/s

Index	*Definition*
‡	activated comples
0	initial
1, 2, 3	number of reaction
eq	equilibrium

M	Michaelis constant
max	maximal
T	total
c	with respect to concentration
t	with respect to time
ref	at reference temperature

Chapter 4

Symbol	*Dimension*	*Definition*
A	m^2	surface area
$\mathbf{A}$	$\mathrm{kg\,s}^{-1}$	coefficient matrix
a, b		constants
A_p	m^2	particle surface area
Bi		Biot number
c	$\mathrm{J\,kg}^{-1}\,{}^{\mathrm{o}}\mathrm{C}^{-1}$	heat capacity
$\mathbf{C}$	$\mathrm{J}\,{}^{\mathrm{o}}\mathrm{C}^{-1}$	capacity matrix
C_a	$\mathrm{mol\,m}^{-3}$	molar concentration of species component a
c_m	$\mathrm{kg\,kg}^{-1}\,{}^{\mathrm{o}}\mathrm{M}^{-1}$	moisture capacity
c_p	$\mathrm{J\,kg}^{-1}\,{}^{\mathrm{o}}\mathrm{C}^{-1}$	heat capacity of particle
C_η		*turbulence constant*
D_a	$\mathrm{m}^2\,\mathrm{s}^{-1}$	mass diffusion coefficient of species component a
E	$\mathrm{J\,mol}^{-1}$	activation energy
$\mathbf{f}$	W	thermal load vector
f_I	$\mathrm{N\,m}^{-3}$	volumetric body force
F_I	N	Cartesian component of particle force
Fo		Fourier number
h	$\mathrm{W\,m}^{-2}\,{}^{\mathrm{o}}\mathrm{C}^{-1}$	surface heat transfer coefficient
H	$\mathrm{J\,kg}^{-1}$	static enthalpy
h_{fg}	$\mathrm{J\,kg}^{-1}$	latent heat of evaporation
h_{ma}	$\mathrm{m\,s}^{-1}$	surface mass transfer coefficient
k	$\mathrm{W\,m}^{-1}\,{}^{\mathrm{o}}\mathrm{C}^{-1}$	thermal conductivity
K	$\mathrm{m}^2\,\mathrm{s}^{-2}$	turbulent kinetic energy
$\mathbf{K}$	$\mathrm{W}\,{}^{\mathrm{o}}\mathrm{C}^{-1}$	stiffness matrix
$k_{f,b}$		forward (f) or backward (b) rate constant
k_m	$\mathrm{kg\,m}^{-1}\,\mathrm{s}^{-1}\,{}^{\mathrm{o}}\mathrm{M}^{-1}$	moisture conductivity
L	m	characteristic length
M	$\mathrm{kg\,mol}^{-1}$	molecular weight
m, n, o, p, q, r		exponents
m_p	kg	mass of particle
$\mathbf{N}$		shape function
$n_\perp$		outward normal to surface
p	Pa	pressure

Q	$\mathrm{W\,m^{-3}}$	volumetric heat generation
$\mathbf{Q}$	units $\mathrm{s^{-1}}$	source term vector
R	$\mathrm{J\,mol^{-1}\,K^{-1}}$	universal gas constant
r		residual
r'_a	$\mathrm{kg\,m^{-3}\,s^{-1}}$	rate of production of species component a
r_a	$\mathrm{mol\,m^{-3}\,s^{-1}}$	rate of production of species component a
R_c	$\mathrm{mol\,s^{-1}}$	reaction rate
S_ϕ	units $\mathrm{m^{-3}\,s^{-1}}$	source of quantity ϕ
T	°C	temperature
t	s	time
$\mathbf{U}$	$\mathrm{m\,s^{-1}}$	velocity vector
$\mathbf{u}$		nodal temperature vector
u_j	$\mathrm{m\,s^{-1}}$	Cartesian velocity component
u_{pi}	$\mathrm{m\,s^{-1}}$	Cartesian component of particle velocity
V	$\mathrm{m^3}$	volume
X_a		mass fraction
x_i, x, y, z	m	Cartesian coordinate

Greek

Ω		object domain
Γ		boundary of object
Γ	$\mathrm{kg\,m^{-1}\,s^{-1}}$	diffusivity of quantity ϕ
ϕ	units $\mathrm{kg^{-1}}$	transported quantity per unit mass
Ψ	°M	moisture potential
δ	$\mathrm{^oC^{-1}}$	thermo gradient
ζ		ratio of vapour diffusion to total moisture diffusion
ρ	$\mathrm{kg\,m^{-3}}$	density
ε		emission coefficient
	$\mathrm{m^2\,s^{-3}}$	turbulent energy dissipation rate
σ	$\mathrm{W\,m^{-2}\,K^{-4}}$	Stefan-Boltzmann constant
ϑ		dimensionless temperature
η	$\mathrm{kg\,m^{-1}\,s^{-1}}$	dynamic viscosity
ε_c		convergence error

Index	*Definition*
0	initial condition
∞	ambient condition

Chapter 5

Symbol	*Dimension*	*Definition*
$\partial^2 J/\partial P_i\, \partial P_j$		partial derivative of juiciness for all combinations of the i^{th} and j^{th} model parameter
C_i		i^{th} model copy
E_i		parameter estimate of the i^{th} model parameter
J		juiciness
P_i		i^{th} model parameter

Index	*Definition*
i, j	running index

Chapter 8

Symbol	*Dimension*	*Definition*
E		Expected value
F		Fisher information matrix
J		Jacobian matrix
M		Model structure
MSE		Mean square error
N	CFU/ml	Cell density
n		Number of observations
n_0	ln CFU/ml	natural logarithm of the initial cell density
n_{t_s}		Number of sampling moments
n_{r_i}		Number of replicate measurements
n_t		Total number of data points
n_p		Number of model parameters
n_{max}	ln CFU/ml	Natural logarithm of the asymptotic cell density
P		Covariance matrix
p		Parameter vector
$\mathbf{p}_{ls}$		Least square estimated parameter vector
$\mathbf{p}_{\mathrm{ml}}$		Maximum likelyhood parameter vector
SSE		Sum of squared errors
s^2		Variance
T_{min}	°C	Theoretical minimum temperature
T_{opt}	°C	Theoretical optimum temperature
T_{max}	°C	Theoretical maximum temperature
t_i	s, min or h	Sampling moments
y_I		ith observation of n_r replicate measurements
$\overline{y}$		Mean of replicate measurements
w		Weighted factor/value

Greek		
μ_{opt}	1/h	Maximum specific growth at T_{opt}
μ_{max}	1/h	Maximum specific growth rate
λ	h	Lag time
π		Probability density function
σ^2		Variance
$\Im_I$		Identification cost

Index	*Definition*
mean	mean

Chapter 9

Symbol	*Dimension*	*Definition*
DM		dry matter
DMC	%	dry matter content
d*W*/d*t*		rate of dry matter production
LAI		leaf area index
P_g		canopy gross photosynthesis
Rm		maintenance respiration
$\mathbf{Y_g}$		growth conversion efficiency

Chapter 10

Symbol	*Dimension*	*Definition*
C		air CO_2 concentration
DSS		decision support systems
J		objective function
L		accumulative cost
LAI		leaf area index
P_{max}		maximum rate of leaf photosynthesis
PAR		photosynthetic active radiation
p		perturbation vector
RUE		radiation use efficiency
R_g		growth respiration
R_m		maintenance respiration
s		weight of structural dry matter
T		plant air temperature
t		time
u(t)		control vector
VPD		vapour pressure deficit
x (*t*)		state vector of the plant
w		dry weight

Chapter 11

Symbol	*Dimension*	*Definition*
CP	$g.\,head^{-1}.day^{-1}$	crude protein content
DM		dry matter
DMTP	$g.\ head^{-1}.day^{-1}$	digestible true protein requirement
DUP	$g.\,kg^{-1}$	rumen indigestible protein digested in the lower intestines
ERDP	$g.\,kg^{-1}$	effective rumen degradable protein
FME	$MJ.\ head^{-1}.day^{-1}$	fermentable metabolisable energy
MCP	$g.\,head^{-1}.day^{-1}$	microbial crude protein content
MP	$g.\,kg^{-1}$	microbial protein
MTP	$g.\,head^{-1}.day^{-1}$	microbial true protein content
n		number of observations
QDP	$g.\,kg^{-1}$	quickly degradable protein
SDP	$g.\,kg^{-1}$	slowly degradable protein
u		Theil's inequality coefficient
UDP	$g.\,kg^{-1}$	rumen indigestible protein

Chapter 12

Symbol	*Dimension*	*Definition*
A	ha	farm area
AE		adult equivalent, energy requirement of a cattle class relative to that of an adult animal
Ai		coefficient relating the numbers of animals in cattle class *i* to the number of cows mated
BRi		ratio of the number of animals in the *i*th cattle class to the number of breeders when survival in the class is 100%
CC		farm carrying capacity
CFi		conversion factor to adult equivalents for cattle class *i*
CM		number of cows mated
CULL		preferred yearly replacement percentage of the number of cows mated
I	$kg.\,year^{-1}$	annual intake for an animal unit
LW		liveweight
Ni		number of cows in the remaining cattle classes
PFi		boolean indicating whether cattle class *i* is present
PG	$kg.\,ha^{-1}$	pasture growth
R	$mm.\,year^{-1}$	effective rainfall
SRi		proportion of survival in cattle class *i*

U		safe utilisation
WR		weaning rate
WU	$mm.day^{-1}$	water use over a specified time step
WUE	$kg.ha^{-1}.mm^{-1}$	water use efficiency

Chapter 13

Symbol	*Dimension*	*Definition*
a	m^2/m^3	specific surface area
C_s	mmol/l	actual substrate concentration
k_l		resistance for oxygen
K_s	mmol/l	affinity constant for substrate
t_m		time required to realise 95% of the final concentration in the liquid
Greek		
μ	1/h	specific growth rate of the organism
μ_{max}	1/h	Maximum specific growth of the organism

Chapter 14

Symbol	*Dimension*	*Definition*
p	kPa	CO_2 partial pressure
r		rate of consumption or production by the product
r_{O_2}		O_2 consumption by the product
RQ		respiration quotient; ratio of CO_2 production to O_2 consumption
Index		*Definition*
O_2		oxygen
CO_2		carbon dioxide

Chapter 15

Symbol	*Dimension*	*Definition*
a_w		surface water activity
$\dot{m}$	$kg/(m^2.s)$	water flux
Bi		Biot number
$\boldsymbol{C}$		global capacitance vector
CFD		computational fluid dynamics
C_p	J/(Kg.K)	heat capacity

D_p	m	penetration depth
D_{wa}	m^2/s	diffusivity of water vapour in air
E		EHTD defined as the ratio of the freezing time of an infinite slab of same thickness as the smallest dimension of the object to the freezing time of the object
$\boldsymbol{f}$		thermal load vector
H	J/Kg	enthalpy
h	$W/(m^2.K)$	heat transfer coefficient
H_{fg}	J/Kg	latent heat of evaporation
$\boldsymbol{K}$		global conductance matrix
k	W/(m.K)	thermal conductivity
$K_{i,a}$	W/(m.K)	thermal conductance between tank I and the environment
$K_{i,j}$	W/(m.K)	thermal conductance between tank j and i
k_y	m/s	mass transfer coefficient
n		number of tanks
ODEs		ordinary differential equations
PDE		partial differential equation
q	W/m^2	heat flux
R	m	product's smallest half-dimension
r_f	m	frozen depth
$\boldsymbol{T}$	K	nodal temperature vector
V	m^3	product volume
w	Kg/Kg	moisture content
Y	Kg/Kg	absolute humidity
Z		geometric parameter
ΔH_2	J/Kg	latent plus post cooling heat

Greek

μ		mass transfer resistance factor
ρ	Kg/m^3	density
ν		mass flow correction

Index	*Definition*
a	air
f	frozen
I	nodal value
m	intermediate value
new	at the end of a time step
old	at beginning of a time step
s	surface
sat	value at saturation
w	water

Chapter 16

Symbol	*Dimension*	*Definition*
B		flux density
Bi		biot number
C	1/ml	bacterial concentration
C_p	J/(Kg.K)	specific heat
D	C/m^2	electric flux density
d	m	characteristic dimension
E	V/m	electric field
E	V/m	voltage gradient
F	s	lethality
f	Ghz	frequency
G		shape factor
g	m/s^2	gravity
H	A/m	magnetic field
H	J/kg	specific enthalpy function
h	$W/(m^2.K)$	interfacial heat transfer coefficient
k_t	1/s	reaction rate
L	m	axial length
N		number of bacteria
n		flow behaviour index
P	Pa	pressure
PEPT		positron emitting particle tracking
Q	W/m^3	amount of heat generation per unit volume
R_Q		heating rate
r	m	radial direction
t	s	time
T	K	temperature
u	m/s	velocity component
V	V	voltage
Z	K	increase in temperature
z	m	axial direction

Greek		
α	m^2/s	thermal diffusivity
ε'		dielectric constant
ε''		dielectric loss
μ	Kg/(m.s)	viscosity
λ	W/(m.K)	thermal conductivity
ρ	Kg/m^3	density
	m/s	velocity component
ζ		dimensionless length
ϕ_{act}		porosity
σ	s/m	electrical conductivity

Index	*Definition*
f	fluid
ref	reference
s, p	solid particle
w	wall

Chapter 17

Symbol	*Dimension*	*Definition*
n	–	number
Q	arbitrary	quality
W	–	weight of attribute category
w	–	weight of individual attributes
z	–	normalised attribute intensity

Index	*Definition*
i, j	of component i, j
cat	category

Chapter 18

Symbol	*Dimension*	*Definition*
τ_I		lag for the *i*th cell of the initial population
μ		maximum specific growth rate
A		accuracy factor
B		bias factor
c_{max}		maximum population concentration
E		environment
h_0		amount of 'work' required to prepare for exponential growth
K_P		Michaelis-Menten saturation constant
L		link function
m		curvature parameter characterising the transition to the stationary phase
N		the initial number of a growing cell population
P		a critical substance to initiate growth
p		growth parameter
t		time
T		temperature
v_n		*n*th environmental factor
x		bacterial concentration
x_0		bacterial concentration at time zero
y		natural logarithm of the bacterial concentration

y_{max}	natural logarithm of the maximum population density
z_I	Change in the *i*th environmental factor that induces a two-fold increase in the value of the maximum specific growth rate keeping the other environmental factors constant
$\alpha(t)$	monotone increasing function with values between 0 and 1
λ	duration of lag
ν	rate of the 'work' required to prepare for exponential growth

Chapter 19

Symbol	*Dimension*	*Definition*
ΔE		chromaticity change
A		desirable quality factors
B		undesirable quality factors
C		chroma
C_i		compositional factor
CTI		critical temperature indicator
CTTI		critical temperature-time integrator
E_A	$J.mol^{-1}$	activation energy
E_{AI}	$J.mol^{-1}$	activation energy of the TTI response
E_j		environmental factor
$f(A)$		quality function
$F(X)$		response function for a TTI
FIFO		first in first out system
k		apparent reaction rate
k_I		TTI response rate constant
$k_{I\ ref}$		TTI response rate constant at reference temperature
k_{ref}		reaction rate constant at reference temperature
LSFO		least shelf-life first out
m		apparent reaction order
Q		quality
r		rate
R	$J.K^{-1}.mol^{-1}$	universal gas constant
SLDS		shelf-life decision system
SLR		remaining shelf-life
t		time
T	K	absolute temperature
T_{eff}	K	effective temperature
$T_{eff(TTI)}$	K	effective temperature for the TTI

T_{ref}	K	reference temperature
TTI		time-temperature integrator or indicator
X_c		normalized chroma

Chapter 20

Symbol	*Dimension*	*Definition*
MAP		modified atmosphere packaging
WMS		warehouse management system

Chapter 21

Symbol	*Dimension*	*Definition*
DPP	€	direct product profitability
DPC	€	direct product costs
weekly_sales	–	number of units sold per week
shelf_space	m^3	volume occupied at the shelf per unit

Index	*Definition*
week	per week
cu	per consumer unit
vol	per volume

Index