The Computerized Greenhouse

Automatic Control Application in Plant Production

Edited by
Y. Hashimoto
G.P.A. Bot
W. Day
H.-J. Tantau
H. Nonami

The Computerized Greenhouse

The Computerized Greenhouse

Automatic Control Application in Plant Production

Edited by

Yasushi Hashimoto
Department of Biomechanical Systems
College of Agriculture
Ehime University
Matsuyama, Japan

Gerard P. A. Bot
Institute of Agricultural Engineering (IMAG-DLO)
Wageningen, The Netherlands

W. Day
Process Engineering Division
Silsoe Research Institute
Silsoe, United Kingdom

H.-J. Tantau
Institute for Horticultural Engineering
University of Hannover
Hannover, Germany

Hiroshi Nonami
Department of Biomechanical Systems
College of Agriculture
Ehime University
Matsuyama, Japan

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
San Diego New York Boston London Sydney Tokyo Toronto

ACADEMIC PRESS RAPID MANUSCRIPT REPRODUCTION

This book is printed on acid-free paper. ∞

Academic Press, Inc.
1250 Sixth Avenue, San Diego, California 92101-4311

United Kingdom Edition published by
Academic Press Limited
24–28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

The Computerized greenhouse: automatic control
application in plant production / edited by Y. Hashimoto . . . [et al.].
p. cm.
ISBN 0-12-330590-X
1. Greenhouses—Automatic control—Data processing.
I. Hashimoto, Yasushi, DATE
SB416.C655 1993
635.9'8283'0285–dc20 92-41640
CIP

PRINTED IN THE UNITED STATES OF AMERICA
93 94 95 96 97 98 BC 9 8 7 6 5 4 3 2 1

Contents

Contributors vii
Preface ix

Section 1: System, Modelling, Identification and Control

1. Identification, Estimation and True Digital Control of Glasshouse Systems 3
P. C. Young, A. Chotai, and W. Tych

2. Physical Modelling of Greenhouse Climate 51
G. P. A. Bot

3. Measurement and Parameter Identification Problems for Control of Biotechnical Processes 75
A. Munack

4. Environmental Control in Plant Tissue Culture and Its Application for Micropropagation 95
T. Kozai and B. R. Jeong

Section 2: Optimal Control

5. Optimal Diurnal Climate Control in Greenhouses as Related to Greenhouse Management and Crop Requirements 119
H. Challa and G. van Straten

6. Optimal Control for Plant Production in Greenhouses 139
H.-J. Tantau

7. Optimal Greenhouse Temperature Trajectories for a Multi-State-Variable Tomato Model 153
I. Seginer and A. Sher

Section 3: Computer Applications and Artificial Intelligence

8. Computer Integrated System for the Cultivating Process in Agriculture and Horticulture — Approach to "Intelligent Plant Factory" 175

Y. Hashimoto

9. Using Empirical Knowledge for the Determination of Climatic Setpoints: An Artificial Intelligence Approach 197

R. Martin-Clouaire, T. Boulard, M.-J. Cros, and B. Jeannequin

10. Computer Opportunities in Agriculture and Horticulture 225

W. Day

Section 4: Relating Plant Sciences

11. Environmental Signals, Receptors, Physiological and Genetic Regulation in Plants 243

A. W. Naylor

12. Physiological Regulation of Water Transport in Controlled Environments 267

J. S. Boyer and H. Nonami

13. Growth Regulation in Plant Factories and Greenhouses from a Physiological Viewpoint 303

H. Nonami

Index 333

Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

G. P. A. Bot (51), Institute of Agricultural Engineering, (IMAG-DLO), Wageningen, 6700 AA, The Netherlands

T. Boulard (197), INRA, Station de Bioclimatologie, 84140 Montfavet, France

J. S. Boyer (267), College of Marine Studies, University of Delaware, Lewes, Delaware 19958.

H. Challa (119), Department of Horticulture, Wageningen Agricultural University, 6708 PM Wageningen, The Netherlands

A. Chotai (3), Centre for Research on Environmental Systems, IEBS, University of Lancaster, Lancaster, LA1 4YQ, United Kingdom

M.-J. Cros (197), INRA, Laboratoire d'Intelligence Artificielle, BP27, 31326 Castanet-Tolosan, France

W. Day (225), Process Engineering Division, Silsoe Research Institute, Wrest Park, Silsoe, Bedford, United Kingdom

Y. Hashimoto (175), Department of Biomechanical Systems, College of Agriculture, Ehime University, Tarumi, Matsuyama 790, Japan

B. Jeannequin (197), INRA, Domaine Horticole d'Alenya, Le Mas blanc, 66200 Elne, France

B. R. Jeong (95), College of Horticulture, Chiba University, Matsudo, Chiba 271, Japan

T. Kozai (95), College of Horticulture, Chiba University, Matsudo, Chiba 271, Japan

R. Martin-Clouaire (197), INRA, Laboratoire d'Intelligence Artificielle, BP27, 31326 Castanet-Tolosan, France

A. Munack (75), Federal Agricultural Research Centre, FAL, D-3300 Braunschweig, Germany

A. W. Naylor (243), Department of Botany, Duke University, Durham, North Carolina 27706

H. Nonami (267, 303), Department of Biomechanical Systems, College of Agriculture, Ehime University, Tarumi, Matsuyama 790, Japan

I. Seginer (153), Technion, Department of Agricultural Engineering, Haifa 32000, Israel

A. Sher (153), Technion, Department of Agricultural Engineering, Haifa 32000, Israel

H.-J. Tantau (139), Institute for Horticultural Engineering, University of Hannover, Herrenhöuserstr 2, D-3000 Hannover 21, Germany

W. Tych (3), Centre for Research on Environmental Systems, IEBS, University of Lancaster, Lancaster, LA1 4YQ, United Kingdom

G. van Straten (119), Systems and Control Group, Department of Agricultural Engineering and Physics, Wageningen Agricultural University, Bomenweg 4, 6703 HD Wageningen, The Netherlands

P. C. Young (3), Centre for Research on Environmental Systems, IEBS, University of Lancaster, Lancaster, LA1 4YQ, United Kingdom

Preface

The aim of this book is to open up new research areas in automatic control engineering for crop production systems in horticulture and agriculture and to explain the principal ideas of advanced automatic control in plant production to engineers, plant scientists, agronomists, and horticulturists.

This book had its beginnings in the IFAC/ISHS Workshop on Mathematical and Control Application in Agriculture and Horticulture, which was held between September 30 and October 3, 1991 in Matsuyama, Japan. This workshop was unique in that it was an international meeting sponsored by two major international academic organizations dealing with completely different fields of study, i.e., the International Federation of Automatic Control (IFAC) and the International Society for Horticultural Science (ISHS). This shows that a new interdisciplinary study has begun to emerge in recent years, and this meeting had a significant role in forming a framework for a new research area in the "new" agriculture and horticulture. The workshop brought together scientists from laboratories in various countries to exchange information about control applications in horticulture and agriculture and to discuss new problems and new technology to solve them.

This book expands upon results obtained in the workshop and provides generalized applications of the topic discussed. In addition, chapters in the book include reviews of related topics. Thus the book is of interest not only to specialists in control engineering but to students who may be interested in learning the history of the development of control applications in agriculture. It should also be of interest to plant scientists who are working toward practical applications of the techniques for crop management that are described in the book.

Furthermore, the book addresses mathematical applications, control engineering and information science in relation to a wide range of issues in agricultural and horticultural science, including the use of computers and related topics in plant sciences. Control applications in the production of horticultural crops in greenhouses is a particular focus. Control and automation in this industry have close parallels with process and factory automation in the manufacturing industries.

In Section 1, fundamental engineering concepts of system, modelling, identification and control for agricultural systems are introduced. Further, such ideas are applied to the field of biotechnology.

The most widely-used practical applications for engineering concepts in agriculture involve optimal control. Thus, optimal control and its applications are discussed in Section 2.

Although sophisticated high technology has not yet been used in practical agricultural production, several attempts have been made in recent years by using computers for artificial intelligence. An idea for using artificial intelligence for agricultural applications is introduced in Section 3.

Engineering for agricultural production optimizes the physiological function of plants grown in agricultural facilities. In Section 4, control mechanisms of plants are discussed from the viewpoints of plant physiologists.

In this book, concepts of automatic control are derived from both engineering viewpoints and plant physiological viewpoints. It is hoped that cooperation between such vastly different scientific fields will lead to fully computerized greenhouses. When computerized greenhouses are realized in the future, efficient use of energy and labor will increase drastically, and the quality of the product will also be improved. We hope that this book will help to develop a new field of science from the fields of engineering and agriculture.

Yasushi Hashimoto
Hiroshi Nonami

SECTION 1
System, Modelling, Identification and Control

Chapter 1

IDENTIFICATION, ESTIMATION AND TRUE DIGITAL CONTROL OF GLASSHOUSE SYSTEMS

Peter Young,
Arun Chotai and Wlodek Tych

Centre for Research on Environmental Systems,
Institute of Environmental and Biological Sciences,
University of Lancaster,
Lancaster LA1 4YQ
England.

1. INTRODUCTION

Although the rapid development of the digital computer during the 1970's stimulated the move towards Direct Digital Control (DDC), it failed to make too many inroads into the underlying faith of most traditional control systems designers in continuous-time system design methods. And even today, despite the tremendous developments in microcomputers over the 1980's and the now common use of microprocessors for the implementation of control systems, the majority of practical automatic digital controllers used in industry are still based on the mechanical digitisation of continuous-time designs, such as the ubiquitous Proportional-Integral-Derivative (PID) controller. In this paper, we try to promote an alternative philosophy; one based on the idea of True Digital Control (TDC), as propounded in a number of our previous publications (Young et al, 1987b, 1988, 1991a; Young, 1989, 1991; Chotai et al, 1990, 1991b), and show how it can be applied to the modelling and control of glasshouse systems.

The TDC philosophy rejects the idea that a digital control system should be initially designed in continuous-time terms. Rather, it demands that the control systems analyst considers the design from a digital, sampled data standpoint, even when fast sampling, near continuous-time operation is required. The TDC design procedure consists of the following three major steps:-

(1) **Recursive identification and estimation** of discrete-time models based on the analysis of either planned experimental data, or via model reduction from data generated by (usually high order) continuous or discrete-time simulation models.

(2) **Off-line TDC system design and initial evaluation** based on the models from step (1), using an iterative application of an appropriate discrete-time design methodology, coupled with closed loop sensitivity analysis based on Monte-Carlo stochastic simulation.

(3) **Implementation, testing and evaluation** of the control system on both simulated and the real process; in the case of self-tuning or self-adaptive control, employing on-line versions of the recursive estimation algorithms used in step (1).

The TDC design philosophy is best exemplified in a new control systems design procedure developed at Lancaster over the past few years. This is based on the definition of a special Non-Minimum State-Space (NMSS) representation of systems whose dynamic behaviour can be characterised by transfer functions in either the backward shift (z^{-1}) or the discrete-differential (δ) operator. Although simple to implement in practice, the resultant Proportional-Integral-Plus (PIP) controller is attractive because it exploits the power of State Variable Feedback (SVF). Moreover, since it has been developed directly within the context of glasshouse horticulture, we believe that the PIP approach has particular potential for application to horticultural and agricultural systems.

Later, we explore the practical potential of PIP by describing how fixed gain and self-adaptive PIP systems have been applied to the control of environmental temperature in a glasshouse and nutrient levels in a Nutrient Film Technique (NFT) system used in glasshouse horticulture. First, however, we outline the theoretical background to the various identification, estimation and control procedures on which these PIP system designs are based and briefly describe a Computer Aided Control System Design (CACSD) package which has been developed for a wide range of computer platforms and is designed to facilitate the practical application of the TDC procedures.

1.1 PIP vs Conventional Control System Design

Most conventional control systems used in horticulture and agriculture are either analog or utilise some form of Direct Digital Control (DDC). In the latter case, the control system design is normally in the form of digitised versions of Proportional–Integral–Derivative (PID) or PI controllers, which are tuned manually on commission. Such tuning is quite often based on trial–and–error procedures and rarely involves any reference to mathematical models of the system to be controlled. Moreover, the controllers may well require manual re–tuning if the system dynamics change over time. Such conventional controllers can also suffer from various problems: for example, since they operate at fairly rapid sampling frequencies, they can experience stability problems if the controlled system is characterised by sizable pure time delays of the kind encountered in some glasshouse systems.

Unlike conventional DDC controllers, the PIP system utilises the alternative TDC design philosophy: namely, it is model–based, it actively exploits the digital mechanisation directly to achieve better control performance and it is not based on a digitised continuous–time design. This involves several innovations: first, the design is always based on a discrete–time mathematical model of the system, normally obtained by prior data–based modelling using recursive identification and estimation procedures; secondly, because it utilises state variable feedback, the PIP controller is inherently more powerful and flexible than the conventional PI and PID alternatives in achieving such objectives as closed loop pole assignment or optimal control; thirdly, the PIP controller has a structure which avoids common control system problems, such as integral wind–up; and finally, it is relatively robust to pure time (transport) delays because, in the z^{-1} operator version, it uses control sampling intervals that would be considered quite coarse by conventional standards.

Of course, these advantages are obtained at some cost; not in terms of implementational complexity, since the PIP system can often be very little more complicated than the conventional three term controller, but in relation to the design process itself. In particular, this involves the use of more sophisticated theoretical analysis and access to specially designed CACSD programs that assist in the application of this analysis to practical design problems. In subsequent Sections of this paper, we outline this theoretical background and describe briefly the True Digital Control (TDC) program developed at Lancaster.

2. TF MODELS IN THE z^{-1} AND δ OPERATORS

In all model-based control system design procedures, the form of the models and the associated theoretical background is of paramount importance. Most continuous-time procedures are based either on a Laplace transform transfer function model or state-space differential equations. It is not surprising, therefore, that TDC designs utilise the discrete-time equivalents of these models. It is possible to unify the model analysis in terms of a general operator (see e.g. Goodwin et al, 1988; Middleton and Goodwin, 1990). However, it is felt that the analysis is more transparent if we consider separately the two major forms of the transfer function model in the discrete-time domain: namely, the well known transfer function in the backward shift (z^{-1}, or equivalently the forward shift, z transform) operator; and the alternative, but less well known, delta (δ or discrete differential) operator transfer function.

2.1 Backward Shift (z^{-1}) TF Model

The general, discrete-time, z^{-1} operator transfer function (TF) representation of an nth order single input, single output (SISO), discrete-time system, with a sampling interval of Δt time units, is normally written in the following form,

$$y(k) = \frac{B(z^{-1})}{A(z^{-1})} u(k) \tag{1}$$

where $A(z^{-1})$ and $B(z^{-1})$ are the following polynomials in z^{-1},

$$A(z^{-1}) = 1 + a_1 z^{-1} + ... + a_n z^{-n}$$

$$B(z^{-1}) = b_1 z^{-1} + b_2 z^{-2} + ... + b_m z^{-m}$$

In general, no prior assumptions are made about the nature of the transfer function $B(z^{-1})/A(z^{-1})$, which may be marginally stable, unstable, or possess non–minimum phase characteristics.

However, if the input–output behaviour of the system is characterised by any pure time delay of τ sampling intervals[1], then this is accommodated by assuming that the first τ coefficients of the $B(z^{-1})$ polynomial, i.e.b_1, b_2, ... , b_τ , are all zero.

2.2 The Discrete Differential (δ) Operator TF Model

A interesting alternative to the z^{-1} operator TF model is the following "discrete differential operator" model, which was revived recently under the title delta (δ) operator by Goodwin and his co–workers (see Middleton and Goodwin, 1990),

$$y(k) = \frac{B(\delta)}{A(\delta)}\, u(k) \qquad (2)$$

where $A(\delta)$ and $B(\delta)$ are polynomials of the following form,

$$\begin{aligned} A(\delta) &= \delta^{p} + a_1\,\delta^{p-1} + \dots\dots + a_p \\ B(\delta) &= b_1\,\delta^{p-1} + \dots\dots + b_p \end{aligned}$$

with the index $p = \max(n,m)$ and the δ operator, for the sampling interval Δt, defined as follows in terms of the forward shift operator z,

$$\delta = \frac{z-1}{\Delta t} \;\; ; \text{ i.e. } \; \delta x(k) = \frac{x(k+1) - x(k)}{\Delta t}$$

Note that, for convenience, we have retained the nomenclature for the parameters a_i, b_j used in the z^{-1} operator case; but the parameter values will, of course, take on different values (see remark (3) below)

Remarks

1. As $\Delta t \rightarrow 0$, the delta operator reduces to the derivative operator

---

[1] Note that although, for convenience, the TF in (1) has an explicit unity time delay, the subsequent analysis can be developed also for a TF with zero time delay (i.e. $B(z^{-1}) = b_0 + b_1 z^{-1}$,..., $b_m z^{-m}$).

(s =d/dt) in continuous time (i.e $\delta \rightarrow s$).
2. Given a polynomial of any order n in the z operator, this will be exactly equivalent to some polynomial in d, also of order n.
3. As a consequence of 2., we can easily move between the z and δ operator domains. For example, a δ design can be implemented in practice by converting it back to the z^{-1} domain, so avoiding direct differencing (although this may not necessarily be the best form of implementation). Also, the δ operator model coefficients are related to forward z operator coefficients by simple vector matrix equations (see Chotai et al, 1990)
4. One attraction of the δ operator model to those designers who prefer to think in continuous-time terms is that it can be considered as a direct approximation to a continuous-time system. For example, it is easy to see that the unit circle in the complex z plane maps to a circle with centre $-1/\Delta t$ and radius $1/\Delta t$ in the complex δ plane; so that, as $\Delta t \rightarrow 0$, this circular stability region is transformed to the better known left half of the complex s plane. For very rapidly sampled systems, therefore, the δ operator model can be considered in almost continuous-time terms, with the pole positions in the δ plane close to those of the 'equivalent' continuous-time system in the s plane; and with the TF parameters directly yielding information on factors such as the approximate natural frequency and damping ratio.

3. MODEL IDENTIFICATION AND PARAMETER ESTIMATION

In control systems design, it makes good sense to use recursive methods model parameter estimation. Not only do these allow for the modelling of nonstationary systems, but they can also be implemented on-line for self-tuning and self-adaptive control applications. Of the many recursive methods that are now available for the estimation of z^{-1} operator transfer function model parameters, only one can also be applied directly to δ operator models. This is the Simplified Refined Instrumental Variable (SRIV) procedure (Young, 1984, 1985) which exploits special adaptive prefiltering, both to achieve good estimation performance and, in the δ operator case, to avoid numerical differentiation. It is a simplification of the full Refined (optimal) Instrumental Variable (RIV) algorithm first suggested by Young (1976), but it has the advantage of not requiring simultaneous modelling of the noise characteristics. As such, the SRIV algorithm provides a simple yet sophisticated vehicle for identification and estimation within the TDC context.

3.1 The SRIV Algorithm for $z^{-1}$and δ operator TF models

If we consider first the z^{-1} operator case, the adaptive prefiltering which characterises the SRIV algorithm can be justified qualitatively by considering the following stochastic form of equation (1),

$$y(k) = \frac{B(z^{-1})}{A(z^{-1})} u(k) + e(k) \qquad (3)$$

where e(k) is a zero mean, serially uncorrelated sequence of random variables with variance σ^2; and the TF is assumed to be stable, i.e. the roots of the characteristic equation A(z)=0 all lie within the unit circle of the complex z plane. This equation can be written in the following alternative vector form, which is linear-in-the-parameters $\{a_i , b_j\}$ of the TF model,

$$y(k) = z(k)^T a + \eta(k)$$

where,

$$z(k)^T = [-y(k-1), \ldots\ldots , -y(k-n) \;\; u(k-1), \ldots\ldots , u(k-m)]$$

$$a = [a_1 \;\; a_2, \ldots\ldots , a_n \;\; b_1, \ldots\ldots , b_m]^T$$

and $\eta(k)$ is a noise variable defined as follows in relation to the original white noise e(k),

$$\eta(k) = e(k) + a_1 e(k-1) + \ldots\ldots + a_n e(k-n)$$

Most estimation problems are posed in a manner such that the variable to be minimised has white noise properties. Thus, a sensible error function is the response or prediction error, $\hat{e}(k)$,

$$\hat{e}(k) = y(k) - \frac{\hat{B}(z^{-1})}{\hat{A}(z^{-1})} u(k)$$

where $\hat{B}(z^{-1})$ and $\hat{A}(z^{-1})$ are estimates of the TF polynomials $A(z^{-1})$ and $B(z^{-1})$. Unfortunately, this is nonlinear in the unknown parameters and so the estimation problem cannot be posed directly in simple, linear terms. However, the problem becomes linear-in-the-parameters if we assume prior knowledge of $A(z^{-1})$ in the form of an estimate $\hat{A}(z^{-1})$: then the error equation can be written in the form,

$$\hat{e}(k) = \frac{1}{\hat{A}(z^{-1})}\left[\hat{A}(z^{-1})\,y(k) - \hat{B}(z^{-1})\,u(k)\right]$$

which can be rewritten as,

$$\hat{e}(k) = \hat{A}(z^{-1})\,y^*(k) - \hat{B}(z^{-1})\,u^*(k)$$

where,

$$y^*(k) = \frac{1}{\hat{A}(z^{-1})}\,y(k) \quad ; \quad u^*(k) = \frac{1}{\hat{A}(z^{-1})}\,u(k) \qquad (4)$$

are "prefiltered" variables, obtained by passing y(k) and u(k) through the prefilter $1/\hat{A}(z^{-1})$.

With this reasoning in mind, the ordinary recursive Instrumental Varaiable (IV) algorithm (e.g. Young, 1984) can be applied iteratively to estimate the model parameter vector a, with the variables y(k), u(k) and the instrumental variable $\hat{x}(k)$ replaced, at each iteration, by their adaptively prefiltered equivalents $y^*(k)$, $u^*(k)$ and $\hat{x}^*(k)$, respectively, and with the prefilter parameters based on the parameter estimates obtained at the previous iteration (see Young, 1976, 1984, 1985). The main recursive part of this SRIV algorithm takes the form,

$$\hat{a}(k) = \hat{a}(k-1) + g(k)\,\{\, y^*(k) - z^*(k)^T\,\hat{a}(k-1) \,\} \qquad \text{(i)}$$

where,

$$g(k) = P(k-1)\hat{x}^*(k)\,[\, 1 + z^*(k)^T P(k-1)\hat{x}^*(k) \,]^{-1} \qquad \text{(ii)} \quad (5)$$

and

$$P(k) = P(k-1) + g(k)\, z^*(k)^T P(k-1) \qquad \text{(iii)}$$

where P(k) is related to the covariance matrix P*(k) of the estimated parameter vector $\hat{a}(k)$ by the equation,

$$P^*(k) = \sigma^2 P(k) \qquad (6)$$

and an estimate $\hat{\sigma}^2$ of the variance σ^2 can be obtained from an additional recursive equation based on the squared values of a suitably normalised recursive innovation sequence (see Young, 1984; p.100; Young et al, 1991b).

This complete recursive algorithm (5) can be considered simply as a modification of the well known Recursive Least Squares (RLS) algorithm, with the data vector z*(k) replaced alternately by $\hat{x}^*(k)$. At the jth iteration, the prefiltered instrumental variable $\hat{x}^*(k)$ required in the definition of $\hat{x}^*(k)$ is generated by adaptively prefiltering the output of an "auxiliary model" of the following form,

$$\hat{x}^*(k) = \frac{1}{\hat{A}_{j-1}(z^{-1})}\hat{x}(k) \quad ; \quad \hat{x}(k) = \frac{\hat{B}_{j-1}(z^{1})}{\hat{A}_{j-1}(z^{1})}u(k) \qquad (7)$$

adaptive prefilter adaptive auxiliary model

where $\hat{A}_{j-1}(z^{-1})$ and $\hat{B}_{j-1}(z^{-1})$ are estimates of the polynomials $A(z^{-1})$ and $B(z^{-1})$ obtained by reference to the parameter estimates obtained by the algorithm at the end of the previous (j−1)th iteration. The prefiltered input and output variables are obtained in a similar manner. Such a recursive–iterative or relaxation approach normally requires only two to three iterations to converge on sensible estimates of the parameters.

The δ operator version of the SRIV algorithm (5) is applied to the vector version of the δ operator TF model, i.e.,

$$\delta^p y(k) = z(k)^T a + \eta(k) \qquad (8)^2$$

where now,

$$z(k) = [\delta^{p-1}y(k), \delta^{p-2}y(k), \ldots, y(k), \delta^{p-1}u(k), \ldots , u(k)]^T$$

---

[2] Note that we choose here to select δ^p y(k) as the variable on the left hand side of this equation: this is rather arbitrary but, from experience, we have found that it appears to have some advantages. An alternative would be to select y(k).

with a and $\eta(k)$ defined accordingly. The resulting algorithm is algebraically identical to the z^{-1} version: it is simply necessary to define the data vectors appropriately, i.e.,

$$z^*(k)=[\delta^{p-1}y^*(k)\ ,\ \delta^{p-2}y^*(k)\ ,\ ..,\ y^*(k)\ ,\ \delta^{p-1}u^*(k)\ ,\ ..,\ u^*(k)]^T$$

$$\hat{x}^*(k)=[\delta^{p-1}\hat{x}^*(k),\ \delta^{p-2}\hat{x}^*(k),\ ..\ ,\ \hat{x}^*(k),\ \delta^{p-1}u^*(k),\ ..\ ,\ u^*(k)]^T$$

with $y^*(k)$ in the innovation term $\{y^*(k) - z^*(k)^T\hat{a}(k-1)\}$ replaced by $\delta^p y^*(k)$. Here, the star superscript again indicates that the variables have been adaptively prefiltered, this time by the δ operator prefilter.

This prefilter now performs two important functions: first it helps to improve the statistical efficiency and yield lower variance estimates; second, the elements of $z^*(k)$ and $\hat{x}^*(k)$ can be obtained directly from the adaptive prefilter, thus avoiding direct multiple differencing of the input and output signals, with its attendant problems of noise amplification. Used in this manner, the prefilters are seen as direct descendants of the "state variable filters" proposed by the first author in the early nineteen sixties for the estimation of continuous-time system models (Young, 1964, 1965, 1966, 1969a, 1969b, 1970)).

3.2 Model Order Identification

Model order identification is extremely important in TF modelling for control system design since it avoids possible over-parameterisation and uncontrollable modes. A successful identification procedure based on IV estimation (see Young, 1989) is to choose the model which minimises the following identification statistic,

$$YIC = \log_e\left\{\frac{\sigma^2}{\sigma_y^2}\right\} + \log_e\{NEVN\} \qquad (9)$$

where,

σ^2 is the sample variance of the model residuals e(k)
σ_y^2 is the sample variance of the measured system output y(k) about its mean value.

while NEVN is the "Normalised Error Variance Norm" (Young et al, 1980) defined as,

$$NEVN = \frac{1}{np} \sum_{i=1}^{i=np} \frac{\hat{\sigma}^2 p_{ii}}{\hat{a}_i^2}$$

Here, in relation to the TF models (1) and (2), np is the total number of parameters estimated, i.e. n+m for model (1) and 2p for model (2); $\hat{a}_i^2$ is the estimate of the ith parameter in the parameter vector a; while p_{ii} is the ith diagonal element of the P(N) matrix, where N is the sample size (so that $\hat{\sigma}^2 p_{ii}$ is an estimate of the error variance associated with the ith parameter estimate after N samples).

The first term in (9) provides a normalised measure of how well the model explains the data: the smaller the variance of the model residuals in relation to the variance of the measured output, the more negative the first term becomes. Similarly, the second term is a normalised measure of how well the parameter estimates are defined for the npth order model: clearly the smaller the relative error variance, the better defined are the parameter estimates in statistical terms, and this is once more reflected in a more negative value for the term. Thus the model which minimises the YIC provides a good compromise between model fit and parametric efficiency: as the model order is increased, so the first term tends always to decrease; while the second term tends to decrease at first and then to increase quite markedly when the model becomes over-parameterised and the standard error on its parameter estimates becomes large in relation to the estimated values (in this connection, note that the square root of $\hat{\sigma}^2\ p_{ii}\ /\ \hat{a}_i^2$ is simply the relative standard error on the ith parameter estimate).

One caveat is necessary: like all currently available model identification criteria, the YIC should be treated only as an objective aid to order identification and should always be considered in association with other criteria in final model order selection. For example, it is possible for the YIC to favour low order models under certain circumstances when a priori information on the system would strongly suggest that a somewhat higher order model is preferable from a control standpoint. Such practical considerations, based on prior information that is not being utilised by the data-based YIC statistic, may well justify selection of a model with a low rather than minimum YIC.

4. PIP CONTROL DESIGN FOR THE z^{-1} OPERATOR MODEL

The special non–minimal state–space (NMSS) representation which we associate with the z^{-1} transfer function model (1) is defined by the following state vector,

$$x^T = [y(k)\ y(k-1)\ ...\ y(k-n+1)\ u(k-1)\ ...\ u(k-m+1)\ z(k)]$$

where z(k) is an "integral of error" term at sampling instant k,

$$z(k) = z(k-1) + \{y_d(k) - y(k)\}$$

in which $y_d(k)$ is the reference or command input to the servomechanism system. This integral–of–error term is introduced to ensure type 1 servomechanism performance in the resultant PIP control system (Young and Willems, 1972). The NMSS representation is then obtained directly in the following form,

$$x(k) = F\,x(k\text{-}1) + g\,u(k\text{-}1) + d\,y_d(k) \qquad (10)$$

where,

$$F = \begin{bmatrix}
-a_1 & -a_2 & .... & -a_{n-1} & -a_n & b_2 & b_3 & ..... & b_{m-1} & b_m & 0 \\
1 & 0 & .... & 0 & 0 & 0 & 0 & ..... & 0 & 0 & 0 \\
0 & 1 & .... & 0 & 0 & 0 & 0 & ..... & 0 & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\
0 & 0 & .... & 1 & 0 & 0 & 0 & ..... & 0 & 0 & 0 \\
0 & 0 & .... & 0 & 0 & 0 & 0 & ..... & 0 & 0 & 0 \\
0 & 0 & .... & 0 & 0 & 1 & 0 & ..... & 0 & 0 & 0 \\
0 & 0 & .... & 0 & 0 & 0 & 1 & ..... & 0 & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\
0 & 0 & .... & 0 & 0 & 0 & 0 & ..... & 1 & 0 & 0 \\
a_1 & a_2 & .... & a_{n-1} & a_n & -b_2 & -b_3 & ..... & -b_{m-1} & -b_m & 1
\end{bmatrix}$$

$$g = [\ b_1\ \ 0\ \ .....\ \ 0\ \ 1\ \ 0\ \ 0\ \ .....\ \ 0\ \ 0\ \ -b_1\ ]^T$$

$$d = [\ 0\ \ 0\ \ .....\ \ 0\ \ 0\ \ 0\ \ 0\ \ .....\ \ 0\ \ 0\ \ 1\ ]^T$$

This is a rather natural, albeit unconventional, state-space form for TF models. Also, the definition of this particular NMSS state vector is significant in control terms: it means that the State Variable Feedback (SVF) control law involves only the directly measurable input and output signals, together with their past values, all of which can be stored in the digital computer. As a result, any SVF control system design (e.g. pole assignment or optimal) does not need to resort to the complication of a state reconstructor (i.e. observer), since the effective output matrix in control terms is the I_{m+n} identity matrix. The conditions for controllability of the NMSS model are given by the following theorem.

Theorem (Wang and Young, 1988). Given a single input-single output system described by (1), the non-minimal state space representation (10), as described by the pair {F,g}, is completely controllable if, and only if, the following two conditions are satisfied:

(i) the polynomials $A(z^{-1})$ and $B(z^{-1})$ are coprime

(ii) $b_1 + b_2 + \ldots\ldots + b_m \neq 0$

The coprimeness condition is equivalent to the normal requirement that the transfer function model (1) should have no pole-zero cancellations. The second condition avoids the presence of a zero at unity which would cancel with the unity pole associated with the integral action; it does not apply if the z(k) state variable is removed (i.e. if the control system does not employ integral action).

4.1 The Control Algorithm

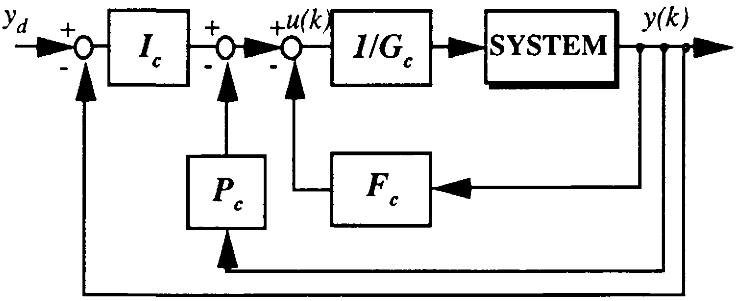
In the context of the NMSS system representation, the automatic control objective is to design an SVF control law with gain vector $k = [\, f_0 \, , f_1 \, , \ldots, f_{n-1} \, , g_1 \, , \ldots \, , g_{m-1} \, , k_I \,]$, i.e.,

$$u(k) = -k^T x(k)$$

or, in expanded form,

$$u(k) = -f_0 y(k) - f_1 y(k-1) - \ldots - f_{n-1} y(k-n+1) - g_1 u(k-1) - \ldots - g_{m-1} u(k-m+1) - k_I z(k)$$

such that either the closed loop poles are at preassigned positions in the complex z-plane;



P_c - Proportional control		(gain f_0)
I_c - Integral control		(gain k_I)
F_c - Feedback filter		$f_1 z^{-1} + f_2 z^{-2} + \ldots + f_{n-1} z^{-n+1}$
G_c - Input filter		$1 \;+ g_1 z^{-1} + \ldots + g_{m-1} z^{-m+1}$

Fig.1 The Proportional-Integral-Plus (PIP) control system

or the system is optimised in an Linear-Quadratic (LQ) sense (or Linear-Quadratic-Gaussian (LQG) in the stochastic situation).

In the pole assignment case, the closed-loop system block diagram takes the Proportional-Integral-Plus (PIP) form shown in Fig.1. The closed-loop TF associated with this block diagram can be written,

$$y(k) = \frac{D(z^{-1})}{CL(z^{-1})}\, y_d(k)$$

where,

$$D(z^{-1}) = k_I B(z^{-1})$$

and,

$$CL(z^{-1}) = (1-z^{-1})[G_c(z^{-1})A(z^{-1}) + F_c(z^{-1})B(z^{-1}) + f_0 B(z^{-1})] + k_I B(z^{-1})$$

The closed-loop characteristic polynomial $CL(z^{-1})$ can now be expanded and the coefficients for like powers of z^{-i} equated to those the desired closed loop characteristic polynomial,

$$d(z^{-1}) = 1 + d_1 z^{-1} + d_2 z^{-2} + \ldots\ldots + d_{m+n} z^{-(n+m)}$$

which is chosen to ensure the assignment of the closed loop poles to designer–specified positions in the complex z plane. This results in the following set of linear, algebraic equations,

$$\sum(n,m) \,.\, k = S_c - f \qquad (11)$$

where k is the SVF control gain vector, i.e.,

$$k = [\, f_0 \,,\, f_1 \,,\, \ldots,\, f_{n-1} \,,\, g_1 \,,\, \ldots \,,\, g_{m-1} \,,\, k_I \,]$$

while S_c and f are the following vectors,

$$S_c = [d_1 \,,\, d_2 \,,\, \ldots\ldots \,,\, d_{n+m-1} \,,\, d_{n+m} \,]^T$$

$$f = [a_1 - 1 \,,\, a_2 - a_1,\, \ldots\ldots \,,\, a_n - a_{n-1},\, -a_n \,,\, 0 \,,\, 0 \,,\, \ldots\ldots \,,\, 0]^T$$

and $\sum(n,m)$ is the (n+m)x(n+m) matrix shown in Fig.2 (see Young et al.; 1987b). Provided the controllability conditions of the above theorem are satisfied, this set of linear, simultaneous equations can be solved to yield the unique set of SVF control gains which define the vector k.

$$\begin{bmatrix}
b_1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & b_1 \\
b_2-b_1 & b_1 & \cdots & \vdots & a_1-1 & 1 & \cdots & \vdots & b_2 \\
\vdots & b_2-b_1 & \cdots & \vdots & \vdots & a_1-1 & \cdots & 0 & \vdots \\
\vdots & \vdots & & 0 & \vdots & \vdots & & 1 & \vdots \\
b_m-b_{m-1} & \vdots & \cdots & b_1 & \vdots & \vdots & \cdots & a_1-1 & b_m \\
-b_m & b_m-b_{m-1} & \cdots & b_2-b_1 & a_n-a_{n-1} & \vdots & \cdots & \vdots & 0 \\
0 & -b_m & \cdots & \vdots & -a_n & a_n-a_{n-1} & \cdots & \vdots & \vdots \\
\vdots & 0 & \cdots & \vdots & 0 & -a_n & \cdots & \vdots & \vdots \\
\vdots & \vdots & & \vdots & \vdots & 0 & & \vdots & \vdots \\
\vdots & \vdots & & b_m-b_{m-1} & \vdots & \vdots & & a_n-a_{n-1} & \vdots \\
0 & 0 & \cdots & -b_m & 0 & 0 & \cdots & -a_n & 0
\end{bmatrix}$$

Fig.2 The $\sum(n,m)$ matrix

In the LQ or LQG situation, the SVF gains are computed to minimise the following quadratic cost function,

$$J = \sum_{i=1}^{i=\infty} x(k)^T Q\, x(k) + q_u u(k)^2 \qquad (12)$$

where Q is an (n+m)x(n+m) diagonal matrix, with elements defined as follows,

$$Q = \text{diag.}[q_1\ q_2\ \ldots\ q_n\ q_{n+1}\ q_{n+2}\ \ldots\ q_{n+m-1}\ q_{n+m}]$$

with,

$$q_1 = q_2 = \ldots = q_n = q_y;\ q_{n+1} = q_{n+2} = \ldots = q_{n+m-1} = q_u;\ q_{n+m} = q_z$$

Here q_y, q_u and q_z are the partial weightings on the output, input and integral-of-error variables in the NMSS vector x.. These partial weightings are defined as,

$$q_y = \frac{W_y}{n}\ ;\ q_u = \frac{W_u}{m}\ \text{and}\ q_z = W_z \qquad (13)$$

so that the total weightings on the output y(k), input u(k) and integral-of-error z(k) are W_y, W_u, and W_z. These three weighting variables[3] are then chosen in the usual manner to achieve the desired closed loop performance. The optimum SVF gains for the selected weighting values are obtained by computing the steady state solution of the associated, discrete-time, matrix Riccati equation, given the NMSS system description (F, g) and weighting matrices (Q, R). Note that, in this case, R = q_u because of the special, NMSS formulation.

---

[3] Note that this simplification to yield only three user-specified weighting terms is for both convenience and simplicity of conrol system design. However, it is possible to apply different weightings to all the state variable terms in the cost function, so allowing for differential weighting into the past.

5. PIP CONTROL DESIGN FOR THE δ OPERATOR MODEL

Following a similar design philosophy to that used for the z^{-1} operator model, the δ operator TF model can be represented by the following NMSS equations,

(14)

where,

$$g = [\, b_1 \;\; 0 \;\; \ldots\ldots \;\; 0 \;\; 1 \;\; 0 \;\; 0 \;\; \ldots\ldots \;\; 0 \;\; 0 \;\; 0 \,]^T$$

$$d = [\, 0 \;\; 0 \;\; \ldots\ldots \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; \ldots\ldots \;\; 0 \;\; 0 \;\; 1 \,]^T$$

and,

$$F = \begin{bmatrix} -a_1 & -a_2 & \ldots & -a_{P-1} & -a_P & b_2 & b_3 & \ldots & b_{P-1} & b_P & 0 \\ 1 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 1 & 0 & 0 \\ 0 & 0 & \ldots & 0 & -1 & 0 & 0 & \ldots & 0 & 0 & 0 \end{bmatrix}$$

In this formulation, however, the control variable is denoted by v(k), which is defined as follows in terms of the control input u(k),

$$v(k) = \delta^{P-1}\, u(k) \qquad (15)$$

with the associated state vector x(k) defined as,

$$x(k) = [\delta^{p-1}y(k), \delta^{p-2}y(k), .., \delta y(k), y(k), \delta^{p-2}u(k), \delta u(k), u(k), z(k)]^T$$

In these equations, z(k) is, once again, the "integral of error" state, which is now defined in terms of the the inverse delta operator, or digital integrator δ^{-1}, i.e.,

$$z(k) = \delta^{-1} \{y_d(k) - y(k)\}$$

and, as before, $y_d(k)$ is the reference or command input at the kth sampling instant.

5.1 The Control Algorithm

As in the z^{-1} operator case, the SVF control law is defined in terms of the state variables. In this δ operator situation, however, these are the output and input and their discrete differentials up to the appropriate order, as well as the integral of error state z(k). This control law can be written in the form,

$$v(k) = - k^T x(k) \qquad (16)$$

where now,

$$k^T = [\ f_{p-1}\ \ f_{p-2}\ \ ...\ f_0\ \ g_{p-2}\ g_{p-3}\ \ ...\ g_0\ k_I\ ]$$

is the SVF control gain vector for the δ operator model NMSS form.

The block diagram of this δ operator PIP control system is similar to the z^{-1} operator system in Fig.1, with appropriate changes in the form of the control filters F_c and G_c. In order to avoid the numerical differentiation inherent in these δ operator filters and the consequent danger of noise amplification, however, the controller will normally be converted into an alternative, practically realisable, form prior to implementation. This can be achieved by either (a) converting it back to the z^{-1} domain; (b) manipulating it into the form of realisable ratios of δ operator polynomials; or (c) invoking the separation theorem and introducing an NMSS observer (i.e. a state reconstruction filter or Kalman filter in the stochastic case), with the SVF control law implemented directly in δ operator terms and the state estimates replacing the discrete differentials.

6. THE TDC COMPUTER AIDED CONTROL SYSTEM DESIGN (CACSD) PACKAGE

The efficacy of any modern control systems design procedure depends to a large extent on the availability of suitable computer software which can assist the control systems analyst in the various stages of the design process. Our major objectives in developing such a Computer Aided Control System Design (CACSD) package (see Tych et al, 1991a,b) for PIP control system design have been to produce a user-friendly, but powerful, set of integrated design tools, which are as portable as possible and will also prove popular to the widest possible cross-section of the control and systems community.

Of course, it is not possible to guarantee that such objectives will be met completely but, by integrating the whole package, which we have called TDC, within the popular Matlab/Simulab™ program (Moler et al, 1987), we believe we have designed a system which will go some way to satisfying the current tastes of many control systems analysts. Moreover, in shielding the user from the Matlab command line by means of a well designed but simple Graphical User Interface (GUI), while still allowing easy access to Matlab if this is desired, we believe we have provided a good combination of a well designed, systematic procedure for day-to-day design studies, and a flexible tool for more novel research exercises.

6.1 The Graphical User Interface (GUI)

The GUI of the package takes full advantage of the Matlab graphical interactive capabilities. It is organized as a two level interface, with the main menu calling lower level menus. The fact that the user interface is almost entirely written in Matlab's macro language, constrains the flexibility and "look" of the GUI to some extent but gives portability to most of the computing platforms where Matlab/Simulab has been implemented. As shown in Fig.3, the GUI shields the user not acquainted with Matlab from its command line, whilst giving easy access to Matlab's commands, Simulab, other toolboxes, and the workspace, if this is required.

6.2 Fortran / C Routines

In the TDC package some numerical computation (e.g. the recursive identification and estimation; and Monte Carlo closed loop sensitivity analysis) is performed by external, compiled programs written in either Fortran or C. These programs then communicate with Matlab by means of disk files, so speeding up the calculations, often to a considerable degree. In some cases, as in the SUN implementation, where a user interface more flexible than that provided by Matlab is required, the programs also take advantage of the more versatile interactive facilities. Clearly, these latter components of the TDC system are not portable to other machines, where alternative arrangements prove necessary. However, such special facilities have been kept to a minimum.

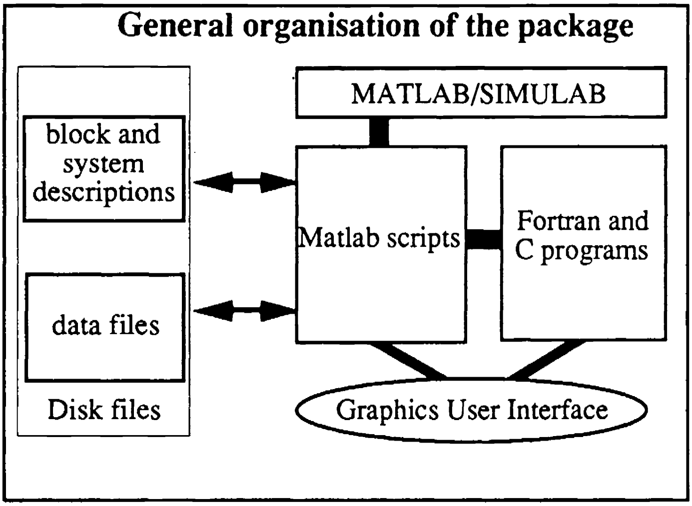


Fig. 3 General organisation of the TDC pachage

6.3 Internal Organisation of the Package

The TDC package is designed as a quasi-object-oriented, high level specialized programming language using the Matlab environment. The commands are issued using either a system of menus and input panels, as shown in Fig.4, or via the Matlab command line. The two main classes of objects defined in the system are data and systems. Data are defined as one or more series of samples, while a system consists of blocks (with parameters) and structure (connections between the blocks, inputs and outputs).

6.4 Control SYSTEM DESIGN AND EVALUATION Tools

The control system design is implemented according to the TDC approach based on PIP control system designs. A high level macro control leads the user through the design procedure, which includes the following options,

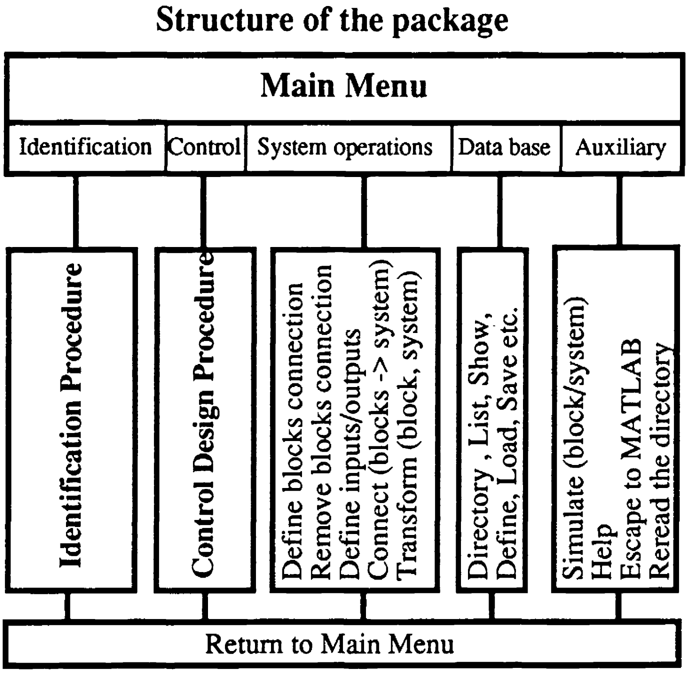


Fig. 4 Structure to the TDC package

- choice between pole assignment and LQ–optimal control;
- optional cancellation of numerator effects for minimum phase systems;
- graphical specification of the closed loop pole locations or entry of LQ cost function weights;
- standard evaluation (time and frequency responses, pole locations) of closed loop behaviour;
- Monte Carlo (MC) closed loop sensitivity analysis based on the uncertainty associated with the open loop model parameter estimates.
- Evaluation of design using Simulab.

The associated interactive control system design panel, in the delta operator mode, is shown in Fig.5. Buttons are highlighted to indicate the current status (i.e. the selections made by the user); and the graphical output appears in the adjacent graphical window.

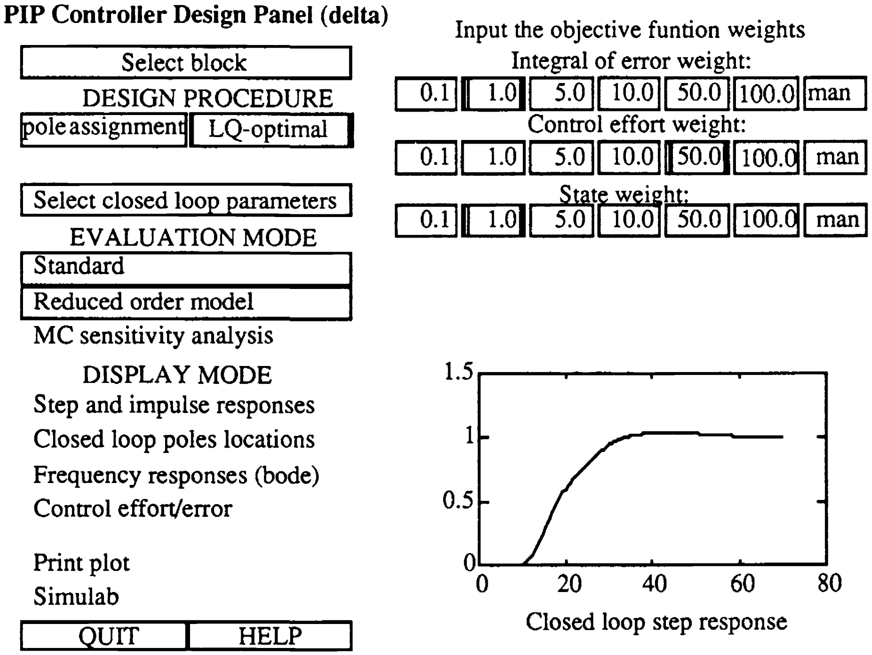


Fig.5 The PIP control design panel in the delta operator mode, after specifying the optimal ctiterion weights {1,50,1} for LQ optimal design: the closed loop step response obtained with these weights is automatically generated in the panel, as shown.

7. PRACTICAL APPLICATIONS IN GLASSHOUSE HORTICULTURE

In this section, we consider three practical examples which epitomise the potential of TDC control system design in glasshouse horticulture. The first is concerned with the relatively simple problem of glasshouse temperature control. Here, the object is to show how a simple PIP controller of similar form and complexity to a conventional sampled data PI controller can provide significantly improved temperature regulation.

In the second example, we consider the self-adaptive control of a non-linear heater-fan system. Although not strictly a glasshouse control problem, this example provides a reasonable, experimentally convenient, test-bed for greenhouse temperature control system designs. In particular, the non-linear heater-fan system can be considered analogous to a small scale heating-ventilation system, with behavioral similarities to greenhouse environmental control systems but operating on a much faster time-scale. This similarity means that the self-adaptive PIP control approach utilised in the example could form the basis for self-adaptive greenhouse temperature control system design, particularly where only "bang-bang" type of control operation is possible.

The third example is more complex and shows how higher order PIP controllers can be applied successfully to the control of the re-circulatory flow system of a Nutrient Film Technique (NFT) system used in glasshouse horticulture.

7.1 Example 1: Temperature Control in a Glasshouse

Many researchers (Saffell and Marshall, 1983; Hurd and Graves, 1984; Bailey, 1985; Miller et al., 1985) have suggested control methods which can save greenhouse heating energy by reducing the temperature setpoint when heat losses are high (windy, no thermal screen), and increasing it when the losses are lower. In order to obtain appreciable energy conservation with such methods, the accuracy of temperature control must be good, so that the setpoint can be reduced without any risk that the temperature will fall below an allowed limit anywhere within the crop. The PIP design is perfectly suited to these requirements: indeed, its inherent use of state variable feedback means that the controller can satisfy the kind of optimal Linear-Quadratic-Gaussian (LQG) criteria that can be related directly to factors such as energy utilisation and conservation. Also, it is straightforward to introduce self-tuning or self-adaption into the PIP design, so ensuring that it retains this optimal operation despite any changes that may occur to the controlled system over the passage of time.

Fig. 6 shows the changes of air and heating pipe temperature in a single glass greenhouse at the AFRC Silsoe Research Institute in the U.K. These changes, which were induced by pseudo-random binary step adjustments of the control valve under open loop conditions, indicate predominantly first order response between the valve and the air temperature. This observation is confirmed by the modelling analysis for this system reported by Davis and Hooper (1991). For a sampling interval of 10 minutes and using the full RIV algorithm, they found that the overall discrete-time model between valve aperture and internal air temperature is of the form,

$$T(k) = \frac{1.22z^{-1}}{1 - 0.922z^{-1}} u(k) + \xi(k) \qquad (17)$$

where T(k) is the internal air temperature in °C, u(k) is the control input (the fractional valve aperture of the heating system) and $\zeta(k)$ is a term to account for the noise in the system. From equation (17), we see that the natural (open loop) system is indeed first order, with a time constant of 123 minutes, steady state gain of 15.6, and pure time delay of 10 minutes. The sampling interval of 10 minutes was selected following model identification studies carried out both at Silsoe (Davis, 1986, Davis and Hooper, 1991) and Lancaster. In contrast to the situation at shorter sampling intervals, this sampling rate yields data-based z^{-1} operator models with well-defined parameter estimates and allows, most conveniently, for the dead-time characteristics of the process.

For the system in equation (17), the PIP design conveniently reduces to a PI design. In contrast to the conventional PI controller used previously in the greenhouse, however, the PIP design has numerous advantages: its PI structure exploits the power of SVF; it is based directly on the discrete-time model of the heating system and not on the digitisation of a continuous-time design; and the vagueries of manual tuning are replaced by objective, model-based synthesis.

The NMSS model of (17) in noise free case is given by

$$\begin{bmatrix} T(k) \\ z(k) \end{bmatrix} = \begin{bmatrix} 0.926 & 0 \\ 0.926 & 1 \end{bmatrix} \begin{bmatrix} T(k-1) \\ z(k-1) \end{bmatrix} + \begin{bmatrix} 1.22 \\ -1.22 \end{bmatrix} u(k-1) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} y_d(k)$$

and the SVF control law takes the form,

$$u(k) = -f_0 T(k) - k_I z(k)$$

where is f_0 the proportional gain and k_I is the integral gain. As a result, the closed loop TF in the noise free case is given by,

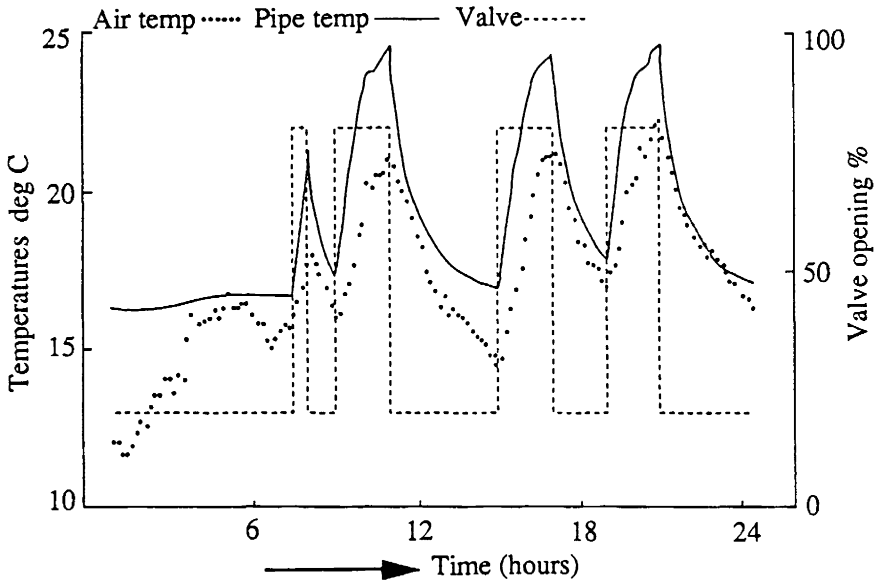


Fig. 6 changes of air and heating pipe temperature in a single glass greenhouse at the AFRC Research Institute in the U.K.

$$T(k)=\frac{1.22k_I z^{-1}}{1+(1.22f_0+1.22k_I-1.926)z^{-1}+(0.926-1.22f_0)z^{-2}}y_d(k)$$

where we see that the closed loop characteristic polynomial can be completely defined by appropriate selection of the control gains f_0 and k_I, as predicted by the NMSS control theory.

If the desired closed-loop polynomial is now defined as[4],

$$d(z^{-1}) = 1 - 0.5z^{-1}$$

then the system closed loop response will be dominated by a first order mode with unity steady state gain to input commands $y_d(k)$,

---

[4] Note that here, for simplicity, we choose to specify first order dynamics for the closed loop; in general, it appears that more robust control is achieved if the order of $d(z^{-1})$ is chosen equal to the natural closed loop order of the NMSS system (in the case 2).

and a time constant of 14.4 minutes, which can be contrasted with the open loop time constant of 123 mins. The resultant control gains are computed straightforwardly as $f_0 = 0.759$ and $k_I = 0.41$.

Unlike the PIP system, the conventional controller used in the glasshouse is a digitised, continuous-time, PI control system operating on the much smaller sampling interval of one minute. It is typical of commercially available controllers (although some controllers make adjustments even more frequently; see Sanger, 1985) and, while it was probably not tuned optimally, the control parameters were set at what were considered to be reasonable values.

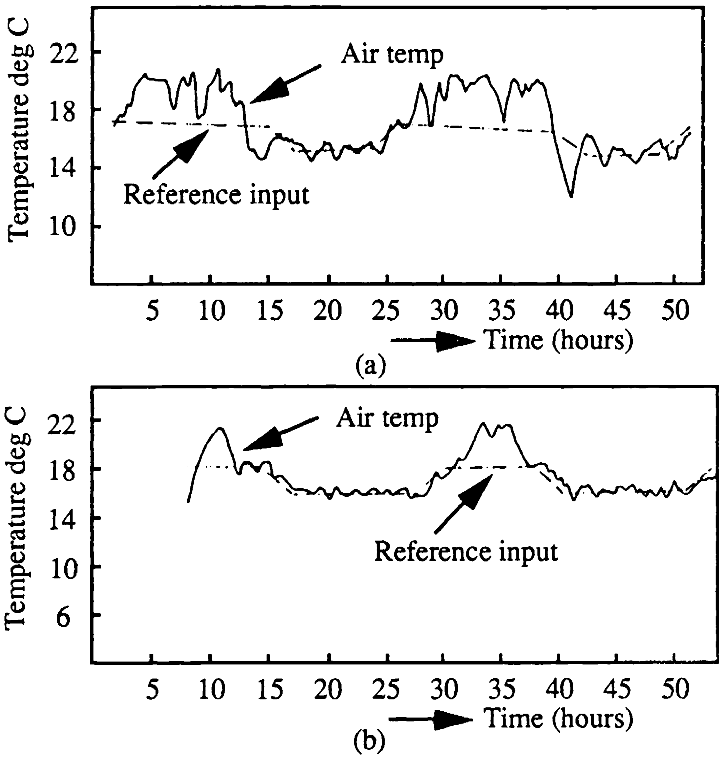


Fig. 7 Comparison of (a) PIP; and (b) conventional PI glasshouse temperature controllers.

The performance of the two controllers is compared in Fig. 7, which shows them responding to similar, but not identical, set point changes over a period of some 55 hours. The failure of both systems to maintain the set points during parts of the day-time is due to the nature of the system: only heating action is available so, during the day when natural temperatures rise because of solar irradiance, the controllers apply their minimum, zero input most of the time. However, at other times, and particularly during the cold night hours, the controllers are fully operative. It is here that we see the superiority of the PIP control system, which quickly compensates for the disturbances and maintains the desired temperature level throughout the night.

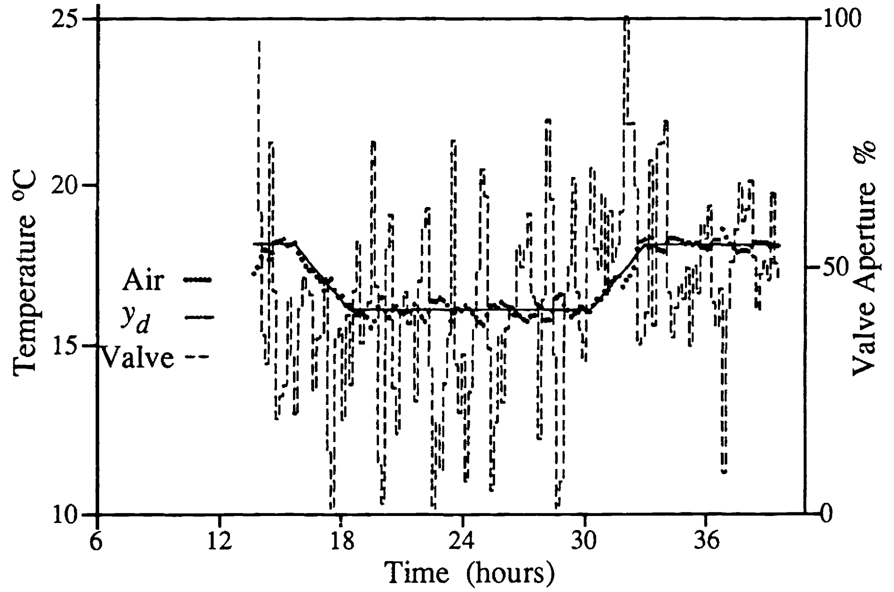


Fig. 8 PIP control of glasshouse temperature using feedback of both air and pipe temperature (from Davis and Hooper, 1991)

In contrast, the conventional PI system, despite its more rapid sampling frequency, is quite sluggish and oscillatory in its response to both set point changes and disturbances. This apparently paradoxical behaviour is quite easy to understand. First, the rapid sampling rate makes the system more sensitive to the time-delays in the system and it is well known that conventional, continuous-time PI controller leads to oscillatory behaviour when significant time delays are present. Second, the PIP pole assignment design in this case is intended to provide rapid and critically damped response and achieves this because of its model-based design and powerful SVF control law.

For some insulated greenhouses, it has been found (Davis and Hooper, 1991) that still better and more robust PIP control is obtained if heating pipe temperatures are included in the PIP feedback design. It is straightforward to extend the PIP design procedure to accomodate additional measured variables of this type: the overall system is then of higher order and control law simply incorporates present and past values of the heating pipe temperature, in addition to the air temperature measurements. Fig.8 shows the response of such a system and more details are given in Davis and Hooper (1991).

7.2 Example 2: Self-adaptive PIP Control of a Nonlinear Heated Bar System

The PIP approach described above can also form the basis for the design of controllers for certain nonlinear systems characterised by piecewise linear or bilinear behaviour. This is demonstrated by recent research at Lancaster on the adaptive temperature control of a nonlinear heated bar system, in which the bar is heated by an electrical element and cooled by a small fan. In this self-adaptive control (SAC) system, an on-line PIP pole assignment control algorithm is utilised to implement pulse-width modulation (PWM) of the "bang-bang" control signals to either the electric heater or cooling fan, which provide the mechanism for controlling the bar temperature. Adaption is based on a novel recursive time variable parameter (TVP) estimation algorithm capable of tracking the extremely rapid variation of model parameters that occur when the control signal changes sign (see, Chotai et al, 1991a and Behzadi, 1989).

This rather novel SAC system is based on a earlier adaptive system for an airborne vehicle (Young, 1981). It is a computationally efficient solution, in the sense that a single recursive estimator is able to track the changes in the parameters of what are, in effect, two, quite different dynamic systems: namely, the bar temperature when controlled rather sluggishly by the heater, and the same temperature controlled much more rapidly by the fan.

Clearly a SAC system based on two separate recursive estimation algorithms would provide an alternative but computationally somewhat slower and more complex solution.

A block diagram of the overall SAC heater-fan system is shown in Fig.9. Certain details of the design that are specific to the heated bar example should first be noted. Most importantly, the first order models for the heater and fan sub-systems are different in structural terms: they are defined at different sampling rates; and they have different order numerators and time delay characteristics. In particular, the heater sub-system model is defined for a sampling interval of 10 secs., with a pure time delay of two sampling intervals (20 secs), and two numerator coefficients b_2 and b_3; while, for the fan sub-system model, the sampling interval is 3 secs., the time delay is only one sampling interval (3 secs.), and there is only one numerator polynomial coefficient β_1 .

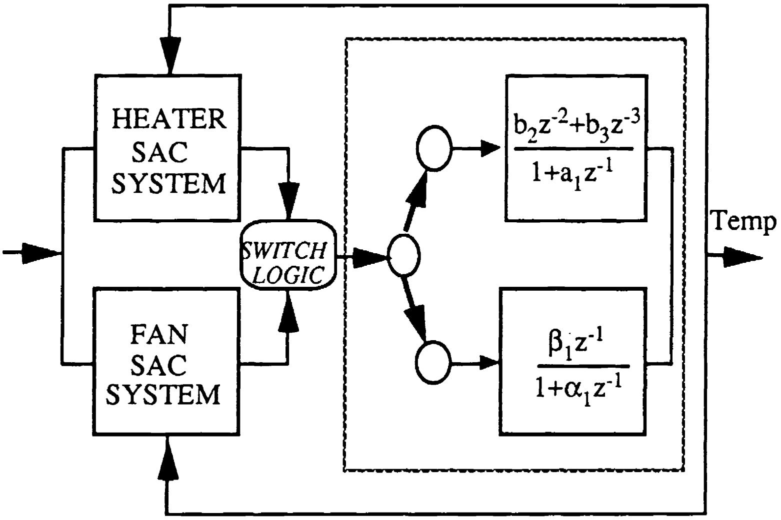


Fig. 9 Block diagram of the Heater/Fan self-adaptive control (SAC) system

Since the time constants of both system models remain reasonably constant, they have been assumed stationary for the purposes of the present design (although this is not essential and could be assumed otherwise in a more general situation; see Behzadi, 1989, Young et al., 1988). In other words, the denominator polynomial coefficients a_1 and α_1 in each of the sub-system models are frozen at their initial, off-line, estimated values and only the rapidly time variable numerator polynomial terms (i.e. the b_2 and b_3 parameters for the heater system and β_1 for the fan system) are recursively updated in real-time.

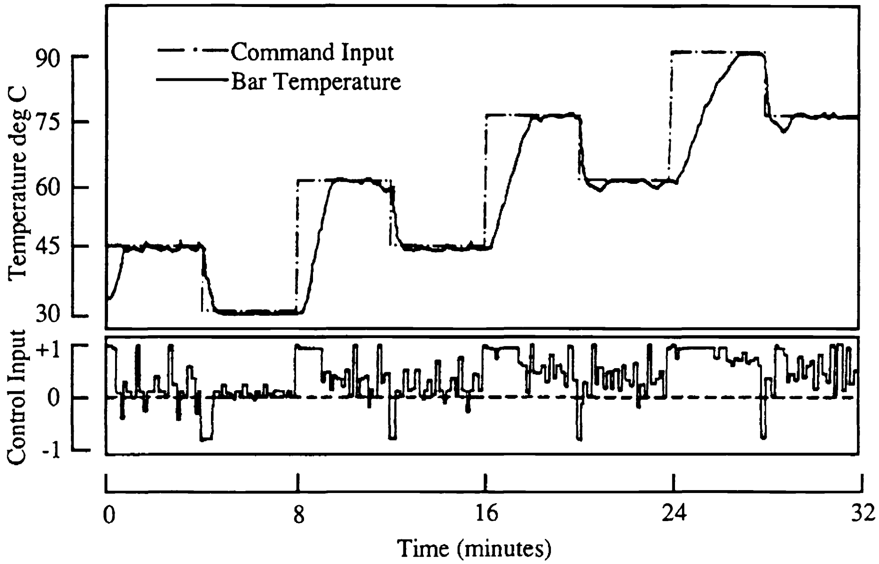


Fig. 10 Self-adaptive PIP control of a nonlinear heated bar system

Fig.10 shows a typical run using the SAC system based on the single recursive estimator: here, the system is responding to a "staircase" type input, chosen to move the bar temperature over a fairly wide range. In this manner, the adaptive action is required not only to allow for the rapidly changing dynamics associated with reversals in the control action, but also to compensate for the longer term, slower changes in the system dynamics which arise either because of the increased heat losses at higher bar temperatures, or the increased cooling effect of the fan at lower temperatures.

These results do not illustrate the performance of the control system in the face load disturbances or input signal saturation. In both cases, however, the system performs well: the inherent integral action ensures that the effects of constant load disturbances are rejected; and the PIP control law (Young et al., 1987b, Behzadi, 1989) avoids integral wind-up when large command inputs lead to input signal saturation. Further details of this heated bar control system are given in Behzadi (1989) and Chotai et al (1991a), including those obtained with the more complex SAC implementation involving two separate recursive algorithms, one for each sub-system. However, this involved greater complexity and yielded no improvement in system performance.

7.3 Example 3 : Control of an Nutrient Film Technique System

The Nutrient Film Technique (NFT) is a water culture system, normally used in protected environments like glasshouses, where a shallow stream of nutrient solution is recirculated over the bare roots of the growing plants in order to provide adequate water, nutrient and aeration. In the NFT system considered here, the plants are sown in small blocks of a substrate called "rockwool" and are placed in gently sloping channels. The nutrient solution is collected in a tank, or settling trench, and then recirculated by an electrically driven pump back into the channels, so forming a positive feedback loop. An extensive exposition of the NFT method is given in Graves (1983).

In the NFT, temperature, ion concentration and acidity of the solution are monitored and controlled in some manner. The NFT system requires the addition of up to twelve nutrients, although most of them are needed in only small quantities. On the other hand, nutrients like potassium and nitrogen are taken up by the plants in comparatively large quantities. Such NFT culture is particularly successful for tomato plants (Cooper, 1967), which can tolerate quite a range of nutrient concentrations but whose fruit quality can suffer at low concentrations (Winsor and Massey 1978).

The NFT system with its positive feedback mechanism can, of course, be modelled with various levels of complexity. An important first stage in modelling is often the development of an adequate computer simulation model. In the present study a relatively high order, simulation model of NFT system was constructed using a combination of experimental investigation (based on tracer experiments to model the flow dynamics) and normal, physically-based, mathematical modelling (Young et al 1987a; Behzadi 1989). Most practical evaluation of the PIP designs, however, was carried out on a small scale physical model of the NFT flow system which was constructed at Lancaster following difficulties experienced during initial experiments on the full scale NFT system. In particular, the extremely slow dynamics of the full scale system necessitated unacceptably long experimental times, which interfered with the normal glasshouse operation and seriously limited the nature of the control studies.

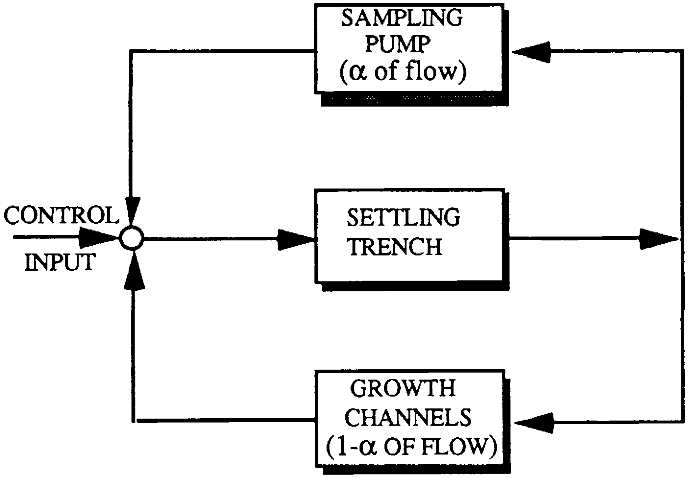


Fig. 11 Simplified block diagram of the NFT system.

Fig.11 shows the simplified block diagram of NFT sysytem, which is dominated, in a dynamic sense, by the nutrient delivery flow system, in which the nutrient mixture is circulated continuously through the parallel growth channels and the catchment trench. It was found that the simplest mathematical model which replicates the physical characteristics (pure advective time delay and dispersive time constant) of this flow system is the aggregated dead zone (ADZ) model developed at Lancaster to model the transportation of solutes in river channels (Beer and Young, 1983; Young and Wallis, 1986; Wallis et al., 1989;). In the laboratory, a physical version of the ADZ model can be constructed as the combination of a continuous stirred tank reactor (CSTR), to simulate the dispersive time constant, and an associated pipeline, to simulate the advective time delay. The time constant (or residence time) of the CSTR is then obtained from the relationship,

$$T = \frac{V}{Q} \qquad (18)$$

where, T is the time constant (sec), Q is the flow rate (ml /sec) through the CSTR, and V is "active mixing volume" (AMV) of the tank (see Young and Lees, 1992). The advective time delay is defined by Q and the pipe length.

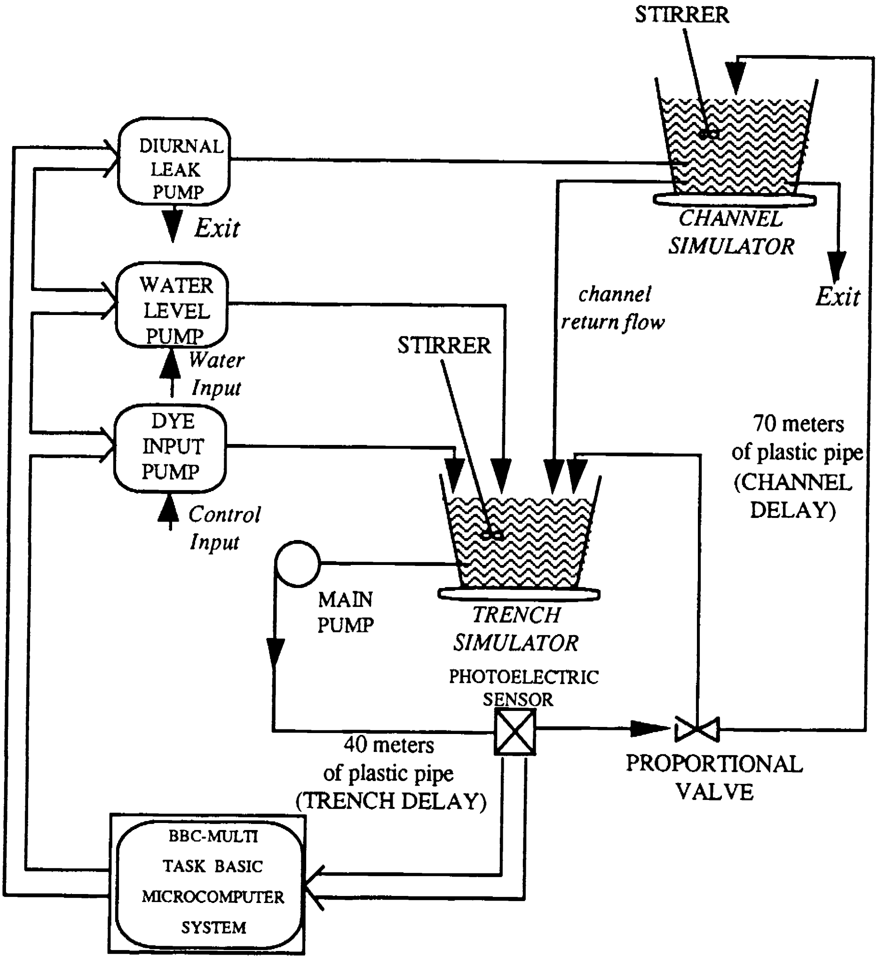


Fig. 12 The schematic diagram of the physical model of NFT flow system

A diagram of the complete model flow system is shown in Fig.12. The long time delays required to simulate the advective transport processes associated with the channels and the trench of the full scale NFT system are constructed from considerable (70 metre and 40 metre) lengths of plastic pipe. These connect the two CSTR's which simulate the dispersive characterists of the channels and trench, respectively. The pump rate can be varied to simulate major variation in the pure time delays; and a sinusoidal leak of liquid from the system is incorporated to simulate the diurnal losses of nutrient that would occur on the full size NFT plant, due to uptake by the growing plants. The nutrient itself is simulated by a black dye: the dye delivery system, which constitutes the main control signal to the system, is in the form of an electrically driven peristaltic pump; while the concentration of dye, which represents the system output, is measured by a photoelectric sensor positioned as shown in Fig.12.

The physical model in Fig.12 is "dynamically" similar to the full system in flow terms, but has much shorter dominant time constants and time delays (in the ratio 1:36, with a plant sampling interval of 25 seconds being equivalent to 15 minutes in the full system). The system dynamics may be easily altered by changing the flow rates, the volume of solution in any of the tanks, the system leak (simulating the rate of uptake), the return flow ratio valve, or any combination of these factors. The system has several advantages over the numerical computer simulation models: for example, it represents a full analog simulation of the flow system, it is subject to realistic inter-sample stochastic disturbances and non-linearities; and it requires control system implementation which is similar in many respects to that of the full NFT system.

Both open and closed loop experiments have been conducted on the physical flow model, with data acquisition and subsequent digital control based on a BBC microcomputer, using a multi-tasking operating system. The system input (the microcomputer generated signal to the peristaltic pump), and output (the dye concentration measured by the photoelectric system) provide the main data used in subsequent analysis and modelling. Various other data, such as the sinusoidal leak control signal and the flow rate of water through the system, are also monitored throughout the experiments.

The data from the experiments are processed in two ways. First, they are stored on disk in the BBC micro and presented graphically on the visual display, both for on-line monitoring of the experiment and, following a "screen dump" to the printer, for later reference. Secondly, they are transferred to a more powerful DOS based microcomputer, by Kermit, for later analysis and modelling studies using the microCAPTAIN (Young & Benner 1989) and TDC program packages . Since on-line recursive estimation algorithms are programmed in the BBC for self-tuning/adaptive control studies, it is also possible to monitor the parameter estimation results during the open or closed loop experiments.

7.3.1 PIP Control System Design based on the z^{-1} Operator Model

All the modelling analysis is based on input-output data monitored at a sampling interval of 1.25 mins. Full details of the identification and estimation results are given in Young et al.(1988) and Behzadi (1989). The most effective control model was identified as the following 5th order TF,

$$y(k) = \frac{G\,(b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3})}{1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + a_4 z^{-4} + a_5 z^{-5}}\,u(k) \qquad (19)$$

where:

$a_1 = -1.092$; $a_2 = 0.329$; $a_3 = -0.09$; $a_4 = -0.032$; $a_5 = -0.1$;

$b_1 = 0.009$; $b_2 = 0.07$; $b_3 = -0.032$ and $G = 3.44$.

Here G is the "system gain", which is defined as the overall scaling effect between the actual concentration levels in the system, and their numerical representations in the digital computer program. In this case, the black dye used to represent the nutrient is actually a suspension which leads to the thickening of the input solution towards the bottom of the supply container. In this sense, it is very difficult to exactly match the concentration of the input solution on replacement of the dye supply; or even maintain a uniform control input concentration during an experiment unless continuous, vigorous stirring is used. The system gain, G of the pilot plant is, therefore, generally unknown and time variable and has to be estimated by manual tuning, in a fixed gain controller, or by recursive estimation in a self-tuning or self-adaptive control system.

As might be expected, the impulse response of the model (19), shown in Fig.13, is similar to that of the high order computer simulation model and the real NFT flow system. The long term response is dominated by the first order mode with a very large time constant. This arises from the low losses in the system: except for the sinusoidal leak, all dye entering the system is retained, so that the system is always quite close to being an integrator. The shorter term behaviour is more interesting and is dominated by the circulatory flow system, with the three decaying peaks on the impulse response indicating that the initial impulsive input is being transported and progressively dispersed around the system.

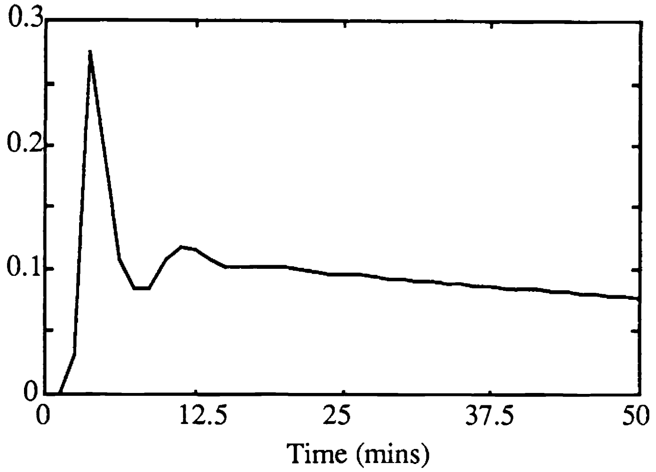


Fig. 13 Impulse response of NFT model (19).

Self-Tuning and Self-Adaptive Control

The NFT system is known to undergo dynamic changes, either because of the plant growth in the channels and the settling tank, or the changing behaviour associated with plant nutrient uptake. Our experimentation with NFT systems so far has indicated that such changes can result in radical model structure alterations, ruling out any assumption of the global reliability of any fixed gain controller. In other words, a "general "solution to the NFT system problem is unlikely from off-line modelling and control system design exercises, particularly if tight control is required.

In practice, therefore, a fixed gain approach may well result in the long term deterioration of the control system performance, to a degree which will depend on the complexity of the particular control system design strategy adopted. In such cases, unless the poorer performance can be tolerated in practice (as may be the case for the growth of some plants), a more sophisticated self-tuning or self-adaptive control system design must be adopted. Such a design can be achieved in two ways; either the system may be regularly remodelled off-line and the computer-based part of the control system re-designed if significant changes are encountered; or a complete, on-line, recursive model parameter estimation procedure can be used to continually update the PIP design. In this latter case, the implementation can involve self-tuning control (STC) which is reactivated at regular intervals; or full self-adaptive control (SAC), where the model parameters are all recursively updated, and new control gains computed, at regular intervals (e.g. each sampling instant).

Self-tuning and self-adaptive PIP controllers rely on the specification of a system model structure within the computer program. The parameters of this model structure, which is initially defined during the off-line modelling experiments that are an essential part of TDC design, are then estimated recursively based on present and past values of the measured system input and output signals. This time series analysis of the input-output data is normally carried by the recursive least squares (RLS) method; or alternatively, in the presence of heavy noise by the recursive instrumental variable (IV) approach). Using these recursive parameter estimates, a new PIP feedback gain vector is then computed, and the control system is suitably updated to reflect the changes in the gain vector.

The fifth order model (19), was found to be the best control model representation of the pilot scale NFT flow system. Based on this knowledge, therefore, a similar fifth order estimation model structure is specified for the STC system design, where the estimated parameter vector â is composed of the estimatews of the 8 unknown model parameters, i.e.,

$$\hat{a} = [\hat{a}_1 \ \hat{a}_2 \ \hat{a}_3 \ \hat{a}_4 \ \hat{a}_5 \ \hat{b}_1 \ \hat{b}_2 \ \hat{b}_3]$$

and the NMSS state vector x(k), is of the form (Chotai et al., 1991a; Young et al., 1987b, 1988),

$$x(k)=[y(k) \ y(k-1) \ y(k-2) \ y(k-3) \ y(k-4) \ u(k-1) \ u(k-2) \ z(k)]^T$$

where z(k) is the integral-of-error state defined by the following equation,

$$z(k) = z(k-1) + \{y_d(k) - y(k)\}$$

while $y_d(k)$ is the reference or command input at the kth sampling instant. With this model structure, we obtain a feedback control system with a 8-dimensional control gain vector, which is continuously updated as the recursive parameter updates are obtained from the RLS estimation algorithm. In this NMSS context, the control system design can be based on any SVF approach. These include pole assignment or optimal LQ or LQG control, as described above. The latter would provide a more natural setting for control system design if energy-efficient solutions were being sought.

Practical Results

Figs. 14 and 15 show two typical examples of the self-tuning PIP controller applied to the NFT pilot scale plant: in Fig. 14 the tuning is applied to all 8 parameters of the fifth order model. The initial fluctuations in the input signal are deliberately created to "excite" the system during the initial start-up and provide richer information to the recursive estimation algorithm.

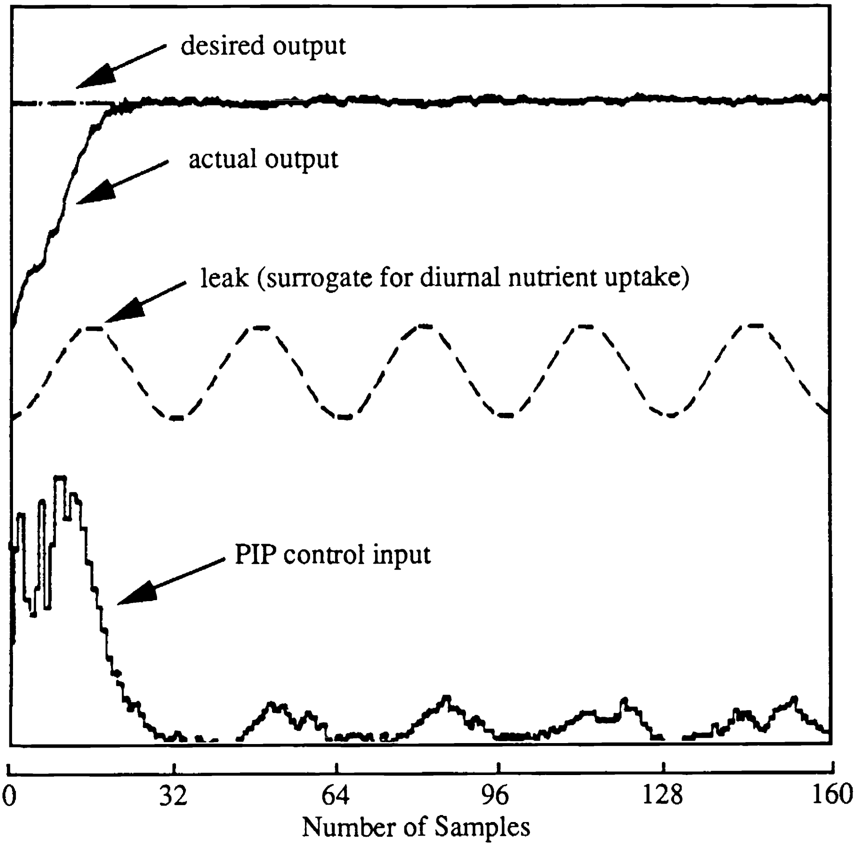


Fig. 14 Self-tuning PIP control of NFT pilot scale process: full model tuning

The results in Fig. 15 are obtained with a much simpler adaptive gain control system, in which only the gain G in (19) is recursively updated, with the other parameters set at their a priori, off-line estimated values. Clearly, both controllers maintain tight control of the set point despite the sinusoidal leak, although the full adaptive system is somewhat superior (note how well the control input counteracts the sinusoidal losses). Full self-adaptive control in the longer term is similarly effective but is difficult to illustrate because of the slow rates of change associated with the model parameters of the NFT system.

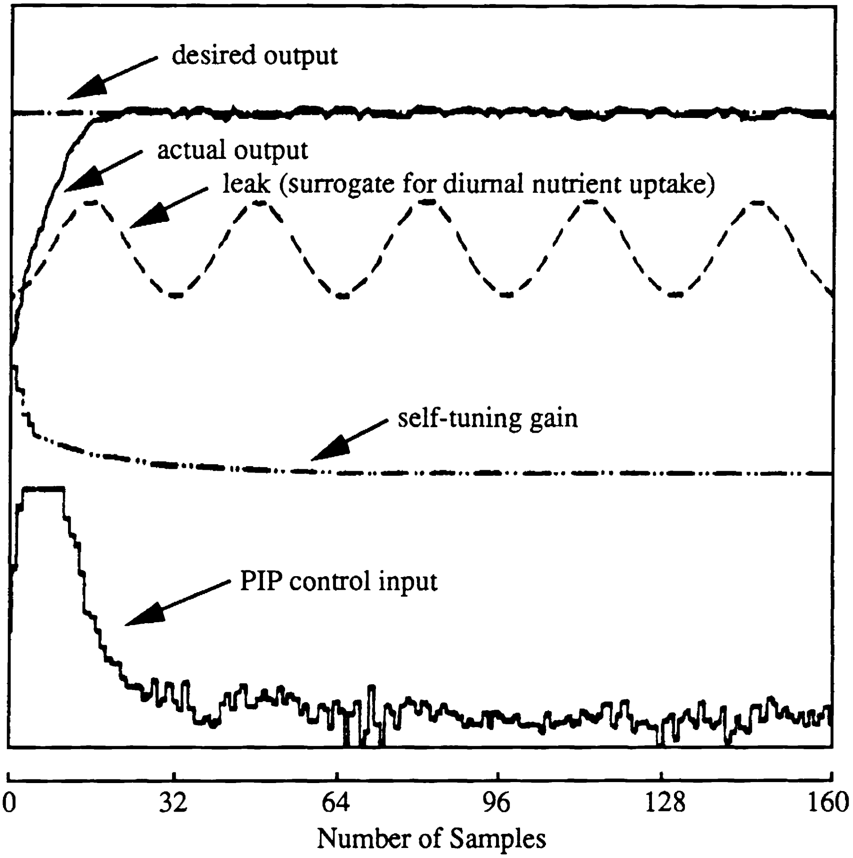


Fig. 15 Self-tuning PIP control based on the δ Operator Model

7.3.2 PIP Control System Design based on the δ Operator Model

For a sampling interval Δt defined as unity, for convenience, the δ operator SRIV algorithm in the TDC package identifies a 3rd order model (i.e. p=3 in equation (2)), with either zero or unity pure time delay. In the former case, the estimated model has the following non-minimum phase (nmp) form,

$$y(k) = \frac{-0.0228\,\delta^2 + 0.0219\,\delta + 0.1662}{\delta^3 + 1.1156\,\delta^2 + 0.6271\,\delta + 0.00411}\,u(k) + \xi(k)$$

where $\zeta(k)$ represents the noise on the system, including the effects of a sinusoidal leak introduced to simulate nutrient uptake by the plants. The covariance matrix P*(N) associated with the parameter estimates in this model has the form,

$$10^{-4}\begin{bmatrix} 24.10 & 14.91 & 0.1094 & -0.3926 & -0.2933 & 0.4180 \\ 14.91 & 27.41 & 0.1946 & 0.5077 & -0.6210 & 0.7586 \\ 0.1094 & 0.1946 & 0.0017 & 0.0039 & -0.0046 & 0.0058 \\ -0.3926 & 0.5077 & 0.0039 & 0.1610 & 0.0197 & 0.0145 \\ -0.2933 & -0.6210 & -0.0046 & 0.0197 & 0.0819 & -0.0177 \\ 0.4180 & 0.7586 & 0.0058 & 0.0145 & -0.0177 & 0.0217 \end{bmatrix}$$

where the sample size N=160. This model has similar dynamic characteristics to the previously estimated z^{-1} operator models, except that these former models had an explicit time delay of one sampling interval: here the nmp characteristics are effectively acting as a surrogate for this time delay, and the model was selected here in order to demonstrate that the PIP design procedure is applicable to nmp systems.

The initial PIP design results for this model, as obtained from the TDC package, are given in Fig.16. Here the control system is designed on an optimal LQ basis, with q_y,=1.0; q_u=0.5 and q_z=0.1. The closed loop step responses and pole locations are shown in 16(a) and (b), respectively. Parts (c) and (d) of the Figure illustrate the sensitivity of the design to the uncertainty on the parameter estimates: these were obtained using Monte Carlo stochastic simulation analysis, with the control system gains set at the nominal values obtained in the LQ design (and based on the nominal parameter estimates given above), but with the model parameters for each stochastic realisation selected from a Gaussian distribution with mean defined by the nominal parameter estimates and covariance defined by the associated covariance matrix P*(N) given above.

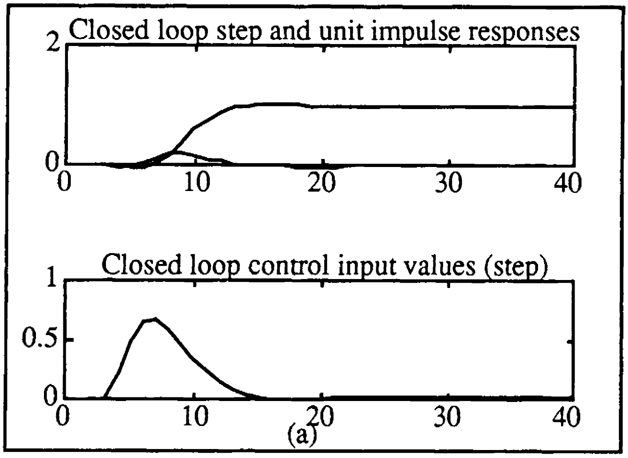


Fig. 16(a) TDC design results: Closed loop step and impulse responses

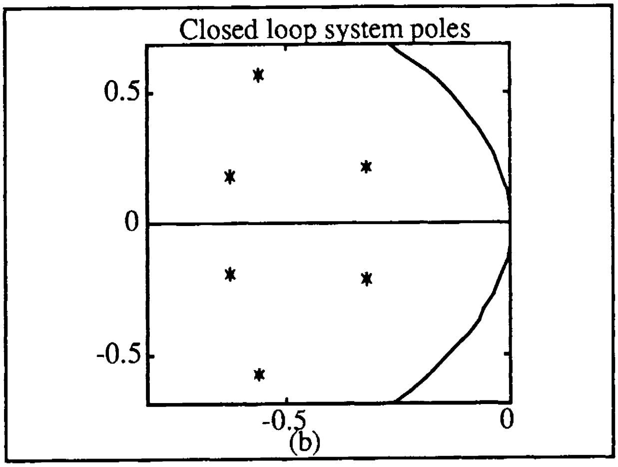


Fig. 16(b) TDC design results: Closed loop pole locations

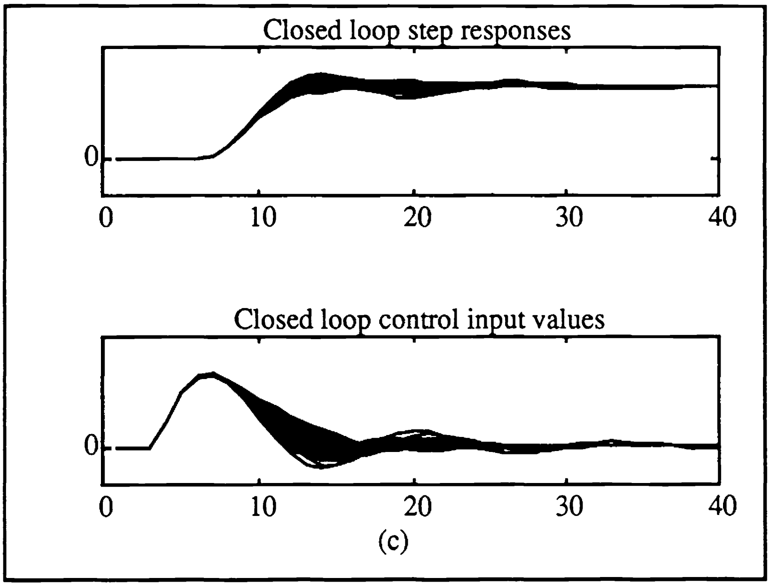


Fig. 16(c) TDC design results: Closed loop Monte Carlo Sensitivity analysis showing ensemble of step and impulse responses

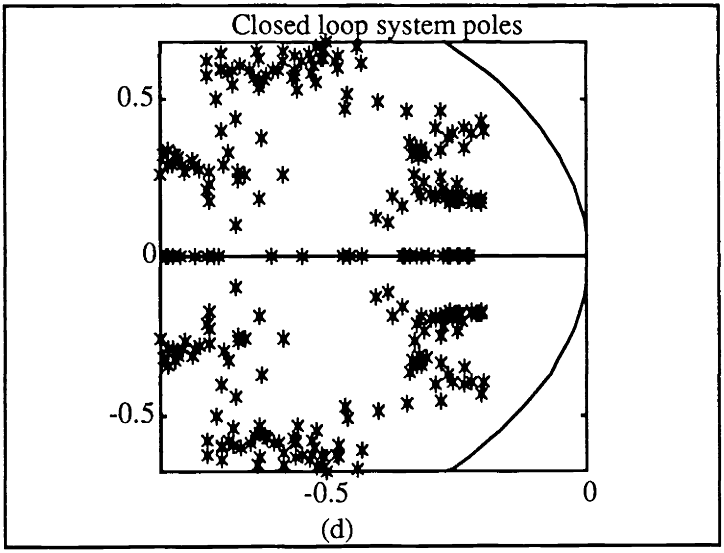


Fig. 16(d) TDC design results: Closed loop Monte Carlo Sensitivity analysis showing ensemble of closed loop pole positions ('stochastic root loci')

The resulting ensemble of step/impulse responses and "stochastic root loci" given in Figs. 16(c) and (d), respectively, all indicate that, while the nominal design has very good response characteristics, as shown in Figs.16(a) and (b), it is a rather sensitive to the inherent uncertainty in the model. In particular, there is clearly some tendency for the oscillatory mode (which arises from the recirculation in the flow system) to become excited during a small proportion of the stochastic realisations. This suggests that this initial design should be rejected and the design procedure repeated, with modified weightings in the performance index, in order to obtain greater stability margins and less sensitivity to uncertainty.

This example of δ operator PIP control system design for the NFT system is included merely for illustration: in this example, the coarsely sampled, z^{-1} operator design is quite adequate, as shown in Figs. 14 and 15, and provides a good basis for implementation in the glasshouse. The d operator design, suitably modified to increase its robustness to uncertainty, performs similarly but appears to offer no obvious advantages over the z^{-1} operator design . However, it would provide a more suitable design approach if a higher frequency sampling strategy was selected.

8. CONCLUSIONS

This paper has outlined a new True Digital Control (TDC) philosophy for the design of automatic control systems and has demonstrated its potential within the context of glasshouse horticulture. The TDC approach is best exemplified by the Proportional-Integral-Plus (PIP) design procedure developed at Lancaster over the last few years. The PIP control system provides a sophisticated alternative to the ubiquitous PI and PID controllers that have so dominated control system design for the past half century. In its simpest form, the PIP controller is similar in structure to the conventional PI alternative, but its model-based design, which exploits a special Non-Minimum State Space (NMSS) formulation of the model, ensures that the complexity of the controller always matches the complexity of the system being controlled, so leading to inherently more powerful and robust designs.

In this chapter, only single input-single output systems described by transfer functions in either the backward shift or the delta operator have been discussed. Since the PIP procedure is based on the powerful NMSS concepts, however, it can be extended straightforwardly to multivariable and stochastic systems (Wang, 1988; Chotai et al, 1991d, 1992). And, as the examples in the Chapter have demonstrated, when combined with on-line, recursive techniques, such as the SRIV estimation algorithms, it provides an excellent basis for the design of self-tuning and self-adaptive control systems (e.g. Young et al, 1988, 1991a,b; Young, 1989; Chotai et al, 1991a).

ACKNOWLEDGEMENTS

The research described in this paper was supported by the U.K. Science and Engineering Research Council (SERC) under grant no. GR/E 8156.2 and is being continued under SERC grant no. GR/G 5386.6 and the U.K. Agriculture and Food Research Council (AFRC). under grant no. LR89/314. The authors are most grateful for this support and also to colleagues at the AFRC Engineering Institute, Silsoe, who collaborated closely in the research programme and are responsible for the detailed glasshouse temperature control system design.

REFERENCES

Bailey, B. J. (1985) Wind dependent control of greenhouse temperature, Acta Horticulturae, 174, 381-386.

Behzadi, M.A., (1989) True digital control of greenhouse systems, PhD Thesis, Univ. of Lancaster.

Beer, T. and P. C. Young (1983) Longitudinal dispersion in natural streams, American Society of Civil Engineers Journal of Environmental Engineering, 109, 1049-1067.

Chotai, A., P.C. Young and W.Tych (1990). A non-minimum state space approach to true digital control based on the backward shift and delta operator models, appears in M.H. Hamza (ed.) Proc. IASTED Conference, Acta press, Calgary, 1-4.

Chotai, A., and P.C. Young. (1991) Self-adaptive and self-tuning control of a Nutrient Film Technique (NFT) system, in Y. Hashimoto and W. Day (eds.), Mathematical and Control Applications in Agriculture and Horticulture, Pergamon

Press, Oxford, 33–39.

Chotai, A., P. C. Young and M. A. Behzadi. (1991a). The self–adaptive design of a nonlinear temperature control system, Proc. I.E.E. , Pt.D., Special Issue on Self Tuning Control, 138, 41–49.

Chotai, A., P.C. Young and W.Tych (1991b). A Non–Minimum State Space approach to True Digital Control Based on d Operator Models, IEE Conference Publication No. 332, Control 91, 567–573

Chotai, A., P.C. Young, P. Davis and Z. Chalabi, (1991c) True Digital Control of Glasshouse Systems, in Y. Hashimoto and W. Day (eds.), Mathematical and Control Applications in Agriculture and Horticulture, Pergamon Press, Oxford, 41–45.

Chotai, A. Young, P.C. & Tych, W. (191d) Proportional–Integral–Plus (PIP) Controllers for Multivariable Discrete–Time Systems, in 8th International Conference on Systems Engineering Proceedings, Coventry Polytechnic, England, pp. 397–404.

Chotai, A. Young, P.C. & Tych, W. (1992). Dynamic Decoupling, Pole–Assignment and Servomechanism Design for multivariable NMSS Discrete–Time Systems, to appear in Sixth IMA conference on Control: Modelling, Computation and Information, UMIST.

Cooper, A. J. (1967) Tiered through and nutrient film methods of tomato growing, Annual Report of the Glasshouse Crops Research Institute, 59.

Davis, P. F. (1986) Pers. Comm. relating to confid. paper by Davis and Hooper, Bedfordshire, England.

Davis P. F., and Hooper A W (1991) Improvement of greenhouse heating control. Proc. I.E.E. , Pt.D. 138, 249–255.

Goodwin, G.C., R.H .Middleton, and M. Salgado(1988) A unified approach to adaptive control, in Warwick, K. (ed.): Implementation of self–tuning controllers, Peter Peregrinus, London, 126–139.

Graves, C. J. (1983) The nutrient film technique, Horticultural Reviews, 5, 1–10.

Hurd, R. D. and C. J. Graves (1984) The influence of different temperature patterns having the same integral on the earliness and yield of tomatoes, Acta Horticulturae, 148, 547–554.

Moler C., J. Little and S.Bangert (1987). Pro–Matlab Users Guide, The Mathworks Inc., South Natick, MA.

Middleton R.H. and G.C.Goodwin (1990). Digital Control and Estimation - a Unified Approach, Prentice Hall, New Jersey.

Miller, W. B., R. W. Langhans, and L. D. Albright (1985) Plant growth under averaged day/night temperatures, Acta Horticulturae, 174, 313-320.

Saffell, R. A., and B. Marshall (1983) Computer control of air temperature in a glasshouse, Jnl. of Agricultural Engineering Research, 28, 469-478.

Sanger, T. R (1985) Measurement and control of temperature, humidity and carbon dioxide concentration in greenhouses. In: Royal Agricultural Society of England, Monograph Series No. 4, Farm Electron. Comp., pp. 269-287.

Tych, W., A.Chotai and P.C.Young (1991a). COMPUTER AIDED DESIGN PACKAGE FOR TRUE DIGITAL CONTROL (TDC) SYSTEMS, IEE Conference Publication No. 332, Control 91., 288-293.

Tych, W., Young, P.C. & Chotai, A. (1991b) TDC: A Computer aided control design package for true digital control, appear in H.A. Barker (ed.), IFAC Symposium on computer aided design in control systems, Pergamon press: Oxford, 454-459.

Wallis, S. G., P. C. Young and K. J. Beven (1989) Experimental investigation of the Aggregated Dead Zone model for longitudinal solute transport in stream channels, Proc. Inst. Civ. Engrs. (U.K.), 87, 1-22.

Wang, C. L. (1988) New methods for the direct digital control of discrete-time systems, PhD Thesis, University of Lancaster.

Wang, C.L., and P.C. Young (1988), Direct digital control by input-output, state variable feedback: theoretical background, Int. J. Control, 47, 97-109.

Winsor, G. W. and D. M. Massey. (1978) Some aspects of the nutrition of tomatoes grown in recirculating solution, Acta Horticultural, 82, 121-132.

Young, P.C. (1964) In-flight dynamic checkout, I.E.E.E. Trans. on Aerospace, AS2, 1106-1111.

Young, P.C. (1965) The determination of the parameters of a dynamic process, Radio and Electronic Engineer, 29, 345-362.

Young, P.C. (1966) Process parameter estimation and self adaptive control; appears in P.H. Hammond (ed.). Theory of Self Adaptive Control Systems, Plenum Press, New York, 118-140.

Young, P.C. (1969a) The Differential Equation Error Method of Process Parameter Estimation, Ph.D Thesis, Department of Engineering, University of Cambridge.

Young, P.C. (1969b, 1970) An instrumental variable method for real time identification of a noisy process, Proc. IFAC Congress, Warsaw (also appears in Automatica, 6, 1970, 271-287).

Young, P. C. (1981) A Second Generation Adaptive Autostabilization System for Airborne Vehicles Automatica, 17, 459-470.

Young P.C. (1984). Recursive Estimation and Time Series Analysis, Springer Verlag, Berlin.

Young, P.C.(1985). The instrumental variable method: a practical approach to identification and system parameter estimation, in Barker, H.A. and Young, P.C. (Eds.), Identification and System Parameter Estimation, Pergamon, Oxford, 1-16.

Young, P.C.(1989). Recursive estimation, forecasting and adaptive control, in C.T. Leondes (ed.), Control and Dynamic Systems, 30, Academic Press, San Diego, 119-166.

Young, P.C.(1991) Simplified Refined Instrumental Variable (SRIV) estimation and True Digital Control (TDC): a tutorial introduction, appears in Proceedings of the First European Control Conference, Grenoble, 1295-1306.

Young P.C., and S. Benner (1989). microCAPTAIN 2.0 User Handbook, CRES, Lancaster University.

Young P.C. and M. Lees, (1992) The active mixing volume: a new concept in modelling environmental systems, to appear as chapter in: V. Barnett and R. Turkman (eds.), Statistics in the Environment, J. Wiley, Chichester.

Young, P. C. and S. G. Wallis. (1986) The Aggregated Dead Zone (ADZ) model for dispersion in rivers, Proc. BHRA Int. Conf. on Water Quality Modelling in the Inland Natural Environment, Bournemouth, England, 421-433.

Young, P.C., and J. Willems (1972) An approach to the linear multivariable servomechanism problem, Int. J. of Control, 15, 961-979.

Young, P. C., Behzadi, M.A., and Chotai, A. (1988) Self tuning and self adaptive PIP control systems; appears in K. Warwick (ed.). Implementation of Self-Tuning Controllers. Peter Perigrinus, London, 220-259

Young, P.C., Chotai, A., & Tych, W. (1991a) True Digital Control: A Unified Design Procedure for Linear Sampled Data

Systems, appears as Chapter in K.Warwick, M.Karny and A.Halouskova (eds.), Advanced methods in adaptive control for industrial applications, Lecture Notes in Control and Information Sciences, vol. 158, springer-Verlg: Berlin., pp 71-109.

Young, P.C., A. Chotai, and W. Tych, (1991b) Identification, estimation and control of continuous-time systems described by delta operator models, Chapter in N. K. Sinha and G. P. Rao (eds.), Identification of Continuous-Time Systems, Kluwer Academic Publishers, Dordrecht, 363-418.

Young, P. C., M. A. Behzadi, A. Chotai and P. Davis. (1987a). The modelling and control of nutrient film systems. Chapter in J. A. Clark, K. Gregson and R. A. Scafell (ed.). Computer Applications in Agricultural Environments. Butterworth, London, 21-43.

Young P.C., M.A. Behzadi, C.L.Wang and A.Chotai (1987b). Direct digital and adaptive control by input-output state variable feeback pole assignment, Int. J. of Control, 46, 1867-1881.

Chapter 2

PHYSICAL MODELLING OF GREENHOUSE CLIMATE

Gerard P.A. Bot

Institute of Agricultural Engineering (IMAG-DLO)
P.O. Box 43
6700 AA Wageningen
The Netherlands

I. INTRODUCTION

A greenhouse is designed as light transparent shelter to improve environmental conditions for plant production. The history of constructions protecting plants from a hostile environment allowing out of season production of fruits and flowers goes back to ancient times. Stanghellini (1987) refers to Roman history and cites Seneca who in 63 A.D. complained that the availability of roses in winter from protected cultivation was an aspect of living against Nature. Also Van den Muijzenberg (1980) in his history of greenhouses pointed out that through the centuries many innovations were made to improve the performance of shelters for plant growth. Especially the improvement of light transmission was a leading item. So one could conclude that in our times nothing new is under the sun. An important new aspect however is that modern technology allowed the construction of relatively cheap greenhouses with high light transmission togethe r with the development of continuous and automatic control of the various environmental factors affecting crop growth and development. In this way greenhouse crop production could develope to a high investment, almost industrial activity, though the average scale is still that of a family owned enterprise.

From the physical point of view one could consider a modern greenhouse as a bioreactor in which a crop produces a marketable product from carbondioxide and water containing fertilizer with the aid of shortwave radiation. The set of environmental factors in the "reactor", referred to as the greenhouse climate, affects the exploitation of the growing factor light. So it is extremely important to control the greenhouse climate in order to have a maximal effect of the production factor light. However the performance of the crop in terms of production in interaction with the environment is of an extreme complex nature. More and more information is gathered on the relation between crop growth and environmental conditions as one aspect of production (Bot and Challa, 1991). This opens the possibility of growth optimization to replace the control of a desired temperature and humidity level in the greenhouse by an optimization scheme calculating these levels by the system (Challa and others, 1988). To do this in a proper way, both the production level and the production costs of the reactor have to be known. For the production level the physical environment in the greenhouse has to be quantified. For the production costs especially energy flows connected to energy consumption have to be known. Physical modelling is a tool to quantify these various factors and to understand the interaction of interacting processes.

Physical modelling of greenhouse climate started already in the seventies. Bot (1983) reviewed the literature up till then. However, the information available on the various processes involved was diffuse, not aimed at the before mentioned purpose or lacking at all (Udink ten Cate, Bot and Van Dixhoorn, 1978). Bot (1983) presented a physical model based on in-situ measurements of the main physical processes. Some of these processes are studied later in more detail, like crop transpiration (Stanghellini, 1987) and natural ventilation (De Jong, 1990). The model or adaptations of it are applied nowadays as design tool and in optimization studies (Houter and others, 1989; Houter, 1990; Gijzen and Ten Cate, 1988; Gijzen, Vegter and Nederhoff, 1990; Van Henten and Bontsema, 1991). De Zwart (1992) reported the most recent versions with adaptations and documented the various aspects and details of it in a carefull way.

In this paper the general approach is outlined.

II. GENERAL ASPECTS

The set of momentaneous environmental factors inside the greenhouse affecting crop growth and development is referred to as the greenhouse climate. As the meteorological term climate indicates a long term average of environmental factors and the term weather is used for the momentaneous set, the proper term would be greenhouse weather. However it is common practise to use the term greenhouse climate in this respect.

The difference between greenhouse climate and outside weather (sometimes called greenhouse effect) is mainly caused by two mechanisms:

- The first is the enveloping of air: The air in the greenhouse is stagnant due to the enclosure. So the exchange of the greenhouse air with the surrounding (outside) air is strongly decreased compared to that of the air without envelope. Moreover the local air velocities are small compared to that in the open. The reduction of the air exchange (or ventilation) directly affects the energy and mass balances of the greenhouse air while the smaller local air velocities affect the exchange of energy, water vapour and carbondioxide between the greenhouse air and the greenhouse inventory (crop, soil surface, enclosure and heating system).
- The second is the mechanism of radiation: The inward shortwave radiation (direct from the sun and scattered from the sky and clouds) is decreased due to the light interception by the opaque and transparent components of the greenhouse while the long wave radiative exchange between inside and outside the greenhouse is changed due to the radiative properties of the covering materials. With glass as covering material this leads to the mouse trap theory: glass is (partly) transparent for the incoming shortwave radiation and opaque for the longwave radiation emitted from the interior, so the energy is trapped. However this effect is of minor importance to explain the increased air temperature in the greenhouse. The common name greenhouse effect for the mouse trap theory is therefore misleading. Nevertheless the radiative effects are indispensable to describe the greenhouse climate because they directly affect all energy balances and therefore the inside temperatures.

The general mechanisms may be relatively simple, the impact on the greenhouse climate itself is of a more complex nature. Therefore the main physical processes for the exchange of energy and mass are discussed seperately first together with the physical characterization of the exchange with the crop. Then these processes are combined to a lumped parameter physical model that can be used to simulate dynamically the greenhouse climate.

III. PHYSICAL PROCESSES

III.1 General overview

The various physical processes can be schematised according to Figures 1 and 2. To keep it simple, in this scheme the greenhouse is considered to be a one dimensional system in which only vertical energy and mass (water vapour and carbon-dioxide) flows occur. This can be done for a description of the dynamic behaviour of the system i.e. the various state and flow variables as function of time. This can be needed for the application as a design tool or for optimal control purposes. When the dynamic behaviour can be charactized by only a few state variables, like in most control applications, then the physiscal model can be simplified or modelling approaches defining input-output relations can be applied. In general the experimental verification can be much simpler in this case. If one is interested in the distributed nature of the greenhouse climate rather than the dynamic nature, then another approach has to be followed (Bot, 1989b).

The major greenhouse components in the one-dimensional system are considered to be homogeneous. The main parts are the greenhouse cover, the greenhouse air, the crop and the soil. Of these parts the soil is non-homogeneous, so has to be composed of more layers. A point of discussion will be the homogeneity of the crop. For the calculation of the physical environment including crop transpiration one crop compartment will do. For the calculation of crop photosynthesis, in which radiation has a strong non-linear effect, the radiation distribution in the crop has to be calculated, leading to a multi-layered crop (Gijzen and Ten Cate, 1988).

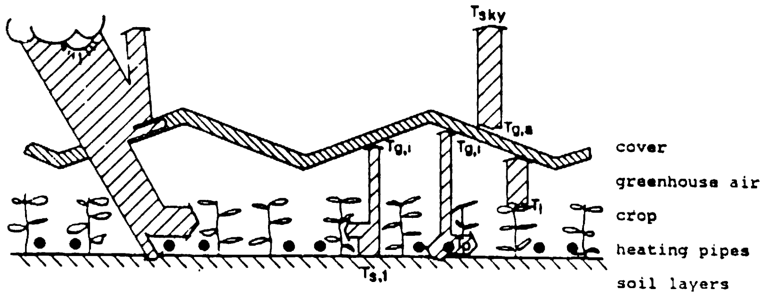


Figure 1. Radiative processes: interception of direct and diffuse solar radiation by the greenhouse cover, the crop and the soil surface (left) and thermal radiative exchange between the various greenhouse components and with the sky (right).

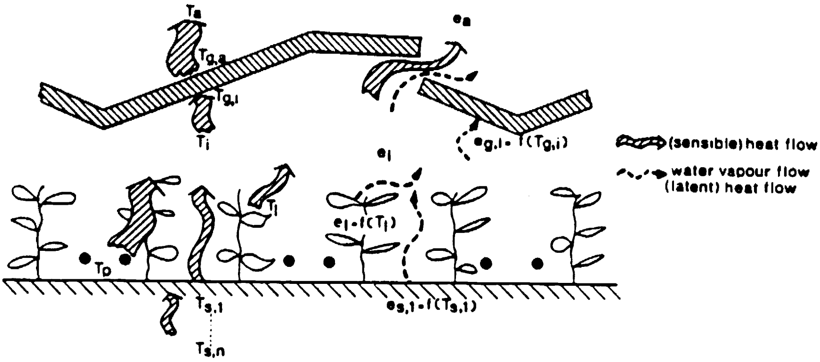


Figure 2. Energy exchange (sensible and latent heat) between the various greenhouse components and to the outside air.

In Figure 1 the radiative processes are outlined, in Figure 2 the various convective and conductive exchange processes. The main processes will be discussed separately in the following sections with the emphasis on physical understanding.

III.2 Radiative exchange

The production in a greenhouse is aimed at an optimal exploitation of incoming direct (originating from the solar position) and diffuse (scattered in the atmosphere and by the clouds) solar radiation as driving force for the photosynthesis process. For 99 % this incoming solar energy flux is within the wavelength region between 300 and 3000 nm. For planth growth a special part of the spectrum in the visible region is of interest, the so called photosynthetic active radiation (PAR) with wave length region in the visible spectrum between 400 and 700 nm. In this region about half of the solar energy is irradiated. Only a small part of the PAR energy is directly converted into the photosynthesis process. The spectral distribution of solar irradiance at earth surface is well known for various meteorological conditions (e.g. Kondratyev (1972)). This general characterization is accurate enough to estimate the ratio between PAR and global radiation for outside coditions, but the effect of the greenhouse cover in the translation to the greenhouse crop situation needs special attention and separate measurement for validation.

The total solar radiation (global radiation) at crop level contributes to the energy balance of the crop so affects crop temperature and transpiration. The energy converted in the photosynthesis process can be neglected in the energy balance over the crop. To translate both direct and diffuse solar radiation at earth level to that inside the greenhouse at crop level is of great importance for both the physical modelling and the quantification of the photosynthesis. The interaction of the solar radiation with the greenhouse cover determines how much radiation is transmitted and available at crop level. This interaction can be calculated from the basic optical laws for reflection, absorbtion and transmission of transparent layers and opaque construction parts. Therefore the optical properties of the transparent materials, the angle of incoming radiation relative to the observed surface and the geometry of the construction have to be known. For the direct light, the angle between radiation and surface follows from the solar position (given by the latitude of the observed greenhouse and the time and date) and of the orientation and geometry of the surfaces. For the diffuse radiation it follows from the distribution of the radiation intensity over the hemisphere, which is different for various meteorological conditions. The most striking difference is that between a clear and cloudy sky. Bot (1983) described the considerations given above in full detail. Recently Critten (1992) reviewed the literature on greenhouse light transmission.

Comparison between the model calculations and measurement of transmission factors gave an agreement for both diffuse and direct radiation within a few percents of transmission. Therefore the discussed light transmission model of greenhouses is an operational part of crop production models (Gijzen and Ten Cate, 1988; Gijzen, Vegter and Nederhoff 1990) and of greenhouse climate models (Bot, 1989a: Houter, 1990).

The thermal radiation exchange from the greenhouse interior to outside is shielded by the cover. The efficiency of this shield depends on the radiative properties in the thermal wavelenght region (about 5000–50000nm for environmental temperatures). Inside the greenhouse the various surfaces exchange thermal radiation. The spectral distribution of the radiated energy is given by Planck's law. For our purposes the energy exchange is of importance. This can be calculated according to the well known non-linear Stefan–Boltzmann relation, implementing emission coefficients to account for the radiative properties of the surfaces and view factors to account for the mutual interception of the radiation between the surfaces:

$$Q_{rad,1,2} = \varepsilon_{12} F_{12} \sigma \{(T_1)^4 - (T_2)^4\} \qquad (1a)$$

where $Q_{rad,1,2}$ is the energy flux density due to thermal radiation from surface 1 to 2 (Wm^{-2}), ε_{12} the combined emissivity between the surfaces (–), F_{12} the view factor from surface 1 to surface 2 (–), σ the Stefan–Boltzmann constant ($5.67\ 10^{-8}\ Wm^{-2}K^{-4}$) and T_1, T_2 the temperatures of surface 1 and 2 respectively (K). The emission coefficients of most surfaces are near to one, but sometimes special low emission materials are used, so it is preferred to set up the equations in a general way. The combined emissivity ε_{12} is found from the individual emission coefficients ε_1 and ε_2. For parallel surfaces, as found in the greenhouse, this is according to:

$$\varepsilon_{12} = (\varepsilon_1^{-1} + \varepsilon_2^{-1} - 1)^{-1} \qquad (1b)$$

For relatively small temperature differences, $d(T^4)$ can be linearized to $(4T^3)dT$ leading to:

$$Q_{rad,1,2} = \varepsilon_{12} F_{12} \sigma (4T^3)(T_1 - T_2) \qquad (1c)$$

where T is the average temperature (K). In this way the energy flux density is related to the temperature difference as the driving force for the energy exchange. The proportionality factor $[\varepsilon_{12} F_{12} \sigma (4T^3)]$ only varies slightly with the temperature level which is expressed in Kelvin. In this way radiative exchange can be compared to the other exchange mechanisms as defined in the following sections.

Of course the view factors for the thermal radiation exchange between the crop and the other greenhouse parts are dependent on the development of the crop. They were determined in a separate study on the physical behaviour of the crop (Stanghellini, 1987). Crop development then is expressed in the physical quantity "leaf area index", in literature abbreviated as LAI. It is defined as the ratio between total leaf area per square meter ground surface and has to be measured seperately. Calculations and measurements of the energy exchange in the thermal wavelength region were in good agreeement.

A special difficulty is the characterization of the radiative exchange between the cover and the hemispere. In reality this is a complex exchange between the cover and the various layers of the atmosphere. The composition of the atmosphere determines absorbtion and emission at various heigts and the atmospheric temperatures at these heights determine the temperature differences with the cover. So for a full description both the atmospheric composition and the temperature as function of the height has to be known. To overcome this complex problem a sky vault temperature T_{sky} is defined as a temperature of a black hemisphere (so emission coefficient equals one) exchanging thermal radiation with the greenhouse cover according to Stefan-Boltzmann to the same extent as the atmpospheric exchange. This sky vault temperature can be measured directly by a pyrgeometer but it is not included in the list of standard meteorological measurements. In literature (e.g. Wartena et al., 1973) correlations of experimental data are reported to relate sky vault temperature to standard meteorological observations as air temperature, humidity and cloudiness, but these relations are only applicable for an average sky temperature over a long period and only valid for regions with the same meteorological characteristics as the region in which the measurements are performed. However for realistic dynamic modelling, sky temperature is a dynamical boundary condition and an important one for the energy budget and the greenhouse climate. Therefore in our experiments it was calculated from real time measurements of the thermal radiation exchange between a surface with known temperature and the hemisphere. Nowadays we measure it continuously.

III.3 Ventilation exchange

The greenhouse enclosure envelopes the interior and prevents mixing of the internal with the external air. Direct exchange of air through openings in the enclosure (leaks and ventilation windows) is called ventilation and can be expressed in terms of volumetric flow (m^3s^{-1}). This volumetric flow carries energy and mass. While this ventilation process affects the most important greenhouse effect, i.e. the enveloping of air, a proper description of the dependency of the ventilation flux on external and internal factors is a prerequisite for a description of greenhouse climate.

In general, ventilation flux is the flow of air from inside to outside and reverse through openings. This flow has to be generated by some pressure difference between both sides of the opening. The pressure difference can be due to the effect of the outside wind (wind effect) or due to the density differences (generated mainly by temperature differences and to a much lesser extent by concentration differences) between internal and external air (temperature efffect). For large greenhouses, with the ventilation windows distributed in the cover, the pressure differences generated by the wind field proved to be of a fluctuative nature. So the ventilation is fluctuative too, but can be described as an effective semi-constant flow from in- to outside and reverse due to an effective pressure difference. The flow through the opening, with known pressure difference between both sides of the opening, is dependent on the flow resistance of the opening itself. In literature on the ventilation of buildings, specific overall relations are deducted for the ventilation rate through specific openings as function of relevant parameters. Bot (1983) adopted the approach in which the wind and temperature effects and the flow characteristics of the openings are discussed seperately. This aproach was based more soundly and validated in more detail by De Jong, 1990. From these studies the effective ventilation volume flux Φ_v (m^3s^{-1}) were expected to be linear proportional to the outside wind speed u at reference height (normally 10 m) and the area of ventilation windows A_0 for any window opening angle ß. Full scale measurements of the ventilation flux with a tracer gas technique proved these relations and validated them for various conditions and window types. Due to the linear proportionalities a dimensionless ventilation number G can be defined as:

$$G = (\Phi_v\ u)/A_0 \qquad (2)$$

In Figure 3 a typical ventilation characteristic is given for a commonly used window type in terms of the relation G(ß).
Also the ventilation due to temperature differences between the in- and exterior can be quantified. In this case the vertical height in the opening or for various openings also the vertical distance between these openings (chimney effect) is of importance. However ventilation due to wind effects will be dominant already at low wind speeds of about 2-3 ms^{-1}. The temperature effect will therefore be relevant under special low wind, high radiation conditions. For the description of this special effect we refer to the literature already mentioned.

With the known dependency of ventilation flux on wind speed, temperature difference between in- and outdoor, window characteristics and window opening the exchanged energy q_{vent} (W) and mass m_{vent} (kgs^{-1}) (water vapour, CO_2, etc) between the greenhouse interior and the ambient can be calculated as respectively:

$$q_{vent} = \Phi_v\ Cap_{air}\ (T_i - T_a) \qquad (3)$$

$$m_{vent} = \Phi_v\ (e_i - e_a) \qquad (4)$$

where Cap_{air} is the volumetric specific heat of the greenhouse air ($Jm^{-3}K^{-1}$), T_i and T_a the in- and exterior temperature respectively (K or °C) and e_i an e_a the in- and exterior concentration of the gas component considered (kgm^{-3}).
So the non-linearity of ventilation exchange is in the non-linear dependency of Φ_v; at constant Φ_v the exchange of energy and mass is linear proportional to the driving force..

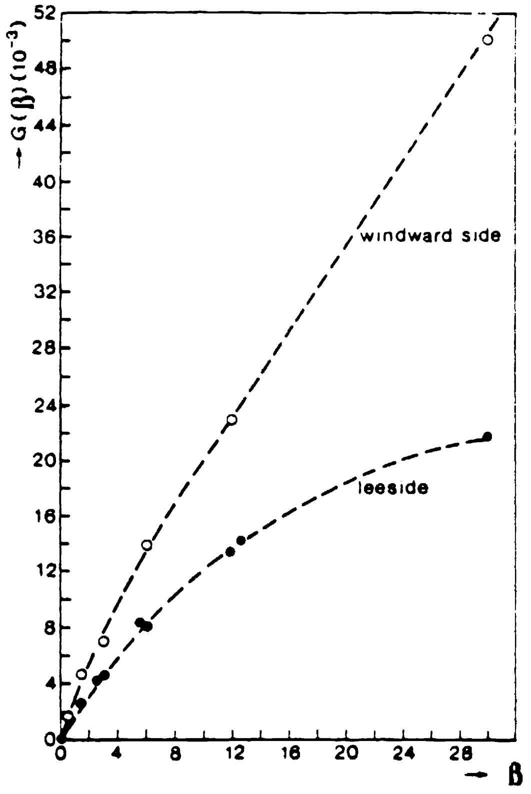


Figure 3. Ventilation number G as function of the window opening angle ß of a typical ventilation window with a width of two glass panes covering half the distance between ridge and gutter and mounted at the ridge in a Venlo-type greenhouse. Relations are given for windows opened at the lee-side and the windside.

III.4 Crop transpiration

All greenhouse activities are aimed at an optimal growth and development of the crop as the supplier of greenhouse products. In the greenhouse climate model the impact of the greenhouse climate on the growth and development of the crop is out of the scope; the vegetation is considered as a well-defined body that exchanges energy, water vapour and carbondioxide with the greenhouse air. The greenhouse climate and the plants interact in this respect. Generally speaking, without plants the greenhouse climate can be characterised as a hot and dry desert climate. With crop it is transformed into a humid and warm (sub)tropical climate. The physical interaction between plants and environment can explain how this transformation is accomplished.

The transport of energy from the leaf is in general defined in the same way as the heat transfer from other surfaces, like the cover. The exchange of water vapour is more complex. Water evaporates in the internal water saturated cavities of the leaves with saturated vapour concentration or pressure as given by the leaf temperature, is transported to the surface of the leaves via the stomata, meeting a stomatal resistance and then transported from the surface to the greenhouse air meeting some boundary layer resistance. The stomata are also the openings for the absorbtion of CO_2 to the leaf tissue to be converted in the photosynthesic process. The stomatal openings and so the stomatal resistances are controlled by plantphysiological processes in a complex way. For a physical model a description of the plant-physiological reactions is too complex. There a stomatal reaction in terms of measurable climate factors is needed. Moreover the effective stomatal resistance of the entire crop, called the crop resistance, has to be known in stead of individual resitances. This crop resistance for was determined by Stanghellini (1987) as a function of the environmental conditions, translating the plant-physiological reactions of the crop to the environment into a phenomenological-physical model. The boundary layer resistance depends on the local air movement and temperature differences and differs quantitatively from the extensively reported outdoor conditions. Also this resistance for greenhouse crops was reported by Stanghellini (1987).

The crop transpiration m_{tr} can be described in terms of the total resistance r_{tot} and the vapour concentration difference (e_l - e_i) as driving force as:

$$m_{tr} = r_{tot}^{-1} (e_l - e_i) \qquad (5)$$

This is a linear relation if the resistance is constant. The strong non-linearity is in the resistance. Often vapour pressure difference p_l - p_i is used as driving force. While water vapour is behaving as an ideal gas, concentration e and pressure p can be easily linked:

$$p = (RTM^{-1})\, e \qquad (6)$$

where R is the universal gas constant (8314 $Jkmol^{-1}K^{-1}$) and m the molar mass ($kgkmol^{-1}$).

To evaporate the water flow from the crop to the greenhouse air the energy E for this evaporation (latent energy) due to the heat of evaporation H has to be accounted for in the energy budget of the crop:

$$E = H\ m_{tr} \qquad (7)$$

III.5 Exchange between the cover and the air

The greenhouse air exchanges energy and water vapour (condensation) to the inner surface of the cover and the cover exchanges energy to the outside air. The mechanism of these exchanges is that of convection. At the inside natural convection is expected due to low local air velocities generated by the appearing temperature differences, at the outside forced convection is expected due to local air velocities generated by the wind field at the outside. To characterize the convective exchange a heat transfer coefficient is defined as the ratio between the heat flux density Q_{conv} (Wm^{-2}) and the temperature difference as driving force for the exchange:

$$Q_{conv} = \alpha\ (T_i - T_{g,i}) \qquad (8)$$

where T_i and $T_{g,i}$ are the air and surface temperature respectively (K) and α the heat transfer coefficient ($Wm^{-2}K^{-1}$).

The concept of the heat transfer coefficient can be applied to the mechnism of radiation also. Then Eq. 1c can be compared to Eq. 8 to read a heat transfer coefficient for radiative exchange. For ventilation, the heat flux is calculated according to Eq. 3 while Eq. 8 stands for the heat flux density. So if ventilation is compared to or combined with the other mechanisms, the α's have to be combined with an exchange area A.

For both the natural and forced convection, α is dependent on fluid properties and system parameters for a particular geometry. These dependencies are expressed in relations between dimensionless numbers. For a detailed description we refer to handbooks for heat and mass transfer e.g. Bird, Stewart and Lightfoot (1960). The coefficient of heat transfer is combined in the dimensionless Nusselt number (Nu). For natural convection the Grashof number is of interest and for forced convection the Reynolds number. This leads for natural convection to relations between the Nusselt and Grashof number and for forced convection between the Nusselt and Reynolds number. In both relations a Prandtl number accounts for some fluid properties. In the greenhouse situation the fluid is air and fluid properties only vary due to temperature dependencies of these properties. These variations can be neglected in a first approach. The main variable factors for a particular geometry then are for natural convection the temperature difference $T_1 - T_2$ as expressed in the Grashof number (Gr) and for forced convection the air velocity u as expressed in the Reynolds number (Re):

$$Gr = g\ (T_1 - T_2)\ T^{-1}\ L^3\ v^{-2} \qquad (9a)$$

and

$$Re = u\ L\ v^{-1} \qquad (9b)$$

where g is the acceleration due to gravity (ms^{-2}), T the mean temperature (K), L the characteristic length of the geometry (m) and v the kinimatic viscosity (m^2s^{-1}). The ratio between Gr and Re^2 indicates if the exchange is due to pure natural or pure forced convection (Morgan, 1975).

The cover of a multispan greenhouse complex has a saw tooth surface geometry. A common Dutch greenhouse type, manufactured in an industrial way, is a so called Venlo-type greenhouse. It has standardised measures. The cover has a span width of 3.2 m and the roof slope can vary between 22 and 26 degrees, so the geometry of the saw tooth surface is known with a characteristic length of about 1.75 m (ridge-gutter distance). The exchange with the cover at both the inside and outside are of interest. At the inside, local air velocities are low ($\approx$ 10 cms^{-1}: Re $\approx 10^4$) and the temperature differences considerable ($\approx$ 10 K: Gr $\approx 10^{10}$), so it can be expected that the heat transfer is due to natural convection. At the outside, forced convection can be expected.

In literature no data on convective heat and mass transfer to and from saw-tooth surfaces could be found. It is not known a-priori if literature data on the transfer to and from flat plates can be applied in this case while the flow field over the surface will be transformed by the saw-tooth surface. Therefore from expriments the convective heat transfer to and from the cover was measured and the flow field over the cover was sampled. This yielded natural convection relations for the heat transfer at the inside and also at the outside for low wind speeds up to about 3 m/s. At higher wind speeds forced convection was found at the outside.

A special remark has to be made on the condensation to the inner surface of the cover. The mechanism of mass transfer is identical to that of heat transfer and the same theory can be applied. This leads to a mass transfer coefficient k (ms^{-1} between mass flux density m ($kgm^{-2}s^{-1}$) and the difference in concentrations of the water vapour (kgm^{-3}) at the surface $e_{g,i}$ and in the free air e_i analogous the Rel.8 for heat transfer:

$$m_{g,i} = k_{g,i} (e_{g,i} - e_i) \qquad (10)$$

The water vapour concentration at the surface is given as the saturated concentration at surface temperature. From the analogy between heat and mass transfer the heat and mass tranfer coefficients α and k respectively are related, so knowing α results in k:

$$k = \alpha \, (Cap_{air})^{-1} \, Le^{2/3} \qquad (11)$$

where Le is the dimensionles Lewis number accounting for the ratio between thermal diffusivity and kinematic viscosity. For air this ratio is near to one.

III.6 Exchange between the heating system and the air

Most commonly a greenhouse is heated by a heating pipe system, distributing hot water from a central boiler. Various pipe arragements are in use, the most common arrangement is that of four 2" pipes per span at a small distance (5–10 cm) to the soil surface. The four pipes are combined in two pairs, with a distance in the pair of about 20 cm and the pairs at a central distance of 1.6 m. In this way the heating pipes can be used in practice as a rail system for the internal transport. The mechanism of heat transfer between the pipes and the air is by convection. The same approach as indicated in section III.5 can be applied here. The characteristic length is the diameter of the pipes.

With local air velocities of 10 cm/s, Re is low (≈250), but

Gr will be $\approx 10^6$ so the value of $GrRe^{-2}$ is ≈ 10 which is of the same order of magnitude as the criterion for pure natural convection around horizontal cylindrs $GrRe^{-2} \geq 14$ (Morgan, 1975). However, literature data on the pure natural convective exchange from horizontal cylinders cannot be used for the convective exchange from the horizontal heating pipes while the above given arrangement differs from the experimental conditions from literature on pure natural convection.

Experiments under greenhouse conditions yielded the heat transfer coefficient, indeed being due to natural convection and differing from literature on pure natural convection under ideal conditions. With the heat transfer between pipes and air described in the same way as Eq. 8 the coefficient of heat transfer being dependent on the temperature difference as given by its dependency on the Grashof number, again introduces non-linearities.

III.7 Exchange with and transport in the soil

In the energy budget of the greenhouse, the exchange with the soil is of minor importance. However, the soil surface exchanges thermal radiation with the other greenhouse components and the energy storage in the soil determines the dynamics of the greenhouse system on a daily base (Bot, 1989b). So the exchange to and the transport in the soil have to be represented in a proper way for the description of the greenhouse climate especially for the daily rhythm. The calculation via the natural convection exchange at the surface and the mechanism of conduction for the transport in the soil has proved to be of sufficient accuracy. The convective exchange is treated already in the sections III.5 and 6. In the mechanism of conduction the driving force for heat transfer is the temperature gradient in the medium. For conduction in the vertical direction z only:

$$Q_{cond} = -l\ (dT/dz) \qquad (12)$$

where Q_{cond} is the heat flux density due to conduction (Wm^{-2}) and l the coefficient of thermal conductivity ($Wm^{-1}K^{-1}$). To be able to calculate the conductive fluxes the thermal conductivity l of the soil has to be known. This will be dependent on the character of the soil and the water content. If the crop is grown in the soil this will be well watered so the water content will not vary much. In-situ measurement is quite cumbersome, data on the thermal conductivity of saturated soils can be found in handbooks on soil physics e.g. Baver et al. (1972). With the crop growing on substrate the soil is covered and is well watered too, due to leakage of nutrient water. So here the same applies. For closed, recirculating watering systems, as are introduced recently to prevent emission of nutrients to the soil and ground water, the soil will be dry. Then l has to be determined for this situation.

IV. DYNAMIC MODEL

With the quantitative description of the main exchange processes, the energy and mass balances can be set up over representative, homogeneous parts of the greenhouse, like those discussed in section III.1. In the first version of the model the parts were the single-layer greenhouse cover, the greenhouse air, the greenhouse crop and a 7-layer greenhouse soil. The successive soil layer thicknesses doubled, and varied from one cm. for the surface layer to 64 cm for the seventh layer. The mass balance for water vapour has to be set up over the air compartment only. In later versions a double cover was introduced represented by a two layer cover. Also a thermal screen was implemented which implicated that two air compartments had to be considered, one between cover and screen and one below the screen. However the general approach was the same as presented here.

The energy balance for the compartment j with temperature T_j, exchanging energy to n neighbouring compartments with temperature T_n, can be represented in general as:

$$V_j \, Cap_j \, dT_j/dt = \Sigma(Q_{j,n} \, A_{j,n}) - E_j + S_j \quad (13)$$

where V_j is the volume of the compartment, $A_{j,n}$ the exchange area between the compartment and the successive neighbouring compartment, S_j the total absorbed solar radiation in compartment j and E_j the energy needed for transpiration in compartment j. The energy fluxes $Q_{j,n}$ due to the various mechanisms can be related to the temperatures mentioned as indicated in rel. 1, 3, 8 and 12.

This leads to the general expression:

$$V_j\ Cap_j\ dT_j/dt = \Sigma((\alpha_{j,n}\ A_{j,n})(T_j - T_n)) - E_j + S_j \qquad (14)$$

Here $\alpha_{j,n}$ is given by the mechanism and will in general contain a non-linearity. From fig. 1a and b the various fluxes to and from each compartment to the neighbouring compartments can be read. In this notation the boundary conditions are considered to be the temperatures of environmental compartments. So the energy balances for all compartments will result in a set of first order differential equations.

The mass balance for the air compartment with concentration c_j can be set up analoguously:

$$V_j\ (dc_j/dt) = \Sigma\ (m_{j,n}\ A_{j,n}) \qquad (15)$$

Here the vapour fluxes due to ventilation, evaporation and condensation have to be considered.
With rel. 10 for the mass flux density, rel. 15 can be rewritten as:

$$V_j\ (dc_j/dt) = \Sigma\ (k_{j,n}\ A_{j,n})(e_j - e_n) \qquad (16)$$

Here $k_{j,n}$ is related to $\alpha_{j,n}$ according to rel. 11.

The evaporative and condensative fluxes $m_{j,n}$ link the mass and energy balances according rel. 7. Moreover the crop temperature and cover temperature determine the saturated vapour concentrations at the respective surfaces so form a second link between the energy and mass balances. So both flow and state variables in the energy and water vapour model are linked.

The response of the greenhouse system to the time varying environmental conditons is then represented by the solution of the set of differential equations. This solution can be found by e.g. forward numerical integration. Algorithms to do so are available in various languages for dynamic simulation varying from simple Euler to more complex Rung-Kutta or other calculating schemes. In our case we applied CSMP (IBM,1966) and TUTSIM (Kraan, 1974) that can be combined with graphical representation of the equations in Bond Graphs. This allows a clear overview of the various state and flow variables in the system together with their relations and interactions. TUTSIM was specially designed to translate the Bond graph representation into a simulation computer programme. For parts of the model this allows an easy analysis and understanding. However for the complete model the non-linearities introduce cumbersome procedures, so finally the complete model is programmed in CSMP. Due to the fact that IBM stopped supporting CSMP an alternative had to be found. Presently ACSL is a promising language for dynamic simulation.

The simulation results are realistic under various outdoor conditions (Bot, 1989a) both in the short and long term response. This is demonstrated in Figure 4 and 5 for a 24 hour period from noon to noon with strong fluctuating shortwave radiation during the first day and a bright second day. Not only the state variables of the system are calculated, like temperatures and concentrations, but of course also the flow variables, like energy and mass flows.

Houter (1990) applied the modeling approach and compared the results with measurements of energy consumption at a large number of growers sites. Also here the results were realistic with an unaccuracy in general within 10%.

The model is succesfully applied and adapted in various studies as reported already at the end of section I.

In control, like the optimization study of Van Henten and Bontsema (1991) it is used to get insight in the complex interactions in the greenhouse system and simplifications are made to adapt it to the application. In simple control, implementation of the behaviour of the mixing valve of the heating system with the characteristics of the heating system itself enables the incorporation of control algorithms for heating. The same holds for the implementation of the characteristics of the motors for opening the ventilation windows which allows the incorporation of control algorithms for ventilation. Then the heating pipe temperature as boundary condition and the measured window opening as a system property can be replaced by a set point for e.g. air temperature and humidity or a set point calculating scheme. In this way the heating and ventilation control is included in the system and control algorithms can be checked with the advantage over simple control models that the response of the various parts of the greenhouse system, e.g. the crop, can be calculated. In this way the model can be applied as a design tool.

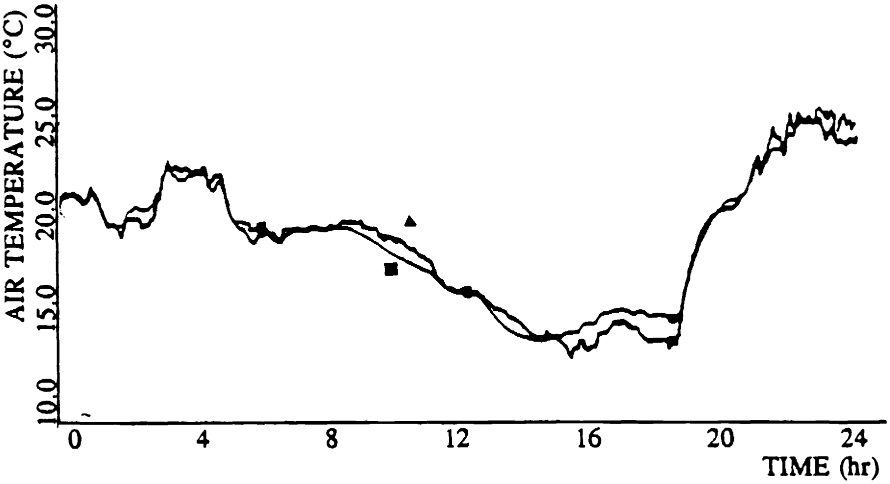


Figure 4. Simulated (■) and measured (▲) greenhouse air temperature during a 24 hr period with varying radiative conditions.

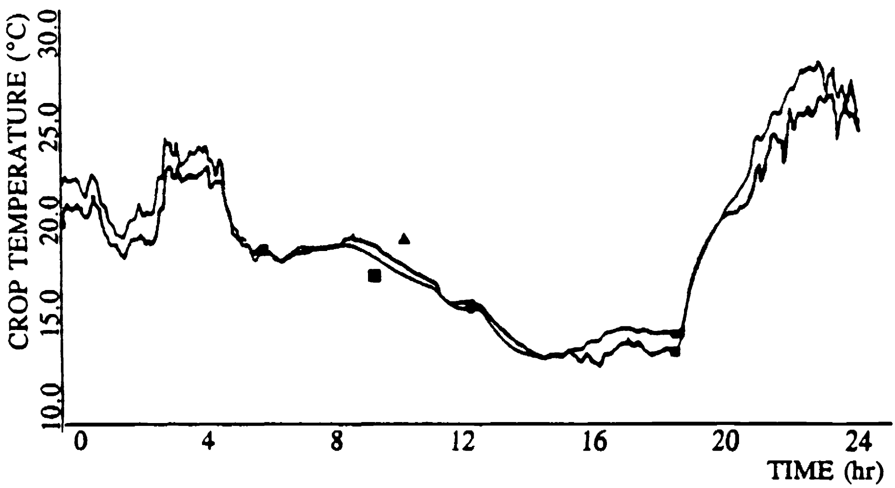


Figure 5. Simulated (■) and measured (▲) crop temperature during a 24 hr. period (same period as Fig. 4).

V. CONCLUSIONS

The greenhouse climate is a complex interaction of various physical processes. Quantification of this climate as affected by the dynamic outdoor weather conditions and the physical properties of the greenhouse itself demands quantification of the various exchange processes. Then combination in a physical model yields a reliable description of the greenhouse climate. This can be applied in practical and theoretical studies on the optimization of greenhouse production and as a design tool in greenhouse studies. For control purposes only mostly the model can be simplified and restricted to relations of interest. Then the extensive model as a representation of the main characteristics of the real system can be used to validate the simplified one. This can prevent cumbersome experimental validation.

REFERENCES

Baver, L.D., Gardner, W.N. and Gardner, W.R. (1972). Soil Physics. Wiley, N.Y. 489 pp.

Bird, R.B., Stewart W.E. and Lightfoot E.N. (1960). Transport Phenomena. J. Wiley and Sons, N.Y. 780 pp.

Bot, G.P.A. (1983). Greenhouse climate: from physical processes to a dynamic model. Ph.D. thesis, Agricicultural University Wageningen. 240 pp.

Bot, G.P.A. (1989a). A validated physical model of greenhouse climate. *Acta Horticulturae* 245: 389-396

Bot, G.P.A. (1989b). Greenhouse simulation models. *Acta Horti culturae* 245: 315-325

Bot, G.P.A. and Challa, H. (1991). Crop growth and climate control of greenhouses. In Y.P.S. Bajaj (Ed.), *Biotechnology in Agriculture and Forestry* Vol. 17, High-Tech and Micropropagation I: 516-539. Springer Verlag

Challa, H., Bot, G.P.A., Nederhoff E.M. and Van de Braak N.J. (1988). Greenhouse climate control in the nineties. *Acta Horticulturae* 230: 459-470

Critten, D.L. (1992) Greenhouse transmission models. Proc.Int. Workshop on greenhouse crop models. Avignon, August

1991. *Acta Horticulturae.* (in press)

De Jong, T. (1989). Natural ventilation of large multispan green houses. Ph.D. thesis, Agricultural University Wageningen. 116 pp.

De Zwart, F. (1992). Physical modelling of the greenhouse climate. *IMAG-DLO Report*, IMAG-DLO, Wageningen (in press) Gijzen, H. and Ten Cate J.A. (1988). Prediction of the response of greenhouse crop photosynthesis to environmental factors by integration of physical and biochemical models. *Acta Horticulturae* 229: 251-258

Gijzen, H., Vegter J.G. and Nederhoff E.M. (1990). Simulation of greenhouse crop photosynthesis: validation with cucumber, sweet pepper and tomato. *Acta Horticulturae* 268: 71-80

Houter, G., Gijzen H., Nederhoff E.M. and Vermeulen P.C.M. (1989). Simulation of CO_2 consumption in greenhouses. *Acta Horticulturae* 248: 315-320

Houter, G. (1990). Simulation of CO_2 consumption, heat demand and crop production of greenhouse tomato at different CO_2 strategies. *Acta Horticulturae* 268: 157-164

IBM (1966). IBM 1130 Continuous System Modelling Programme. *IBM manuals* SH 20-0905 and H 20-0209

Kondratyev, K.Ya. (1972). Radiation processes in the atmosphere. *WMO report 309*, WMO Geneva. 214 pp.

Kraan, R.A. (1974) THT SIM: A conversational simulation programme on a small digital computer. *Journal A* 15: 186-190.

Morgan, V.T. (1975). The overall convective heat transfer from smooth circular cylinders. *Advances in heat transfer* 11: 199-264. Academic Press, N.Y.

Stanghellini, C. (1987). Transpiration of greenhouse crops, an aid to climate management. Ph.D. Thesis Agricultural University Wageningen. 150 pp.

Udink ten Cate, A.J., Bot G.P.A. and Van Dixhoorn J. (1978). Computer control of greenhouse climates. *Acta Horticulturae* 87: 265-272

Van den Muijzenberg, E.W.B. (1980). A history of greenhouses. IMAG Wageningen. 435 pp.

Van Henten, E.J. and Bontsema J. (1991). Optimal control of greenhouse climate. In Y. Hasimoto and W. Day (Eds.), Mathematical and Control Applications in Agriculture and

Horticulture. IFAC Workshop Series, 1991. number 1: 27-32. Pergamon Press

Wartena, L., Palland C.L. and Vossen G.H.L., (1973). Checking of some formulae for the calculation of long wave radiation from clear skies. *Arch. Met. Geoph. Biokl. B* 21: 335-348

Chapter 3

MEASUREMENT AND PARAMETER IDENTIFICATION PROBLEMS FOR CONTROL OF BIOTECHNICAL PROCESSES

Axel Munack

Institute of Biosystems Engineering
Federal Agricultural Research Centre
Braunschweig, Germany

I. INTRODUCTION

Also in the classical biotechnological processes of the food processing industry - like production of beer, wine, baker's yeast etc. - measuring and control techniques play a major role. In the past, these tasks were taken over by man almost entirely, e.g. by taking manual samples, by performing various sequential process steps, by suitable feeding of nutrients to the process or by stopping the process at the 'best' time instant. The rapidly increasing number and amount of biotechnologically produced products, the rising demands for increased product quality - also with regard to legal requirements (GMP) - as well as the permanent pressure to increase the process productivity have led to a continuously growing application of automation and control equipment for biotechnical processes. This trend is supported by the availability of inexpensive electronic components which enable to apply sophisticated algorithms of data processing and control, and facilitate the application of complex on-line procedures to medium-scale plants which in former times seemed to be economically feasible only for large production plants.

The situation in agricultural applications is quite similar. Here, however, the demand for cheap equipment is even stronger, since production units are small and most of the products have to be sold at low prices. The need for protection of the environment and the corresponding new legal restrictions demand for agricultural production processes with lowest possible inputs of fertilizers and agrochemicals. Having in mind, that - at least seen world-wide - the agricultural production must be maintained at the actual output quantity and should even be increased in proportion to the increasing world population, high productivity and low input can be obtained only by sophisticated production methods, using advanced control equipment. Thus, new technologies will find a high level of application in agriculture in the near future.

In this contribution, the emphasis is laid on applications to biotechnical processes. Various parallels will be evident with respect to agricultural applications. At first, the major measurement techniques for bioreactors are presented and discussed. In the more control-oriented part of the paper the use of dynamic process models and the specific aspect of the information content of measurement data for model generation will be considered. It is not the purpose of this paper, to give a complete overview on measurement and control of biotechnical processes (which can be found in textbooks, cf. e.g. Rehm and Reed (1991)); instead, some techniques will be addressed which are most likely to be relevant for control of agricultural processes, too. As application, a distributed-parameter system is chosen, since agricultural production processes also show transport and diffusion/convection phenomena in their distributed compartments soil and air.

II. MEASUREMENT TECHNIQUES FOR BIOTECHNICAL SYSTEMS

A peculiarity of measurements at biotechnical systems lies in the fact, that data from different sources reach the data-logger at different times. One can divide into three kinds of data:

- quasi continuous measurements, which are sampled at a high frequency and without any lag time, like pressure, temperature, pH, optical density, etc.;
- discontinuous measurements, which are sampled with a small lag time, like off-gas analyses and auto-analyzers for measurements of the broth concentrations;
- discontinuous measurements with long lag time, like manually processed analyses for the composition of biomass, and for concentration of substrate and product, in particular for cell-

internal concentrations.

For process automation, one will rely on data of the first two groups, while data of the last group are mainly used for documentation and further calculations after the fermentation is finished. The use of these data during the actual cultivation is under research in some theoretically working groups. In any case, an effective data management is necessary in order to process the different data according to their temporal origin.

DIRECT MEASUREMENTS

In the following, only the first two groups are addressed; Table 1 gives an overview on the various possibilities for measurements at biotechnical processes. The list is by no means exhaustive, since the development of new sensors and analyzers shows a significant progress. However, a critical review exhibits, that many of the important measurement probes for automation of biotechnical processes have not reached the state of reliability and simplicity which is needed for industrial or at least pilot scale applications. Many of the recently developed sensors still show various cross-sensitivities, may only be used in restricted concentration ranges, are subject to disturbances and drift, or are not applicable in situ due to their physical construction. In the moment, automated analytical methods have attained a great deal of interest. Taking probes from the process and carrying out analyses outside the reactor, e.g. by FIA, GC, HPLC, or MS, enables to calibrate the measurement device as often as necessary; the cost of these devices, however, is still too high for a wide-spread application, and the devices need a highly qualified staff in order to be operationable over long time periods.

MODEL-BASED MEASUREMENTS

The very restricted opportunities for direct measurements at biotechnical processes lead to the question whether the very efficient methods of control theory, like observers or filters, may also be applied to fermentation processes. Somewhat problematical in this respect is the fact, that in many cases reliable models for bioprocesses do not exist, since the dynamical equations are nonlinear, and the interdependencies of various major metabolic reactions inside the organisms are not yet fully understood. Nevertheless, efficient and general modelling techniques for

Table 1 Measured variables and measuring principles for biotechnical process

Measured variable	measuring principle
volume	differential pressure ultrasonic devices
foam	contact electrode
stirrer speed	counter
temperature	PT100, thermocouple
gas flow	volumetric flow sensors
pH, redox potential	glass electrodes ion-sensitive FETs
biomass	filtration probe optical density fluorescence specific impedance
dissolved O_2	polarographic probe PTFE-probe + analyzer
dissolved CO_2	PTFE-probe + analyzer
off-gas O_2	paramagnetic detector ZiO_2-sensor mass spectrometer
off-gas CO_2	infrared detector mass spectrometer
ethanol	PTFE-probe + analyzer
glucose	polarimeter enzyme electrode enzyme thermistor
L-lactate	enzyme thermistor
urea	optical biosensor
NADH in cells	fluorescence
morphological characterization	image processing
various substances	FIA, MS, HPLC, CFA

bioprocesses are a very promising research subject for the future, since the formulation of a process model also enables to use sophisticated (predictive) control strategies. Thus, there is a need for off-line structural modelling as well as for on-line and off-line parameter identification.

The first industrial cultivation process driven by an observer was the Pekilo protein process, cf. Halme and Tiussa (1983). Further observer principles for bioprocesses were proposed by Stephanopoulos and San (1984) and by Dochain (1986). The latter author avoids to specify a growth kinetic for the process. This is a well-suited means for process observation, however it does not enable to calculate long-term predictions of the process.

III. MODEL BASED PREDICTIVE CONTROL

A control structure which seems to be well suited for biotechnical processes is the so-called Open-Loop Feedback Optimal Controller (OLFO). This algorithm, cf. Luttmann et al. (1985), relies heavily on a good (structurally correct) process model, and is very similar to the recently established generalized predictive controller types, cf. Clarke et al. (1987). The structure is shown in Fig. 1. The algorithm solves the combined problem of system identification and process optimization separately. As in most adaptive algorithms, the time is divided into time slices, the adaptation intervals. In each interval, the unknown parameters of the system and - maybe - some deterministic disturbances are determined by minimizing an output least squares error criterion which weights the differences of measured system outputs and model outputs of the most recent adaptation interval(s). These identified parameters are extrapolated into the future and serve as a basis for calculation of future control actions, which control the process in an optimal way. Various criteria for optimization are possible. Then the control inputs are set to the calculated optimal ones, but only for the forthcoming adaptation interval. During this interval, the whole identification and optimization computations are carried out again, which may lead to a corrected control input for the following interval, and so on.

It can be stated, that this strategy should be well suited for processes with unknown or slowly time-varying parameters, if a structurally correct model is known and if the on-line estimation of parameters is possible by the available measurements. Therefore the evaluation of the information content of measurements plays a key role in this concept; a problem which is considered in detail in the following paragraphs.

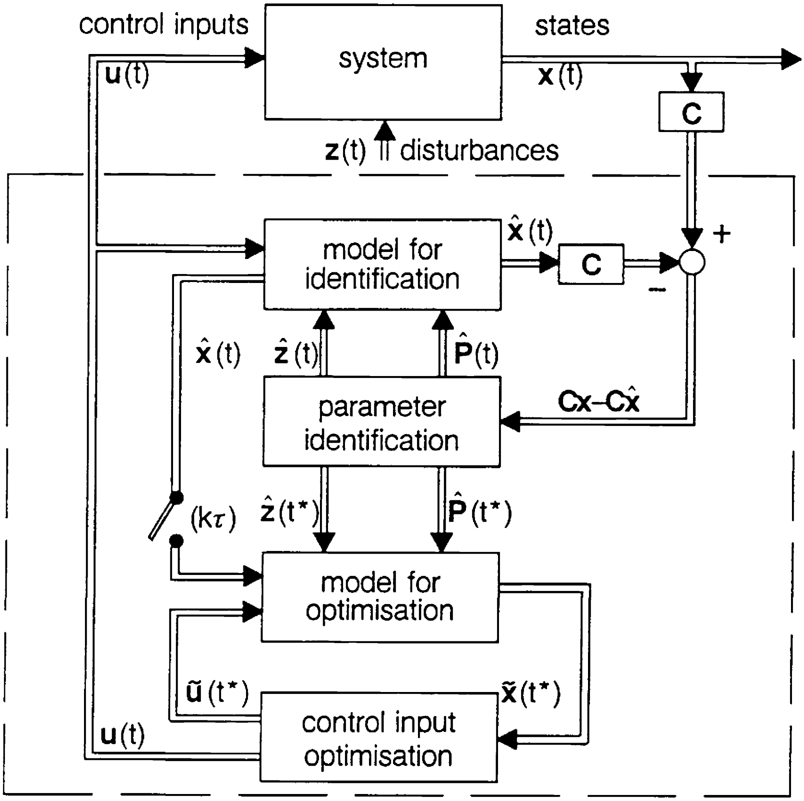


Fig. 1 Structure of the OLFO controller

IV. INFORMATION CONTENT OF MEASUREMENTS

The problem of the information content is directly related to the precision of parameter estimates which may be obtained by using a certain measurement data set. This data set is dependent on (i) the experimental conditions under which the data were obtained, e.g. the process input functions, and (ii) the measurement devices, their precision and – for distributed-parameter systems – their location. The general theoretical framework which allows to deal with these problems is reported here shortly; for a more detailed discussion see Goodwin (1987) and the literature referenced therein.

Let the system be described by the nonlinear differential equation

$$\mathring{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x},t,\mathbf{u},\mathbf{P}) \;;\quad \mathbf{x}(0) = \mathbf{x}_0(\mathbf{P}) \,, \qquad (1a)$$

and the system outputs be given by

$$\mathbf{y}(t_i) = \mathbf{g}(\mathbf{x}(t_i,\mathbf{u},\mathbf{P}),t_i,\mathbf{P}) \,, \qquad (1b)$$

where $\mathbf{x} \in R^n$ denotes the state, $\mathbf{u} \in R^m$ denotes the input, $\mathbf{y} \in R^q$ is the (discrete time) input, and $\mathbf{P} \in R^p$ is the vector of unknown system parameters. As pointed out above, we will treat the situation that a structural correct model is available, which allows for a formulation of the model equations identical to Eq. (1a,b), where only **P**, **x**, and **y** are replaced by the model parameters, states, and outputs $\hat{\mathbf{P}}$, $\hat{\mathbf{x}}$, and $\hat{\mathbf{y}}$, respectively.

With the difference between model and process outputs,

$$\mathbf{e}_i = \mathbf{e}(t_i,\mathbf{u},\hat{\mathbf{P}}) = \hat{\mathbf{y}}(t_i,\mathbf{u},\hat{\mathbf{P}}) - \mathbf{y}(t_i) \,, \qquad (2)$$

the quadratic identification functional

$$J_I(\mathbf{u},\hat{\mathbf{P}}) = \sum_{i=1}^{N} \mathbf{e}_i^T \mathbf{Q} \mathbf{e}_i \;;\quad \mathbf{Q}_i \geq 0 \qquad (3)$$

is obtained. The optimization problem to be solved for parameter identification is now posed as follows:
"Find $\hat{\mathbf{P}}_{opt}$ such that $J_I(\mathbf{u},\hat{\mathbf{P}}_{opt}) \leq J_I(\mathbf{u},\hat{\mathbf{P}}) \; \forall \; \hat{\mathbf{P}} \quad P_{ad}$, and $\hat{\mathbf{P}}_{opt} \in P_{ad}$, where $\hat{\mathbf{P}}_{ad}$ is the set of physically, chemically, or biologically meaningful parameters."

Next we must consider that measurements on a system usually do not represent the undisturbed system's outputs, but will be corrupted by noise. For biotechnical processes, this noise is partly caused by the impreciseness of the measurement device, but most often a larger part is caused by the process itself with imperfect mixing, turbulent flow, gas bubbles, and other inhomogeneities in the reactor. This fact should be kept in mind when respresenting all noise characteristics by a zero-mean, Gaussian white noise process as follows:

Table 2 Criteria to evaluate and optimize the information content of measurements

criterion	definition	interpretation
A-criterion	min (trace **S**) max (trace **F**)	minimization of the variance in the arithmetical mean
D-criterion	min (det **S**) max (det **F**)	minimization of the variance in the geometrical mean
G-criterion E-criterion	min (λ_{max}(**S**)) max (λ_{min}(**F**))	minimization of the largest variance of any parameter combination

$$\mathbf{y}^M(t_i) = \mathbf{y}(t_i) + \mathbf{w}(t_i) \; ; \tag{4}$$

$$E\{\mathbf{w}(t_i)\} = \mathbf{0} \quad , \; i = 1, \ldots, N \; ; \tag{5a}$$

$$E\{\mathbf{w}(t_i)\mathbf{w}^T(t_j)\} = \delta_{ij} \; \mathbf{C}(t_i) \quad , \; i,j = 1, \ldots, N \; . \tag{5b}$$

Due to this noise, the result of the optimization problem, $\hat{P}_{opt}$, becomes a random variable. If a bias-free estimate can be obtained (which will not be possible in most nonlinear situations), one could state that, in the limit, the expectation for $\hat{\mathbf{P}}_{opt}$ could be equal to the true process parameters,

$$E\{\hat{\mathbf{P}}_{opt}\} = \mathbf{P} \; . \tag{6}$$

Futhermore, it can be shown by using the Cramér-Rao inequality, that

$$E\{\hat{\mathbf{P}}_{opt} - \mathbf{P})(\hat{\mathbf{P}}_{opt} - \mathbf{P})^T\} \quad \mathbf{S} = \mathbf{F}^{-1} \; , \tag{7}$$

where **F** is the Fisher information matrix. Thus **S** gives a lower bound on the obtainable covariance of the parameter estimates.

The information matrix is computed by

$$F = \sum_{i=1}^{N} [Y_P^T(t_i,u,P)\, C^{-1}(t_i)\, Y_P(t_i,u,P)] \,, \tag{8}$$

which shows that the optimal weighting of the errors in Eq. (3) is given by $Q_i = C^{-1}(t_i)$. Several criteria were proposed to give a measure for evaluation of the precision which is obtained; they may be applied to the information matrix **F** or to the covariance matrix **S**. An overview is compiled in Table 2. The most simple criterion trace(**F**) → max should not be applied since it may lead to non-informative experiments.

These well-known results are the basis for the application to a biotechnical process in the following paragraph.

V. APPLICATION TO A BIOTECHNICAL PROCESS IN A TOWER REACTOR

A simplified model of a biomass cultivation process is used to demonstrate the various aspects of utilization of the above formulated results. The detailed description of the plant and a derivation of a complete model can be found in the paper by Luttmann et al. (1985); calculations with a reduced model were already reported, cf. Munack (1985). The model equations used here are even more simplified; however, the main results remain the same. This shows, that the results are not heavily model-depending, an apprehension which is often expressed when treating nonlinear systems with the above (linearized) theory. The emphasis in this contribution is laid on a comparison of different criteria and the time-dependent availability of parameter estimates in different phases of the fermentation.

DESCRIPTION OF THE PLANT

The reactor used is a tower reactor, which is gassed from the bottom - a bubble column, cf. Fig. 2. It is filled with liquid, which contains the substrate and all other nutrients. The cells are submerged in the liquid phase; no further mechanical agitation is needed, since the rising gas bubbles give a nice mixing. However, at higher cell concentrations, when the batch process reaches its end, an oxygen limitation occurs at the top of the reactor, since all available oxygen is consumed in the lower part of the reactor. Thus, distinct oxygen profiles can be observed, which lead to the necessity to describe the process by a distributed-parameter model.

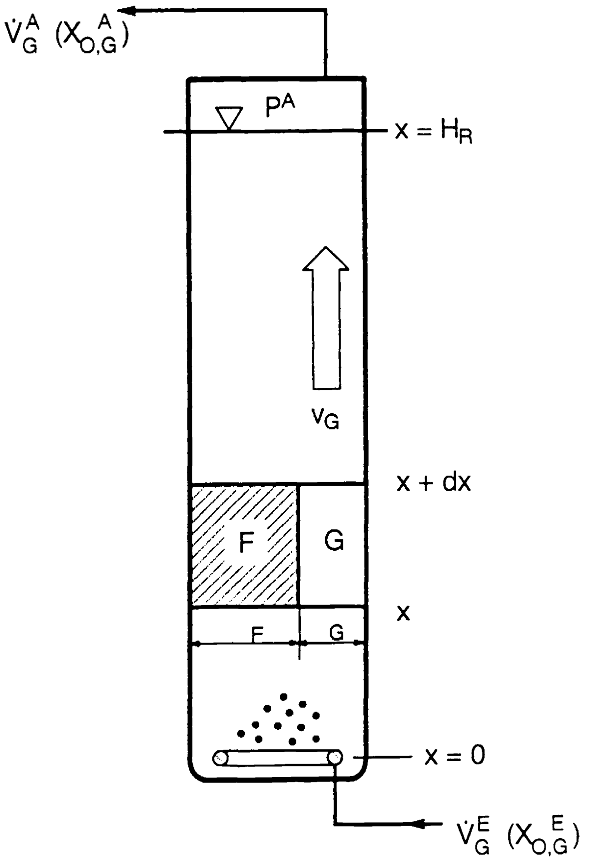


Fig. 2 Schematic diagram of the tower reactor

The simple model used in the calculations is shown in Table 3. It consists of three balances: the first equation describes the dissolved oxygen in the liquid phase of the reactor, the second gives the oxygen molar fraction of the gaseous phase (bubbles), and the third equation gives the biomass concentration.

Measurements can be taken of the dissolved oxygen concentration by sensors which may be installed at any spatial position in the column. Furthermore, the biomass concentration is measured; here the position is not critical, as the biomass is almost well mixed in the reactor. In addition, an off-gas analysis gives the molar fraction of the oxygen contained in the exhaust gas, which

Table 3 Model equations for the tower reactor

Dissolved oxygen in the liquid phase

$$\frac{\partial O_F}{\partial t} = D_F \frac{\partial^2 O_F}{\partial x^2} + k_L a(x) \cdot (\frac{p}{H} X_{O,G} - O_F) - r_{OX}$$

$$\left.\frac{\partial O_F}{\partial x}\right|_{x=0} = 0; \quad \left.\frac{\partial O_F}{\partial x}\right|_{x=H_R} = 0; \quad O_F(t=0) = \frac{p \cdot X_{O,G}{}^E}{H}$$

Oxygen molar fraction in the gas phase

$$\frac{\partial X_{O,G}}{\partial t} = -v_G \frac{\partial X_{O,G}}{\partial x} - \frac{\epsilon_F}{\epsilon_G} k_L a(x) \cdot \frac{RT}{p} (\frac{p}{H} X_{O,G} - O_F)$$

$$X_{O,G}\ (0) = 0.21\ ; \quad X_{O,G}\ (t = 0) \approx 0.21$$

Biomass concentration

$$\frac{\partial X}{\partial t} = D_F \frac{\partial^2 X}{\partial x^2} + r_{OX} Y_{XO} - r_D$$

$$\left.\frac{\partial X}{\partial x}\right|_{x=0} = 0; \quad \left.\frac{\partial X}{\partial x}\right|_{x=H_R} = 0;$$

where $k_L a(x) = k_L a^E - \Delta k_L a \cdot \frac{x}{H_R}$

$$r_{OX} = \frac{\mu_m \cdot O_F}{Y_{XO}\ (K_O + O_F)}\ X\ ; \quad r_D = \mu_D \cdot X$$

may be used to calculate the integral oxygen uptake rate of the microorganisms.

Five parameters have turned out to be unknown. They are time-variant and varying from cultivation to cultivation, too. These are two fluiddynamical parameters, k_La^E and Δk_La, describing the oxygen transfer from gas phase into liquid phase and - roughly - its decrease towards the top of the reactor, as well as three biological parameters, μ_m, Y_{xo}, and K_0, the maximum specific growth rate, the yield coefficient, and the limitation constant in the Michaelis-Menten kinetics.

STATEMENT OF THE PROBLEMS

With respect to the instrumentation of the process, several questions must be solved. These are:

- Which are the best positions for the dissolved oxygen (DO) sensors to be placed in the reactor column?
- How many DO sensors are essential for a sufficient precision of the parameter estimates?
- Which types of the measurements - DO, cell mass, off-gas - are essential or give the highest contribution to the parameter estimates?
- Are parameter estimates possible in all states of the cultivation, which means: Can the OLFO controller identify all parameters by using only measurement data from the preceding adaptation interval?
- Which are the parameters that are worst identified, and are these parameters of great influence on the course of the process (which would result in unreliable process predictions).

RESULTS CONCERNING SENSOR TYPE AND ALLOCATION

To answer these questions, the nominal trajectories of the process were computed as well as the associated sensitivities with respect to parameter changes. Then the information matrices were computed for various sensor configurations.

The results are summarized in Fig. 3 to Fig. 7.

Fig. 3 shows the evaluation of the E-criterion for the information matrix. When only one DO sensor is installed (1), it should be placed near the bottom of the fermenter, since the information content of the signal then is at its highest value. Adding a cell concentration measurement (1+X) gives a result for $\lambda_{min}(F)$ which is more than one decade higher, if the DO sensor is placed near the bottom. Note that this means that the standard deviation of the worst identified parameter combination is reduced by a factor of approx. 3.5. A further decade is gained by taking one DO sensor and the off-gas analysis (1+O), however, now the best position for the DO sensor is on the top of the fermenter. Best identification, of course, is possible by taking all three measurements (1+X+O), where the best position for the DO sensor is again on the top of the reactor. Fixing the first DO sensor on this top position, now a second DO probe is placed. This results in the curve (2+X+O), indicating that the best position of this sensor is at the bottom of the reactor. A third DO sensor (this curve is not shown) wouldn't add very much to the information content of the measurements.

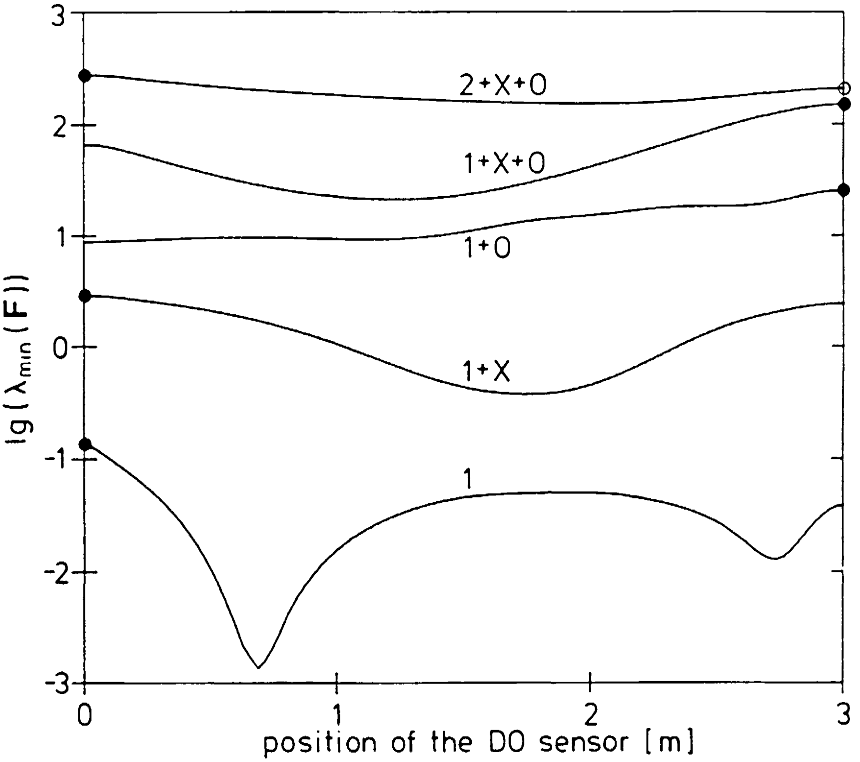


Fig. 3 Results of the sensor allocation problem ($\lambda_{min}(F)$)

One could object that the calculation for the two DO sensor placement problem is not complete, because it starts with the assumption that one sensor is fixed at the top of the reactor and then only the second position is sought. Therefore, the complete problem of allocating two sensors (in combination with cell concentration and off-gas analysis) was treated. The contour plot of the resulting $\lambda_{min}(F)$ functional depending on the two sensor positions is shown in Fig. 4. It proves that in fact an allocation at the bottom and at the top gives best results. Relative maxima are obtained by allocating both sensors at the bottom or both sensors at the top of the reactor, while an allocation of both at 1.25 m of height would be the worst choice. It must be pointed out, that for cases where two sensors are mounted on the same height, there should be a sufficient (radial) distance between them in order to ensure the assumed statistical independence of the noise signals.

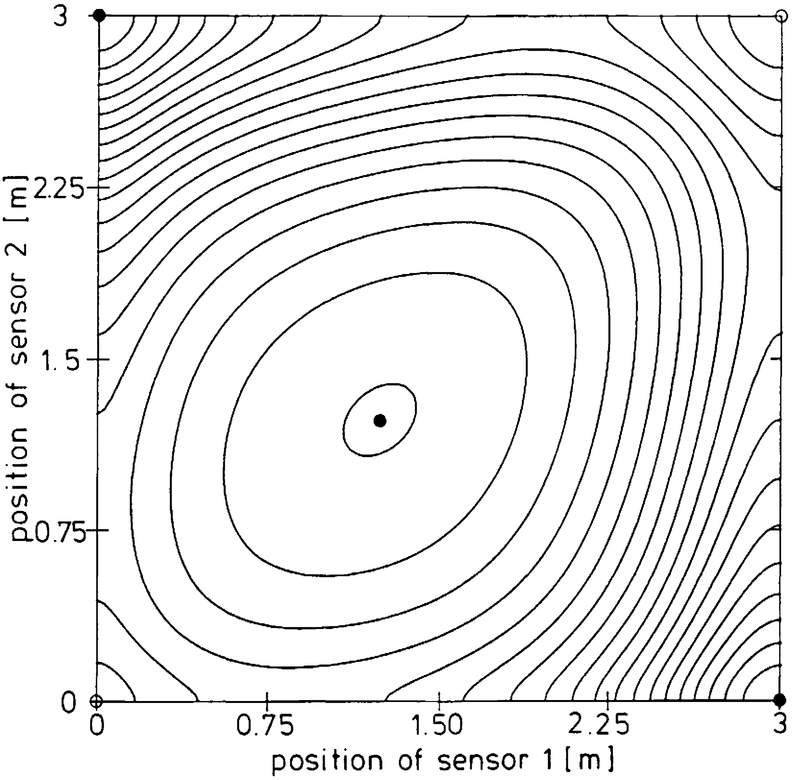


Fig. 4 Contour plot of $\lambda_{min}(F)$for two DO sensors

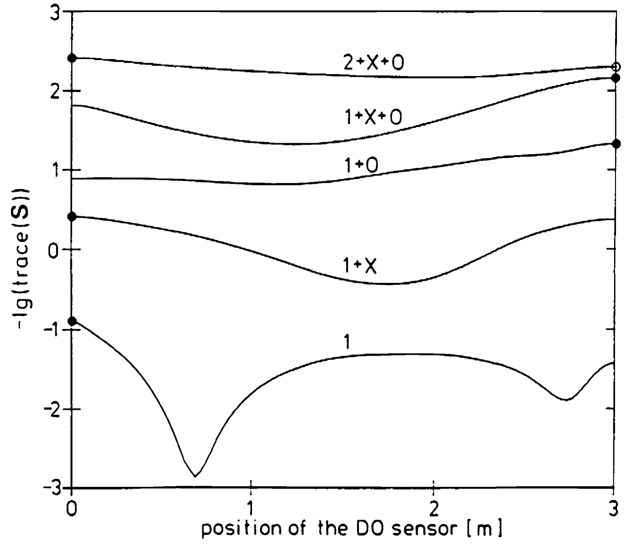


Fig. 5 Results of the sensor allocation problem (trace(S))

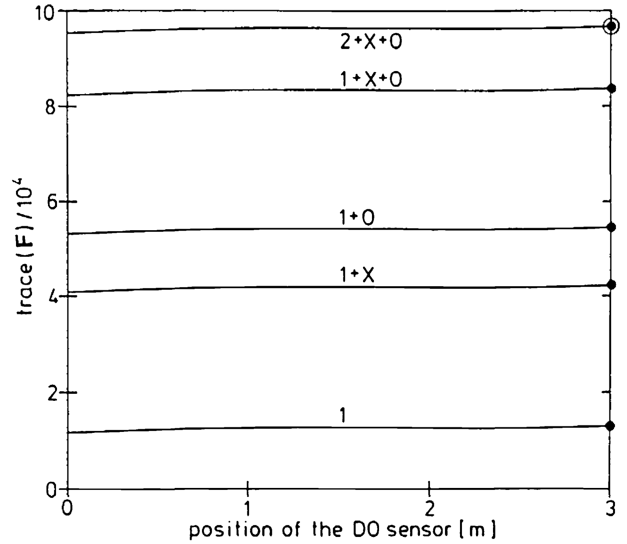


Fig. 6 Results of the sensor allocation problem (trace(F))

A comparison of the results gained by application of the E-criterion with the other criteria gives some further insights into the statements which are possible. Taking the determinant of **F** leads to quite comparable results. The application of the E- and D-criteria to **S** instead of **F** would, of course, give the same results. This is not the case when taking the A-criterion. Fig. 5 demonstrates that trace(**S**) gives quite comparable results to det(**F**) and λ_{min}(**F**); however, Fig. 6 shows that trace(**F**), the most simple criterion, does not provide much information.

The results may be summarized as follows:

- The off-gas analysis is the most informative measurement for parameter identification for the process; in fact, it can be shown that it is superior to eight DO measurements which are optimally placed.
- The best practical instrumentation would comprise cell concentration measurements, off-gas analysis and three DO sensors, which should be placed at the bottom, 40 % of the reactor height and at the top. Then one of the DO sensors could fail during the process, and the other two would still be sufficient.
- This result is "robust" in the sense, that it holds for the simplified model and for the more detailed model as well - indicating that a severe model dependence is not to be expected.

IDENTIFIABILITY OF PARAMETERS IN DIFFERENT STAGES OF THE CULTIVATION

For application of the parameter identification in the OLFO control algorithm, one must ensure parameter identifiability also by inspection of small data sets, e.g. the preceding adaptation interval. This is of great interest, if the plant exhibits time variant parameters or if the model used is quite simple, such that some structural model inaccuracies are compensated by time variant model parameters.

The utilization of the eigenvector components of the information matrix in a supervising system for fermentations was first described by Posten (1990). Concerning the example treated here, Fig. 7 shows the relative standard deviations of the five unknown system parameters, and the standard deviation of the worst identified parameter combination ($\sqrt{\lambda_{max}(\mathbf{S})}$), when only measurement data of the most recent four hours are used. It can be seen, that at the beginning of the cultivation, K_0 is not identifiable - which is easy to understand, since the process is then running under non-limited conditions. During the central part of the cultivation, all parameters are nicely identified. At the end, when the dissolved oxygen profile goes down, identifiability conditions become worse again. Now the coalescence factor $\Delta k_L a$ becomes practically non-identifiable, and the standard deviation of K_0 rises, too. This means, that at the start of the fermentation, one should calculate the process predictions with parameter values from preceding experiments, and at the end, one should not attempt to identify fluiddynamical parameters. Surprisingly, the problem of K_0 identification is not as severe as it was reported for batch processes in stirred-tank reactors, cf. Holmberg (1982). This can be explained by the fact that in the middle stage of the cultivation there is always some part of the reactor where limiting conditions are present, while at the bottom unlimited growth conditions are to be found. This results in relatively nice identifiability conditions for the Michaelis-Menten kinetics. Thus predictive control may be very efficient, since the main parameters (the biological ones) are identified with sufficient precision.

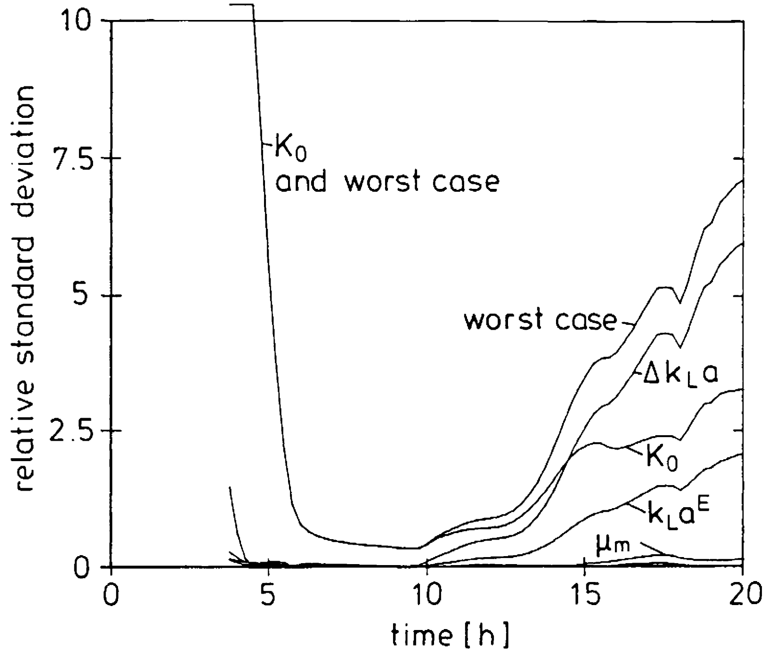


Fig. 7 Relative standard deviations of the unknown parameters during the course of the cultivation (identification horizon = 4 h)

VI. CONCLUSIONS

Measurement facilities for biotechnical processes are somewhat restricted, in particular when relevant biological states are to be measured. Furthermore, biotechnical processes are quite complex and often not reproducibly run. Therefore, model-based measurement techniques, parameter identification techniques for formulation of models and their update, and adaptive predictive control algorithms are well-suited tools for control of these processes. In this contribution, the effects of suitable instrumentation of the plants with sensors and analytical devices were discussed, and the problem of identifiability of the parameters in different stages of the cultivation was addressed. Further research with respect to on-line evaluation of identifiability and automatic modification of input functions in order to guarantee identifiability of the relevant process parameters in each stage of the fermentation is needed.

To keep this contribution concise, no attempt was made to point out the various parallels between biotechnical and agricultural systems. However, it is obvious that greenhouses and animal houses are also described by partial differential equations. Therefore, sensor positioning problems and all addressed related problems arise in the same way. This means that the methods for analysis and design of installations of sensors and sensor systems which were presented here may fruitfully be applied to these agricultural systems, too. The same holds for the OLFO algorithm, which - besides applications to biotechnical processes - was already used for predictive control of an underfloor heating system, cf. Munack (1987).

REFERENCE

Clarke, D.W.; Mohtadi, C.; Tuffs, P.S. (1987). Generalized predic tive control - Part I/II. Automatica, 23, pp. 137-148, 149-160.

Dochain, D. (1986). On-line parameter estimation, adaptive state estimation and adaptive control of fermentation processes. Dissertation, Universit Catholique de Louvain.

Goodwin, G.C. (1987). Identification: Experiment Design. In: System & Control Encyclopedia, Pergamon, Oxford, Vol. 4, pp. 2257-2264.

Halme, A.; Tiussa, E. (1983). Use of computers in Pekilo-protein production. Proc. 1st IFAC Workshop on Modelling and Control of Biotechnical Processes, Pergamon, Oxford. pp. 267-272.

Holmberg, A. (1982). On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities. Math. Biosc., 62, pp. 23-43.

Luttmann, R.; Munack, A.; Thoma, M. (1985). Mathematical modelling, parameter identification and adaptive control of single cell protein processes in tower loop reactors. Advances in Biochemical Engineering, Vol. 32, Springer, Berlin. pp. 95-205.

Munack, A. (1985). On parameter identification for complex bio technical systems. 1st IFAC Symp. on Modelling and Control of Biotechnological Processes, Pergamon, Oxford, pp. 139-145.

Munack, A. (1987). Application of receding horizon adaptive control to an underfloor heating system. IFAC Proceedings Series: "Simulation of control systems", Pergamon, Oxford, pp. 263-268.

Posten, C.; Munack, A. (1990). On line application of parameter estimation accuracy to biotechnical processes. Proc. of the ACC, San Diego, Vol. 3, pp. 2181-2186.

Rehm, H.J.; Reed, G. (1991). Biotechnology - A Multi-Volume Comprehensive Treatise. 2nd ed., Vol. 4, VCH, Weinheim, New York, Basel, Cambridge.

Stephanopoulos, G.; San, K.-Y. (1984). Studies on on-line bioreactor identification. Biotech. and Bioeng., 26, pp. 1176-1218.

Chapter 4

ENVIRONMENTAL CONTROL IN PLANT TISSUE CULTURE AND ITS APPLICATION FOR MICROPROPAGATION

Toyoki Kozai
Byoung Ryong Jeong

Department of Horticulture
Faculty of Horticulture, Chiba University, Matsudo, Chiba 271,
Japan

I. INTRODUCTION

Plant tissue culture is a term used to describe the aseptic culture *in vitro* (in the test tube) of a wide range of excised plant parts for propagation and breeding, biomass production of biochemical products, plant pathology, and preservation and storage of genetic resources. Micropropagation is a specific aspect of plant tissue culture dealing with the aseptic propagation of plants *in vitro* and can be divided into several major stages; stock plant selection and preparation (stage 0), initiation and establishment of an aseptic culture (stage 1), multiplication (stage 2), *in vitro* rooting and conditioning (stage 3), and acclimatization to greenhouse environment (stage 4) (Hartmann et al., 1990).

Micropropagation has many advantages over conventional vegetative propagation methods, such as cutting and grafting, and its use in horticulture, agriculture and forestry is currently expanding worldwide. However, its widespread commercial use is still restricted due to its relatively high production costs resulting mainly from high labor costs, limited growth rates during multiplication, and poor rooting and low survival rates of the plantlet during acclimatization (Kozai, 1991e).

The goal of micropropagation is to obtain a large number of genetically identical, physiologically uniform and developmentally normal plantlets, with a high photosynthetic or photoautotrophic ability (utilizing CO_2 in the air as the main carbon source) to survive the harsh greenhouse conditions, in a reduced time period and at a lowered cost. Development of 1) automated environmental control systems to minimize manual operation and 2) of an improved *in vitro* culture system for increased plant productivity is essential for a significant reduction in production costs.

Recently, extensive research and development has been focused on automation and robotization of micropropagation processes (Aitken-Christie, 1991; Kozai, 1990). Examples are automated liquid medium preparation and feeding system, plant image recognition and processing system, and microcutting and transplanting systems.

On the other hand, only limited research has been conducted for investigation of the effect and control of environmental factors *in vitro* (microenvironment) for increased plant productivity at different culture stages. This is partly due to the fact that conventional culture vessels, such as test tubes and flasks, were small and airtight and control and measurement of the environmental factors were difficult. Read (1990) reviewed literature on the effects of environmental factors on the growth and development of plants during stages 1 and 2 and Dunstan and Turner (1984), Preece and Sutter (1991), and Kozai (1991a) during stage 4.

In this section, microenvironment *in vitro* and its effect on the photosynthesis and growth of the plantlet/explant (plantlet) are discussed. This article is a revised and condensed version of reviews by Kozai (1991c) and Kozai et al. (1992) and the reader is referred to them for further details.

II. SIGNIFICANCE OF ENVIRONMENTAL CONTROL IN PLANT TISSUE CULTURE

The mechanism of environmental changes in a tissue culture vessel is similar to that of the environmental changes in a greenhouse. That is, the interrelationship between a plant and its environment inside and outside the culture vessel is similar to that of a greenhouse plant and its environment inside and outside of the greenhouse. A culture vessel is a miniature greenhouse or a growth chamber in a sense. An explant to be cultured *in vitro* may be thought of as a miniature cutting in conventional vegetative propagation (Read, 1990).

Ecological, ecophysical, ecophysiological, environmental physiological and environmental engineering studies on micropropagation processes should be important (Hughes, 1981), but have long been neglected. Plant tissue culturists have relied upon application of exogenous plant growth regulators (Wilkens, 1988). Few research was conducted on modeling and dynamic simulation of growth of plantlets affected by environmental variables *in vitro*, whereas enormous amount of similar research were conducted on horticultural and agricultural crops grown in the greenhouse and on the field. Developing a growth model for plantlets *in vitro* should be easier than that for greenhouse- or field-grown crops, since many microenvironmental parameters, i.e. air temperature and light intensity, are kept constant over time.

The physical environment *in vitro* in conventional tissue culture systems has been quite different from that in the greenhouse and often encountered undesirable physiological and pathological problems (Debergh and Maene, 1984). The conventional *in vitro* environment are characterized as follow (Kozai et al., 1992): 1) high relative humidity (RH), constant temperature, low photosynthetic photon flux density (PPFD), large diurnal change in CO_2 concentration, presence of sugar, salts and growth regulating substances in the medium at high concentrations, accumulation of C_2H_2 and other toxic substances, and absence of microorganisms. These undesirable environmental conditions often result poor plantlet growth caused by low rates of transpiration, photosynthesis, water and nutrient uptake, CO_2 uptake, and high dark respiration rate.

Traditionally, the plantlets *in vitro* have been cultured under predominantly heterotrophic (on artificially supplied sugar) or photomixotrophic (partly on artificially supplied sugar and the rest on photosynthetically produced carbohydrates) conditions. Explants and regenerated shoots in culture have been considered to have little photosynthetic ability and has been assumed to require sugar in the culture medium as a carbon and energy source.

Recent research (Kozai, et al., 1987; Kozai and Iwanami, 1988; Kozai et al., 1988) has revealed that chlorophyllous explants and shoots, in general, had relatively high photosynthetic ability and that they grew faster in some cases under photoautotrophic conditions than under heterotrophic and photomixotrophic conditions, provided the physical and chemical environments in the culture were properly controlled for photosynthesis (Kozai, 1991b, 1991d). Furthermore, plantlets regenerated from embryos or adventitious buds at a heterotrophic or photomixotrophic phase are expected to smoothly transit into a photoautotrophic phase under controlled environmental conditions. Providing microenvironment for increased photosynthetic and transpirational activities may also be beneficial for the uptake and absorption of mineral elements, which are taken up along the transpirational water flow, i.e. calcium and boron.

Environmental control during micropropagation is important from a practical point of view, especially for the reduction of production costs. Control of microenvironment: 1) promotes plant growth and development, i.e. increases fresh and dry weights, number of nodes and leaf area, 2) enhances rooting and branching, 3) reduces morphological and physiological disorders, such as vitrification (Ziv, 1991) and incomplete development of leaf cuticular wax, and 4) decreases loss of cultured plants resulting from biological contamination by bacteria, fungi or algae. It also enhances uniform plant growth and development and eliminates the need for excessive application of exogenous growth regulating substances. Furthermore, it will encourage more rapid and vigorous plant growth and development during acclimatization stage.

Research to improve *in vitro* environmental conditions for increase and promotion of photosynthetic and transpirational activities by plants in culture is a prerequisite for development of an improved micropropagation system with increased efficiency.

III. SOME CONSIDERATIONS FOR MEASUREMENT AND CONTROL OF THE *IN VITRO* ENVIRONMENT

For the measurement of the microenvironmental factors, many of the existing environmental sensors widely employed in agronomy, horticulture, agricultural meteorology, and ecology can be used. Measurement and control of the greenhouse plant environment are well described by Langhans (1978) and Hanan (1984). For the measurement of the *in vitro* environmental parameters, sensors and the amount and size of sample material should be as small as possible to minimize environment disturbance.

A. Light

Spectral distribution of light from different light sources differs significantly from each other (Bickford and Dunn, 1978). Fluorescent lamps have been the primary light source used in micropropagation, since its spectrum generally matches the requirements of *in vitro* cultures. When the lamps are placed close together, they generally give a relatively uniform horizontal distribution of photosynthetic photon flux density throughout the culture shelf. However, there is a large difference in PPFD between the inside and outside of the culture vessel. The distribution of PPFD in culture vessel on the shelf is largely dependent upon the vessel and closure types, and vessel arrangement on the shelf (Fujiwara et al., 1989).

The light source is usually installed above the vessels and the plantlets in the vessels generally receive downward illumination. As the plantlet grow, increasing amount of light energy is intercepted by the upper parts of the plantlet and only small amount of light energy reaches to the lower parts.

Hayashi et al. (unpublished) compared growth of potato plantlets in vitro cultured for 28 days under downward and sideward lighting conditions (Figure 1). Shoot length was approximately 3.5 cm shorter and dry weight and leaf area per plantlet were 1.8 times greater in the sideward than in the downward lighting treatment. Also there was no difference in plant growth and development affected by location of the culture vessels.

Based on experimental data, it might be better to direct light from the sides by means of optical fibers or other thin or tiny light sources in the future micropropagation systems (Kozai, 1990). With the lateral lighting, the plant will receive an increased amount of light energy evenly throughout all parts even with a decreased amount of electricity consumption for lighting. The lateral lighting promoted plant growth and produced plants with physiologically improved shape (Kozai et al., 1991a).

For control of plant photomorphogenesis, different types of light emitting diodes (LED) can be used to emit either blue, red or far-red light at a low cost. Application of LED for growing plantlets would also be practical (Bula, et al., 1991).

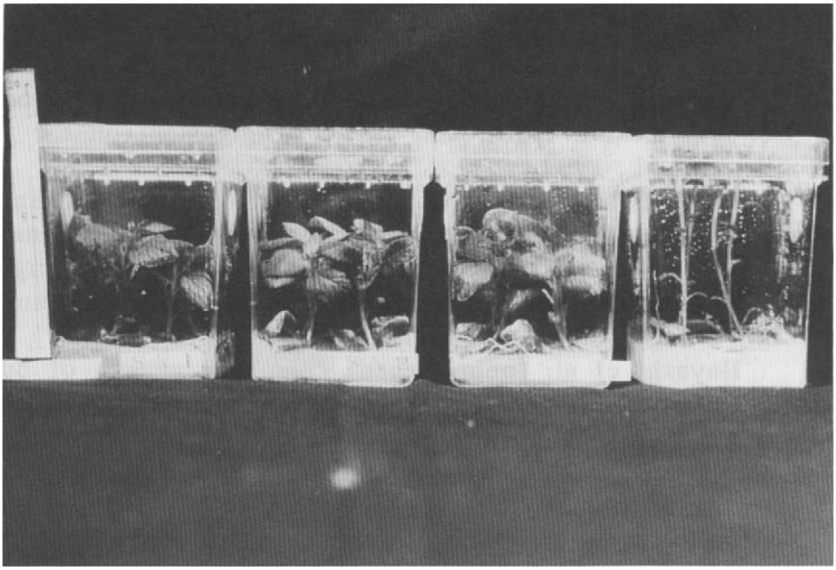


Figure 1. Potato plantlets cultured for 28 days *in vitro* on MS sugar-free medium. Culture vessel in the treatments A1 (bottom), A2 (middle) and A3 (top) were stacked and illuminated using a sideward lighting system. Treatment B were illuminated using a conventional downward lighting system. (Hayashi et al., unpublished).

Morini et al. (1990) tested the effect of different light/dark cycles and found that the growth of peach shoots was significantly greater with 4 hours light and 2 hours dark cycle compared to conventional 16 hours light and 8 hours dark cycle when the same amount of total radiation was supplied. The method of lighting, such as light and dark cycle, light source and spatial arrangement of light sources, is an area of further research.

B. Gas exchange characteristics of the culture vessel

The type of vessel closure affects the gaseous composition as well as the light environment, and hence vitrification and growth of plants in culture. The loose closures was better than the tight ones for reducing vitrification of *Gypsophila paniculata* (Dillen and Buysens, 1989) and carnation plantlets (Hakkaart and Versluijs, 1983) and for promoting the growth of strawberry plantlets (Kozai and Sekimoto, 1988).

The air exchange characteristics of the vessel are best expressed by the number of air changes (infiltration) per hour (E). The E is defined as the hourly air change rate of the vessel divided by the air volume of the vessel (Kozai et al., 1986a). The number of natural air changes for a flat bottom glass test tube (air volume: 45 ml) closed with an aluminum foil cap, plastic formed cap and silicon foam rubber plug were 0.18, 1.5 and 0.6 per hour, respectively (Kozai et al., 1986b). The number can be raised up to 3–6 times by using a gas permeable microporous polypropylene film as part of the vessel closure (Kozai and Sekimoto, 1988).

The number of air exchanges per hour is a physical property of the vessel and is basically constant over time. The gas concentration in the vessel containing a plantlet and medium varies with the gas concentration outside the vessel, the E value and gas production and absorption characteristics of the plant and medium in the vessel. The mathematical relationship among the factors mentioned above was described by Fujiwara et al. (1987). The E can be defined similarly for forced air changes in the vessel.

C. Relative humidity in the vessel and medium water status

The interchange of water in the gaseous and liquid phases among plants, air and root zone (medium) environments in the vessel as well as the characteristics of the air outside the vessel plays an important role in plant growth and development. Direction and rate of water flow are determined by the spatial distribution of water potentials inside and outside the vessel.

Tanaka et al. (unpublished) cultured potato plantlets for 22 days *in vitro* under different RH conditions and observed a greater shoot length with increasing initial RH (Figure 2). In addition, plant fresh weight in the 100 and 93 % RH treatments was greater than that in the 84 and 75 % RH treatments. The specific leaf area ($m^2 \cdot g$ dry weight) decreased with decreasing RH. However, there was no significant difference in plant dry weight among treatments.

Relative humidity is normally high in the culture vessel. Vessel RH can be rather accurately measured with a high polymer sensor (impedance type) if proper and timely calibration is conducted using saturated salts. A compact cylinder type sensor (8 mm in diameter and 35 mm in length) is commercially available at reasonable prices. Most humidity sensors are attached to a thermistor for measurement of both temperature and humidity. Kozai et al. (1990b) used a high polymer sensor to measure RH in the vessel and found a high dependence of vessel RH upon the number of air changes per hour, RH in the culture room and total leaf area of the plantlet.

The water potential of a gelled medium under atmospheric pressure is the sum of the osmotic potential (negative value of osmotic pressure) and matric potential; the matric potential being normally small as compared to the osmotic potential because there is no capillary in the gel. The water (osmotic) potential of liquid medium is the sum of the water potentials created by the basic component (inorganic matter) and sugar (carbon source and some agents used for osmotic potential regulation, i.e. mannitol and sorbitol) (Kozai, et al., 1986b).

The water potential caused by inorganic matter of several widely used culture media were measured (in kPa) −212 (Murashige and Skoog), −89 (Heller), −37 (White), −143 (Gamborg; Miller and Ojima), and −106 (Nitsch; Schenk and Hildebrandt) (Kozai et al., 1986b). The water potential created by mono− (i.e. glucose, fructose, mannitol, sorbitol) and di−saccharide (i.e. sucrose) sugars can be estimated by multiplying −1.46 and −0.78, respectively, to its weight per volume (%) (Kozai et al., 1986b).

Figure 2. Potato plantlets cultured for 22 days *in vitro* on MS agar medium containing 20 $g \cdot l^{-1}$ sucrose and under 4 different relative humidity conditions in the culture vessel, 75, 84, 93, and 100 %. (Tanaka et al., unpublished).

D. Temperature

Although the set point of air temperature in culture room is not changed throughout the day in most cases, the temperature distribution is somewhat uneven in space and over time. The temperature inside the vessel is approximately 1 C higher than that outside the vessel during the photoperiod. Urban and Jaffrin (1990) developed a mathematical model for heat and mass transfer inside culture vessels and applied it to various physical conditions to predict thermal phenomena in the vessel.

Kozai et al. (1991b) examined the effect differences between photoperiod and dark period temperature (DIF) and PPFD levels on morphogenesis and growth of potato plantlets *in vitro* under a CO_2 enriched condition (1300–1500μmol·mol) during photoperiod. With the same daily average temperature of 20C, the air temperature during photoperiod and dark period were set at 25 C/15 C (+10 DIF), 20 C/20 C (0 DIF) and 15 C/25 C (-10 DIF) and PPFD at 74 $\mu mol \cdot m^{-2} \cdot s^{-1}$ (low) and 147 $\mu mol \cdot m^{-2} \cdot s^{-1}$ (high). The shoot length was greater with increasing DIF under both low and high PPFD conditions (Figure 3). Number of unfolded leaves slightly increased with decreasing air temperature difference. However, dry and fresh weights of plantlets were similar between DIF treatments. They concluded that DIF was an efficient way of controlling plantlet height *in vitro* with minimum heating and cooling costs.

E. CO_2, O_2, and ethylene

Concentrations of CO_2, O_2 and ethylene (C_2H_4) can be measured simultaneously using a gas chromatograph with a flame ionized detector (FID). CO_2 concentration can also be measured with an infrared CO_2 analyzer which is convenient for a continuous measurement. However, a continuous measurement requires a relatively large amount of gas sample and may disturb the gaseous environment in the vessel.

There are a few ways of modifying the gaseous components in the vessel: a) addition of agents absorbing or producing a particular gas component, b) control of the gaseous environment in the culture room and use of a gas permeable vessel closure, and c) use of a forced ventilation or air circulation system with a controller. For details, the reader is referred to a review paper by Kozai (1991c).

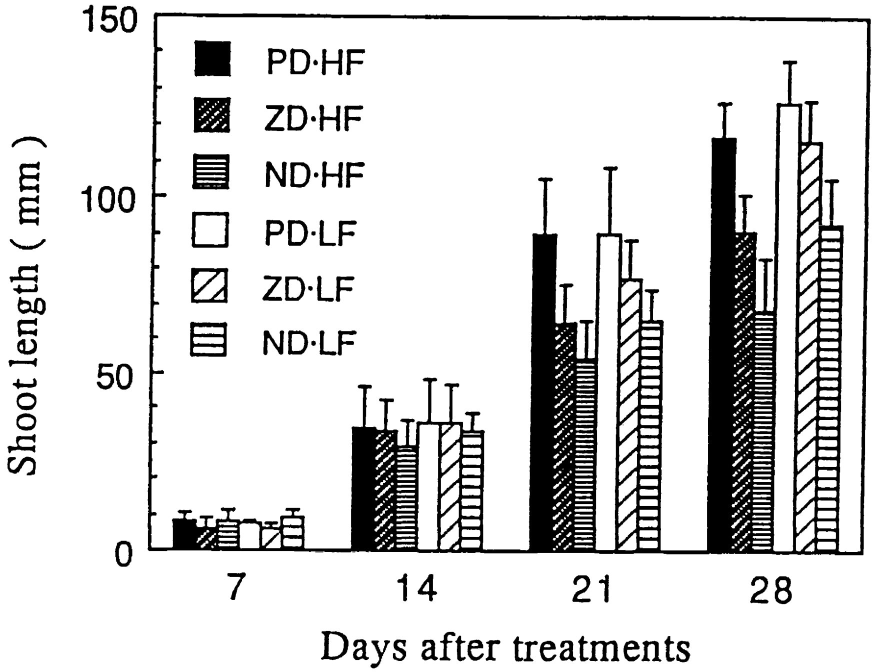


Figure 3. Potato shoot length affected by difference in photoperiod and dark period temperature (DIF) on day 7, 14, 21 and 28 in culture. Daily mean temperature was 20C in all treatments. Plantlets in the PD, ZD and ND treatments were grown under day and night temperatures of 25/15, 20/20, and 15/25 C, respectively. Plantlets were cultured under either low (LF, 74 $\mu mol \cdot m^{-2} \cdot s^{-1}$) or high PPFD (HF, 147 $\mu mol \cdot m^{-2} \cdot s^{-1}$). Vertical bars represent standard deviation from the mean of duplicate. (Kozai et al., unpublished).

F. CO_2 concentration in the vessel during the photoperiod

The decrease in CO_2 concentration in the culture vessel during the photoperiod (C_{in}) was first reported by Ando (1978) in orchids at the rooting stage.

The C_{in} in airtight vessels containing ornamental plantlets decreased to 70–80 ppm in 2 to 3 hours after the onset of the photoperiod (Fujiwara et al., 1987; Infante et al., 1989). They also found that C_{in} was as low as the CO_2 compensation point of C_3 plants and was about 250 ppm lower than the normal atmospheric CO_2 concentration (ca. 340 ppm).

Results of Fujiwara et al. (1987) and Infante et al. (1989) suggest several things. (1) The chlorophyllous plantlet has a photosynthetic ability since the C_{in} decreases sharply after the onset of the photoperiod in the conventional air tight vessels. (2) Insufficient amount of CO_2 supply from the outside air through the vessel cap limits photosynthesis during most of the photoperiod. (3) Plants are forced to develop heterotrophy or photomixotrophy and a raised PPFD will not increase net photosynthetic rate under such low CO_2 conditions. (4) Plants develop photoautotrophy and grow faster under photoautotrophic conditions with high levels of CO_2 and PPFD than under hetero- or photomixo-trophic conditions. (5) The initial growth rate is greater for an explant with a large area of chlorophyllous tissue.

G. Net photosynthetic rate

Estimation of the net photosynthetic rate (NPR) of plantlets *in vitro* is mostly conducted using a small assimilation chamber with a forced air mixing system. Measurement of CO_2 concentration has been made using an infrared CO_2 analyzer (Capellades, 1989; Donnelly and Vidaver, 1984; Grout and Ashton, 1978; Reuther, 1988). The NPR is estimated by multiplying the forced air flow rate though the chamber to the difference in CO_2 concentration going in and out of the chamber. The difference in CO_2 concentration going in and out is relatively small even at a low air flow rate since the chlorophyllous parts of the plantlet and hence their CO_2 uptake rate is small.

In the assimilation chamber method, physical parameters, i.e. gas diffusion coefficient adjacent to the plantlet, and the physiological conditions, i.e. water status of the plantlet in the chamber, may differ from those *in vitro* and *in situ* (in the vessel). The NPR of the *in vitro* plantlet *in situ* may then differ from those in the chamber with identical levels of PPFD, CO_2 and temperature. One must be careful, therefore, to interpret the results obtained with the assimilation chamber method.

Fujiwara et al. (1987) developed a method for numerically estimating values of different parameters involved in photosynthesis of the *in vitro* plantlet *in situ* using the data on the diurnal fluctuation of CO_2 concentration inside and outside the vessel. Kozai and Iwanami (1988) estimated the steady state NPR of the *in vitro* plantlet *in situ* by multiplying the difference in CO_2 concentration inside and outside the vessel to the natural air ventilation rate. Hourly natural ventilation rate is a product of the number of air changes per hour of the vessel, the air volume of the vessel and a unit conversion factor. In these two methods, CO_2 concentrations were measured using a gas chromatograph with a negligible disturbance and the NPR of the *in vitro* plant could be estimated *in situ*. A similar method can be employed for estimation of the dark respiration and transpiration rates.

IV. PHOTOAUTOTROPHIC GROWTH AND DEVELOPMENT OF THE PLANTLET *IN VITRO*

In photoautotrophic micropropagation, the photosynthesis, growth and development of the *in vitro* plantlet are significantly influenced by the physical environmental factors, such as light source and intensity, CO_2 and O_2 concentrations, humidity, air flow rate and temperature. Literature on the effect of environmental factors on photoautotrophic micropropagation was extensively reviewed by Kozai (1991a, 1991b, 1991c). In this section, the environmental effect on the net photosynthetic rate (NPR), growth and development of the plantlet *in vitro* will be briefly reviewed.

A. Photosynthetic response of plantlets *in vitro*

Kozai et al. (1990a) studied photosynthetic response of the *Cymbidium* as affected by CO_2, PPFD and temperature levels and found that the response curves of the *in vitro Cymbidium* plantlet *in* *situ* was similar to those of plants grown in the shade in the greenhouse.

When CO_2 concentration was fixed at approximately 200 ppm, the NPR in the *Primula malacoides* (C_3 plant) plantlet *in vitro* in 1 and 10 % O_2 was approximately 3 and 1.5 times higher, respectively, than that in 21 % O_2 resulted from a reduced photorespiration at lower O_2 concentrations (Shimada et al., 1988).

The NPR of the *in vitro* rose plantlet was increased when cultured on the medium with a lowered sucrose concentration (Capellades, 1989). A similar result was obtained in the potato plantlet *in vitro* (Nakayama et al., 1991). The leaf starch content of the plantlet was increased when cultured on the medium with a raised sucrose concentration, while an increased leaf starch content was associated with a lowered NPR (Capellades, 1989).

The NPR was also affected by the RH (Capellades, 1989). The NPR in strawberry was higher when plantlets were cultured in a vessel with a forced ventilation than with natural ventilation (Kozai et al., 1989). Under natural ventilation, CO_2 diffusion into the stomata was probably restricted since the air was almost stagnant since air movement in the vessel was caused only by natural convection.

The NPR of the plantlet and seedling *in vitro* under saturated PPFD, 340 ppm CO_2 and a leaf temperature of 20 C were similar, regardless of the fact that NPR affected by CO_2 concentration was slightly different (Pospisilova et al., 1987).

Significance of photosynthetic response of the *in vitro* plantlet *in situ* as affected by the *in vitro* physical environment has only been recently recognized and there are many aspects that need further investigation.

B. CO_2 enrichment under high PPFD

Carbon dioxide enrichment under high PPFD (100–200 $\mu mol \cdot m^{-2} \cdot s^{-1}$) was effective for promoting shoot and plantlet growth of potato (Kozai et al., 1988) and tobacco (Mosseau, 1986) when cultured on medium with and without sugar. Kozai and Iwanami (1988) observed enhanced carnation plantlet growth under a condition with CO_2 concentration of 1000–1500 vpm and PPFD of 150 $\mu mol \cdot m^{-2} \cdot s^{-1}$, due a large extent to CO_2 enrichment. This growth promotion was observed in treatments with and without sugar addition in the medium (Figure 4).

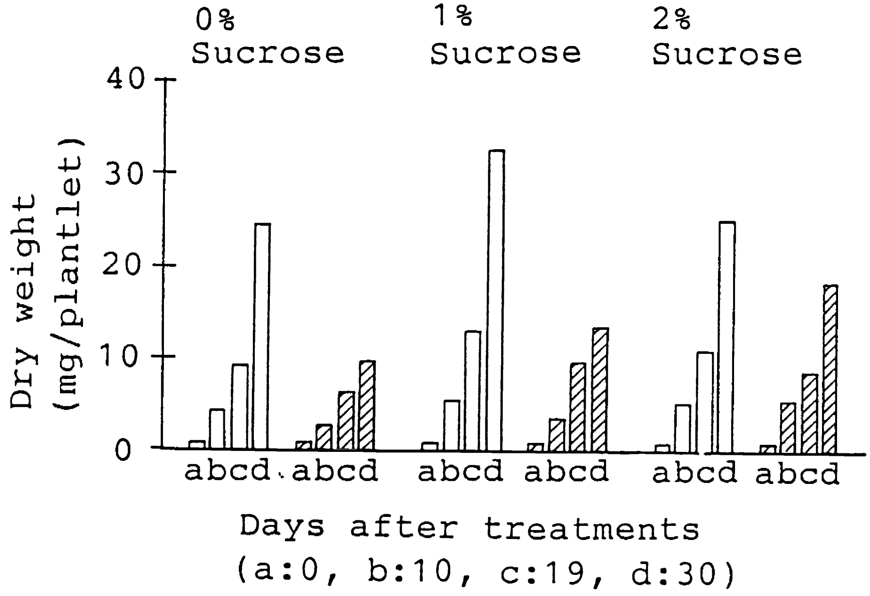


Figure 4. Changes in carnation plantlet dry weight over time as affected by air CO_2 enrichment and medium sugar content under high PPFD. Measurements were made on day 0, 10, 19, and 30. Treatment with CO_2 enrichment produced plantlets with greater dry weight than CO_2 non-enriched treatment. Treatment with 1 % sucrose gave greater plantlet dry weight compared to treatments with 0 and 2 % sucrose. (Kozai and Iwanami, 1988). Open bars represent CO_2 enriched and shaded bars CO_2 non-enriched treatment.

Based upon the data presented above, one can expect an increase in NPR and hence growth and development of the plantlet *in vitro* if CO_2 concentration in the vessel is raised during the photoperiod. There are a few practical ways of raising vessel CO_2 concentration.

1) Use of CO_2 permeable film in the closure

Several reports indicate positive effects of gas permeable film as closure under high PPFD on NPR and growth of the plantlet *in vitro* (Kozai, 1991d). Plantlets of some species derived from leafy single node cuttings grew faster when cultured photoautotrophically in the vessel closed with the gas permeable film than when cultured heterotrophically in the relatively airtight vessel. Under high PPFD even passive CO_2 enrichment, which requires only a minor change in the existing micropropagation facility, will significantly enhance plantlet growth *in vitro*. The percentage of vitrification was also decreased with the use of the gas permeable film, probably resulted from a lowered RH and an increased gas exchange and dehydration of the medium.

2) CO_2 enrichment in the culture room

CO_2 enrichment under high PPFD (100–200 $\mu mol \cdot m^{-2} \cdot s^{-1}$) was effective for promoting the growth of chlorophyllous tobacco (Mosseau, 1986), *Cymbidium* (Kozai et al., 1987), carnation (Kozai and Iwanami, 1988) and potato (Kozai et al., 1988) plantlets regardless of the medium sugar content.

3) A large culture vessel with a CO_2 supply system

Dry weight and NPR of strawberry plantlets cultured on the sugar free liquid medium were greater when cultured in a large vessel with a forced ventilation system under a PPFD of 96 $\mu mol \cdot m^{-2} \cdot s^{-1}$, compared to those of plants cultured using a conventional method (Fujiwara et al., 1988).

However, forced ventilation with atmospheric air or a N_2–O_2–CO_2 mixture reduced propagule weight and shoot number in stage 2 *Rhododendron* cultured in the vessel with 400 ml air volume and under a PPFD of 39 $\mu mol \cdot m^{-2} \cdot s^{-1}$ (Walker et al., 1988).

In those systems described above, not only CO_2 concentration but also RH, ethylene concentration, and gas diffusion in the vessel are modified. Therefore, the changes in growth of plantlets *in vitro* resulting from the use of these systems cannot be attributed totally to CO_2 enrichment. However, in most cases the changes are probably caused primarily by the CO_2 enrichment. More studies on the effect of different gaseous environments and forced ventilation on the photoautotrophic growth and development of the plantlet *in vitro* are needed.

V. ADVANTAGES OF PHOTOAUTOTROPHIC MICROPROPAGATION OVER CONVENTIONAL MICROPROPAGATION METHOD

Some disadvantages and problems of hetero- and photomixotrophic micropropagation are summarized as follow: 1) Addition of sugar as a carbon source in the medium increases the incidence of biological contaminations, and airtight, small vessels are commonly used to reduce these contaminations. Therefore, automation, robotization and computerization of the micropropagation system is practically difficult. 2) The air inside vessel is nearly saturated with water vapor and vessel CO_2 and ethylene concentrations become abnormal. Thus, high PPFD becomes ineffective in promoting plantlet growth. 3) Growth regulators are often necessary for plant regeneration. 4) These undesirable environment conditions induce physiological and morphological disorders, growth retardation, and mutation. 5) Ultimately, unstable production cycle, not uniform plantlet growth and high plantlet death rate during the acclimatization stage raise production costs.

On the other hand, photoautotrophic micropropagation have several advantages over the conventional micropropagation method (Table 1) and may enhance plant productivity and plantlet quality and therefore, considerable reduce production costs. Some of the advantages shown in Table 1 have been experimentally proven. However, still many of those points need to be investigated further, especially on the effect of different environmental factors in the air and medium on photosynthetic growth of the plantlet *in vitro*.

Potential benefits of photoautotrophic micropropagation, using chlorophyllous shoots or nodal cuttings as explants, whether automated or not, seems to be great.

Table 1. Some advantages of photoautotrophic micropropagation

1. Growth and development of plantlets are promoted resulting from improved environmental conditions for normal growth and development.
2. Application of growth regulators and other organic matter is minimized.
3. A larger vessel with environmental control and monitoring systems can be used with decreased incidence of biological contaminations.
4. A loss of plantlets due to biological contaminations is reduced and procedures for rooting and acclimatization are simplified.
5. Physiological, morphological and genetic disorders are reduced, and therefore, plantlet quality is improved.
6. The control of plantlet growth and development by means of environmental control is relatively easier.
7. Automation, robotization and computerization can be practically achieved.
8. Production cycle is stable and costs are lowered.

REFERENCES

Aitken–Christie, J. (1991) Automation, p. 363–388. In: Debergh, P. and Zimmerman, R. (eds.). Micropropagation: Technology and Application. Kluwer Academic Publishers, Dordrecht, Netherlands.

Ando, T. (1978) Gaseous environment in the airtight culture vessel containing orchids. *Ann. Autumn Meet. Jap. Soc. Hort. Sci.* 368–369. (Abstr.)

Bickford, E.D. and Dunn, S. (1978) Lighting for plant growth. The Kent University Press, Ohio, U.S.A.

Bula, R.J., Morrow, R.C., Tibbitts, T.W., Barta, B.J., Ignatius, R.W. and Martin, T.S. (1991) Light–emitting diode as a radiation source for plants. *HortScience* 26: 203–205.

Capellades, M.Q. (1989) Histological and ecophysical study of the changes occurring during the acclimatization of *in vitro* cultures. Dissertation. Gent University, Belgium.

Debergh, P. and Maene, L. (1984) Pathological and physiological problems related to the *in vitro* culture of plants. *Parasitica* 40: 69–75.

Dillen, W. and Buysens, S. (1989) A simple technique to overcome vitrification in *Gypsophila paniculata* L. *Plant Cell, Tissue and Organ Culture* 19: 181–188.

Donnelly, D.J. and Vidaver, W.E. (1984) Pigment content and gas exchange of red raspberry *in vitro* and *ex vitro*. *J. Amer. Soc. Hort. Sci.* 109: 177–181.

Dunstan, D.I. and Turner, K.E. (1984) The acclimatization of micropropagated plants, p. 123–129. In: Vasil, I. (ed.). Cell Culture and Somatic Cell Genetics of Plants. Vol. 1. Academic Press, N.Y., U.S.A.

Fujiwara, K., Kozai, T. and Watanabe, I. (1987) Measurements of carbon dioxide gas concentration in closed vessels containing tissue cultured plantlets and estimates of net photosynthetic rates of the plantlets. *J. Agric. Meteorol.* 43: 21–30.

Fujiwara, K., Kozai, T. and Watanabe, I. (1988) Development of a photoautotrophic tissue culture system for shoots and/or plantlets at rooting and acclimatization stages. *Acta Horticulturae* 23: 153–158.

Fujiwara, K., Kozai, T., Nakajyo, Y. and Watanabe, I. (1989) Effects of closures and vessels on light intensities in plant tissue culture vessels. *J. Agric. Meteorol.* 45: 143–149.

Grout, B.W.W. and Ashton, M.J. (1978) Transplanting of cauliflower plants regenerated from meristem culture. II. Carbon dioxide fixation and the development of photosynthetic ability. *Hort. Res.* 17: 65–17.

Hakkaart, F.A. and Versluijs, J.A. (1983) Some factors affecting glassiness in carnation meristem tip cultures. *Neth. J. Plant Pathol.* 89: 47–53.

Hanan, J.J. (1984) Plant environmental measurement. Bookmakers Guild, Longment, CO., U.S.A.

Hartmann, H.T., Kester, D.E. and Davies, F.T. Jr. (1990) Princip les of tissue culture for micropropagation, p. 459–495. In: Hartmann, H.T., Kester, D.E. and Davies, F.T. Jr. (eds.). Plant propagation: Principles and practices. 5th ed. Prentice Hall, Englewood Cliffs, N.J., U.S.A.

Hughes, K.W. (1981) *In vitro* ecology: Exogenous factors affecting growth and morphogenesis in plant culture systems. *Environ. Expt. Bot.* 21: 281–288.

Infante, R., Magnanini, E. and Righetti, B. (1989) The role of light and CO_2 in optimizing the conditions for shoot proliferation of *Actinidia deliciosa in vitro*. *Physiol. Plant.* 77: 191–195.

Kozai, T., Fujiwara, K. and Watanabe, I. (1986a) Effects of stoppers and vessels on gas change rates between inside and outside of vessels closed with stoppers. *J. Agric. Meteorol.* 42: 119–127.

Kozai, T., Fujiwara, K. and Watanabe, I. (1986b) Relation between the culture medium composition and water potential of liquid culture media. *J. Agric. Meteorol.* 42: 1–6.

Kozai, T., Oki, H. and Fujiwara, K. (1987) Effects of CO_2 enrichment and sucrose concentration under high photosynthetic photon fluxes on growth of tissue-cultured *Cymbidium* plantlets during the preparation stage. *Proc. Symp. Floriezel* 87: 135–141. Arlon, Belgium.

Kozai, T. and Iwanami, Y. (1988) Effects of CO_2 enrichment and sucrose concentration under high photon fluxes on plantlet growth of carnation (*Dianthus caryophyllus* L.) in tissue culture during the preparation stage. *J. Japan. Soc. Hort. Sci.* 57: 279–288.

Kozai, T., Koyama, Y. and Watanabe, I. (1988) Multiplication of potato plantlets *in vitro* with sugar free medium under high photosynthetic photon flux. *Acta Horticulturae* 230: 121–127.

Kozai, T. and Sekimoto, K. (1988) Effects of the number of air changes per hour of the closed vessel and the photosynthetic photon flux on the carbon dioxide concentration inside the vessel and the growth of strawberry plantlets *in vitro*. *Environ. Control Biol.* 26: 21–29.

Kozai, T., Kubota, C. and Nakayama, M. (1989) Net photosynthetic rates of plantlets *in vitro* under natural and forced ventilation conditions. *J. Jap. Soc. Hort. Sci.* 58(1): 250–251.

Kozai, T. (1990) Environmental control and automation in micropropagation. *Proc. 4th Toyota Conf. Automation Biotech.* p. 201–227. Aichi-ken, Japan.

Kozai, T., Oki, H. and Fujiwara, K. (1990a) Photosynthetic characteristics of *Cymbidium* plantlet *in vitro*. *Plant Cell, Tissue Organ Culture* 22: 205–211.

Kozai, T., Tanaka, K., Watanabe, I. and Hayashi, M. (1990b) Measurements of relative humidity and transpiration characteristics of plantlets *in vitro*. *Ann. Meet. Agric. Meteorol.* 24–25. (Abstr.)

Kozai, T. (1991a) Acclimatization of micropropagated plants, p. 127–141. In: Bajaj, Y.P.S. (ed.). Biotechnology in agriculture and forestry 17: High-Tech and Micropropagation I. Springer-Verlag, New York, N.Y., U.S.A.

Kozai, T. (1991b) Autotrophic micropropagation, p. 313–343. In: Bajaj, Y.P.S. (ed.). Biotechnology in agriculture and forestry 17: High-Tech and Micropropagation I. Springer-Verlag, N.Y., U.S.A.

Kozai, T. (1991c) Controlled Environments in conventional and automated micropropagation, p. 213–230. In: Vasil, I. (ed.). Cell Culture and Somatic Cell Genetics of Plants. Vol. 8. Academic Press, New York, N.Y., U.S.A.

Kozai, T. (1991d) Micropropagation under photoautotrophic conditions, p. 447–469. In: Debergh, P. and Zimmerman, R. (eds.). Micropropagation: Technology and Application. Kluwer Academic Publishers, Dordrecht, Netherlands.

Kozai, T. (1991e) Photoautotrophic micropropagation. *In Vitro* Cell. Dev. Biol. 27: 47–51.

Kozai, T., Hayashi, M. and Ochiai, M. (1991a) Effect of the sidewards lighting on the growth and morphology of potato plantlets *in vitro*. *J. Japan. Soc. Hort. Sci.* 60(1): 228–229.

Kozai, T., Kushihashi, S. and Kubota, C. (1991b) Effect of the difference in air temperature between photoperiod and dark period on morphogenesis and growth of potato plantlets *in vitro* under photoautotrophic and CO_2 enriched conditions. *J. Japan. Soc. Hort. Sci.* 60(1): 230–231.

Kozai, T., Fujiwara, J., Hayashi, M. and Aitchen-Christie, J. (1992) The *in vitro* environment and its control in micropropagation, In: Kurata, K. and Kozai, T. (eds). Transplant production systems. Kluwer Academic Publishers, Dordrecht, Netherlands. (In press).

Langhans, R.W. (1978) A growth chamber manual. Comstock Pub., Ithaca, N.Y., U.S.A. pp. 222.

Morini, S., Fortuna, P. Sciutti, R. and Muleo, R. (1990) Effect of different light-dark cycles on growth of fruit tree shoots cultured *in vitro*. *Advances Hort. Sci.* 4: 163–166.

Mousseau, M. (1986) CO_2 enrichment *in vitro*. Effect on autotrophic and heterotrophic cultures of *Nicotiana tabacum*. *Photosyn. Res.* 8: 187–191.

Nakayama, M., Kozai, T. and Watanabe, K. (1991) Effect of presence/absence of sugar in the medium and natural/forced ventilation on the net photosynthetic rates of potato explants *in vitro*. (Submitted to Plant Tissue Culture Let.).

Pospisilova, J., Catsky, J., Solarova, J. and Ticha, I. (1987) Photosynthesis of plant regenerants. Specificity of *in vitro* conditions and plantlets response. *Biol. Plant.* (Praha). 29: 415–421.

Preece, J.E. and Sutter, E. (1991) Acclimatization of micropropagated plants to the greenhouse and field, p. 71–94. In: Debergh, P. and Zimmerman, R. (eds.). Micropropagation: Technology and Application. Kluwer Academic Publishers. Dordrecht, Netherlands.

Read, P.E. (1990) Environmental effects in micropropagation. p. 95–125. In: Ammirato, P.V., Evans, D.A., Sharp, W.R. and Bajaj, Y.P.S (eds.). Handbook of plant cell culture. Vol. 5. McGraw–Hill Pub., New York, N.Y., U.S.A.

Reuther, G. (1988) Comparative anatomical and physiological studies with ornamental plants under *in vitro* and greenhouse conditions. *Acta Horticulturae* 226: 91–98.

Shimada, N., Tanaka, F. and Kozai, T. (1988) Effects of low O_2 concentration on net photosynthesis of C_3 plantlets *in vitro*. *Acta Horticulturae* 230:171–175.

Urban, L. and Jaffrin, A. (1990) Steady state thermal conditions inside plant tissue culture vessels submitted to a constant level of irradiation. *Biotronics* 19: 71–81.

Walker, P.N., Heuser, C.W. and Heinemann, P.H. (1988) Micropropagation: Studies of gaseous environments. *Acta Horticulturae* 230: 145–151.

Wilkins, H.F. (1988) Techniques to maximize cutting production. *Acta Horticulturae* 226: 137–143.

Ziv, M. (1991) Vitrification: morphological and physiological disorders of *in vitro* plants, p. 45–70. In: Debergh, P. and Zimmerman, R. (eds.). Micropropagation: Technology and Application. Kluwer Academic Publishers. Dordrecht, Netherlands.

SECTION 2
Optimal Control

Chapter 5

OPTIMAL DIURNAL CLIMATE CONTROL IN GREENHOUSES AS RELATED TO GREENHOUSE MANAGEMENT AND CROP REQUIREMENTS

Hugo Challa

Dept. of Horticulture, Wageningen Agricultural University,
Haagsteeg 3, 6708 PM Wageningen, The Netherlands

Gerrit van Straten

Dept. of Agric. Eng. & Physics, Wageningen Agricultural
University, Bomenweg 4, 6703 HD Wageningen, The Netherlands

I. INTRODUCTION

The control of agricultural production systems is, in general, a difficult task, due to the complexity of the systems involved and the disturbances by unpredictable factors like the weather, the occurrence of pests and diseases and a lack of tools to control the system. As a result the control of such systems is usually based on empirical knowledge and experience. A more systematic and scientifically based solution is, moreover, hampered by poorly defined criteria for control.

In contrast to outdoor production, the growing conditions in greenhouse cultivation can be controlled to a large extent. By means of ventilators and a heating system, temperature and air humidity can be regulated, and the CO_2 concentration of the air can be increased by means of flue gasses or pure CO_2. In the most sophisticated greenhouses, in addition, there are possibilities to influence the level and the duration of radiation by means of screens and/or supplementary lighting. Also, for many crops it is nowadays feasible to control the root environment with respect to temperature and availability of water and nutrients. Especially in soilless culture the concentration and composition of the nutrient solution can be adapted quickly in relation to the environment and the crop status.

The grower needs these wonderful and sophisticated technical achievements to control the production process. Present control systems, however, do not fully exploit the great potentials of modern greenhouse technology (Challa and others, 1988). Fundamental improvements are feasible in the diurnal climate control, if more knowledge on plant physiological and physical processes could be incorporated in new, more intelligent control systems (Challa, 1990). The solution of this problem is complex for a number of reasons:

- the production process, as will be pointed out later, is a complex of a multitude of processes, taking place simultaneously, reacting with different response times and different response patterns to the environmental factors, and characterised by many interactions
- the grower has several objectives that often require, at least partly, conflicting actions
- climate control is a process that takes place without interference of the grower over appreciable periods of time where complex information has to be dealt with in real time under environmental conditions that may deviate considerably from the predictions
- there is a general lack of quantitative knowledge on the greenhouse/crop system required to translate the objectives of the grower into the proper actions

The problem of how to control greenhouse climate is obviously an optimisation problem, because of the conflicting objectives, the fact that each climate factor is affecting several processes simultaneously (Challa, 1990) and because the balance of costs and economic returns has to be taken into consideration. So far this problem has been dealt with separately from different disciplines, such as engineering, management, or horticulture. The purpose of the present study is to investigate optimal climate control in greenhouses in a wider context, linking these disciplines together.

II. OPTIMAL CONTROL, AN ANALYSIS

A. What is Optimal Control?

Optimal control may be defined as the control that satisfies most the goals set by the grower, which is obviously a definition that in its implicit truth does not contribute to any practical solution. For a theoretical analysis of the problem it is thus necessary to elaborate further on these goals and how they are achieved in the management of the nursery. Furthermore the criteria for control, and the relation with processes to be controlled have to be considered more closely, before investigating optimisation of climate control.

B. Management

Management may be defined as the collection of activities directed to reach certain goals. One of the goals of a grower, as an entrepreneur, in general is to maximise his profit. Climate factors strongly affect the production process as well as the cost of operation. Climate control in greenhouse culture is then one of the tools the grower has to manage his nursery, and thus should be considered as a part of the overall management rather than as an isolated activity.

In management theory different management levels are usually distinguished (Anthony, 1965): strategical, tactical and operational level, depending on the planning horizon (Fig. 1). The argument for this distinction is that, in spite of interactions between different levels, basically the decisions at each level are made independently. E.g. there is no point in considering the question of building a new greenhouse (a strategical decision), or changing to another cultivar (a tactical decision) once the crop is established, and likewise it is not feasible to deal with all possible situations that might occur when making a strategical or tactical plan. On the other hand it is clear that the usual management at the operational level does affect the planning at e.g. tactical level.

With respect to climate control, the management levels distinguished have different implications. At the strategical level decisions on capital investments for equipment determine the technical possibilities for climate control. In addition decisions may be made concerning the long-term policy of the nursery, for

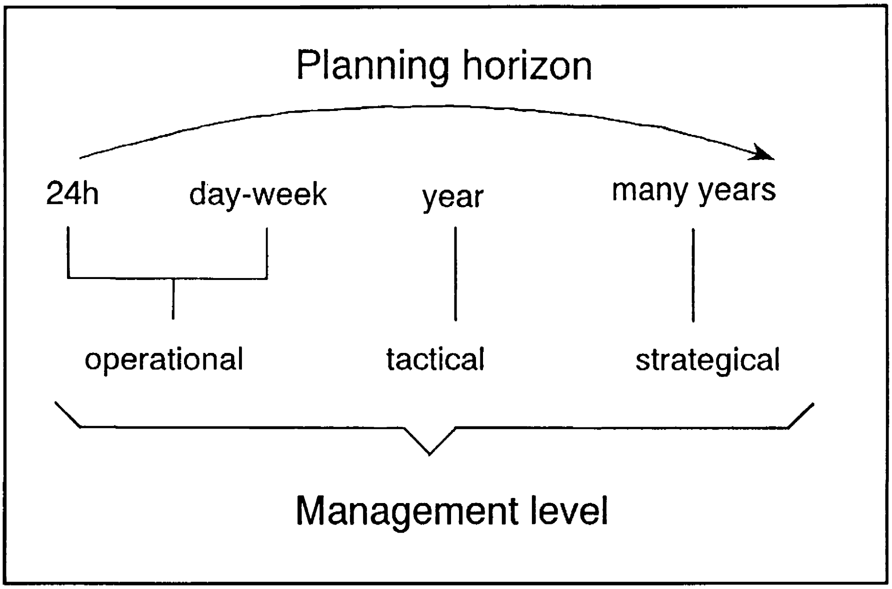


Figure 1. Management levels in relation to the planning horizon; the time scale is indicative rather than absolute.

example with respect to product quality.

Before the start of a new cultivation, or planning cycle a tactical plan is formulated, where the grower decides e.g. what crop and cultivar to cultivate, when to plant or sow the crop, as well as the temperature regime he plans to follow. This plan is, of course, made within the framework of the strategical plan. Connected with the tactical plan is an expectation of average climatic conditions, prices that the grower will receive for his product, as well as a "blueprint" of how the crop will grow, develop and produce as a function of time.

During implementation of the tactical plan considerable deviations may occur, which have to be dealt with at the operational level. Examples are the actual weather conditions, the behaviour of the crop, or unexpected developments in the market. Although deviations of the average do not necessarily have a negative effect, they were not planned and therefore the consequences have to be evaluated and some adaptations may be required. Climate control, because of its close relation with weather conditions and crop response, then is a tool of the operational management.

Operational management is often described as a cycle, progressing in time, of planning-implementation-control (Boehlje and Eidman, 1984). In this cycle "control" refers to the management process and should not be confounded with climate control. It is clear that control of the management process requires criteria. These criteria are derived from the goals set at the tactical and strategical levels and will be further discussed within the scope of the problem of optimal climate control.

C. Criteria for Operational Management in Relation to Climate Control

With respect to climate control the following criteria in operational management can be formulated: physical yield (in kg, or numbers per m^2), crop quality, product quality, timing of the production process, production costs and production risks. These criteria will often give rise to conflicting climate requirements (e.g. yield versus quality, yield versus costs). These conflicts have to be solved, explicitly or implicitly, at the tactical level, but they play also a role at the operational level. The criteria will now be briefly reviewed.

Physical yield. Yield is strongly affected by the climate conditions and as such it is a major criterion for climate control. At the operational level it is important, however, to realise that short-term yield increase may have negative long-term implications (see crop quality), an aspect considered at the tactical level. When yield increase requires extra economic inputs, as may be the case with e.g. CO_2 enrichment, additional yield and associated extra costs have to be compared (Challa and Schapendonk, 1986). Photosynthesis is a major yield determining process, that is primarily sensitive to radiation and CO_2 concentration, but to a lesser extent also influenced by air humidity (Bakker, 1991) and temperature (Challa, 1989).

Crop quality. Especially with long season crops the grower is anxious to keep his crop in a good condition for production. To maintain the productivity of the crop the internal balance between vegetative and generative growth is an important criterion with many crops (Ho and Hewitt, 1986). Temperature is the major climatic factor controlling this balance. Other phenomena affecting crop quality are the occurrence of physiological disorders and of pests and diseases, which are dealt with under "risk prevention".

Product quality. Quality is a concept with a wide scope. In relation to climate control the external quality (e.g. size, weight, shape, colour) and the absence of visible injury are particularly relevant for the selling price. With respect to internal quality keeping-quality and taste are probably influenced by the climatic conditions during cultivation. The relation between climate factors and product quality are highly crop specific and besides are often poorly understood and documented. Temperature and air humidity are important determinants of external product quality.

Timing. The market may show predictable patterns with some crops. Known examples are Christmas (Poinsettia), Easter, Mother's Day, etc. In these cases timing is extremely important. Timing, besides, is also crucial for the cost of production in relation to labour requirements and space utilisation in greenhouses (e.g. potplants). After establishment of the culture the production process can be advanced or delayed to a certain extent through temperature. With daylength sensitive crops flowering can be controlled by the duration of the light period. In practice days are shortened by means of darkening screens, whereas longer days can be provided through low level supplementary radiation.

Production costs. Part of the production costs can be directly attributed to climate control, e.g. heating, CO_2-enrichment, electricity consumption for supplementary lighting. In addition there are indirect effects of climate control due to e.g. cost of labour, or length of the production cycle.

Risk prevention. During cultivation there is a continuous risk of damage to the crop and the product, due to pests, diseases, physiological disorders and environmental stress. Climate control to a certain extent contributes to prevention and control of these risks. Humidity (Hand, 1988) and temperature, but sometimes also radiation have to be kept within certain limits in order to prevent acute problems. Beside these instantaneous reactions there are long-term adaptations of the crop to the climatic conditions that determine its sensitivity to pests, diseases and environmental stress (Levitt, 1980). Though pests and diseases may be controlled by means of pesticides and through biological control it should be realised that in order to attain environment-friendly production techniques adequate climate control is an important tool.

A characteristic of many of the criteria mentioned is the absence of an exact standard and the difficulty of quantification in economic terms. There is a notion of an "ideal" and of unacceptable situations, but in between there is often a gradual range. Moreover the factors involved in the criteria mentioned are common to many of them, which makes an independent control impossible. We believe that these are important reasons why the grower has to be closely involved in the determination of control procedures.

D. Greenhouse Climate and Growth Processes

The production process.

Crop production is a process where, as a first step, radiative energy of the sun is trapped and transformed into chemical energy. This chemical energy subsequently is used to reduce CO_2 molecules and to form the essential building blocks: sugars, amino acids and organic acids (Penning de Vries, 1975). Sugars in addition provide energy for growth and maintenance of the crop. The building blocks are transported to the growing centres in the plant and together with water and nutrients they are used for growth of cells and organs. While growing, the crop is in continuous development: leaves are formed, maturing and ageing, new shoots are developing, the crop may pass from the vegetative to the generative phase and it may form storage organs. This complex of processes is what can be summarised as the production process. In relation to climate control it is worthwhile to classify this complex process with respect to response type and response time.

Response types.

Some crop reactions show a smooth, continuous response to one or more climatic factors within the range normally encountered in greenhouses. Other reactions, usually unwanted reactions, only become manifest if some boundary condition is exceeded. The boundary conditions, beside showing an interaction with other climatic factors, are often affected by pre-conditioning.

This distinction in smooth continuous and strongly non-linear threshold reactions is important, because smooth continuous reactions have always to be taken into account, whereas threshold reactions have to be considered only when the boundary conditions are approached. Given the multiple effects of each of the controllable factors the control problem then can be largely reduced by focusing on continuous responses, while respecting the boundary conditions of the remaining responses.

Response time.

The processes involved in production have largely deviating response times. Although this fact complicates the problem to a certain extent, at the same time it offers a clue to handle this complexity with respect to climate control: for processes with a response time > 24 h it is usually acceptable that, within the response time, in stead of a constant level of the factor in consideration, certain deviations do occur that can be compensated for later on (Cockshull, 1988). Requirements with respect to processes with a slow response time then can be formalised as one or more constraints on these processes or factors, rather than adopting fixed set-points, and in this way more room is left for requirements of processes with a short response time.

Examples of processes with a response time > 24 h are development of the crop, the distribution of assimilates over various plant parts, formation of leaf area, flowers, fruits and tubers, the development of Ca-deficiency, weak/soft plants, vegetative/generative plants, etc. An example of rule based criteria is the prevention of Ca-deficiency, where a suitable model is still lacking (Aikman and Houter, 1990), but where practical experience could be summarised in a set of rules concerning transpiration requirements.

When dealing with processes with a response time < 24 h there is clearly no or only very limited interaction with the grower possible, and compensation in time is not possible either.

Typical processes with a short response time are photosynthesis and transpiration. In addition acute stress, due to temperature (chilling, overheating), radiation (e.g. photobleaching), and extreme water conditions (water stress, tissue damage due to too high water potential) fall in this category. Also the occurrence of condensation on plant parts or high humidity (e.g. poor release of pollen in tomato; Bakker, 1991) may require action within 24 h.

III. OPTIMAL CONTROL, A SYNTHESIS

Because climate control is so intimately linked to management, optimal control should be considered as one aspect of optimisation of the management. At the operational level, the grower uses climate control as a tool to optimise the production process. Knowing the status of the crop, developments of the market, weather predictions and other relevant information, he decides about required actions during the next planning interval, including the desired climate conditions. Optimisation of this process may be enhanced by decision support systems, a research topic of management and computer science (Hofstede, 1992).

During implementation of the operational plan the climate control system has the task to control the greenhouse climate. A new element in the management process described so far is the absence of interaction with the grower for most of the time. Because the actual weather conditions may deviate considerably from expected, climate control should be optimised in real time, that means automatically. The control system, in order to accomplish this task, has to be informed explicitly about the criteria for optimisation and requires, in addition, sufficient information about the processes to be controlled.

Major criteria for climate control have been mentioned before, but cannot be used as such for automatic decisions. In interaction with the grower they should be translated in more specific criteria. In this translation step, the grower is forced to attribute values to these criteria and, while doing so, he implicitly or explicitly makes decisions about the relative importance of different goals. He could (and probably should) be supported in this process by the system by predictions of the consequences of his suggestions.

Operational management, as stated previously, concerns planning over periods of days to weeks. Ultimately, the goal of the operational management would be the optimisation of economic yield, but since the weather plays an important role in climate control and reliable weather forecasts cover only one to a few days, it is desirable to look for possibilities to detach minute-by-minute automatic optimal control from medium and long term optimisation. In view of the inability to forecast the weather, and besides interaction with the grower normally takes place at least once a day, it is reasonable to consider one full 24 h day/night cycle as a suitable basis for unsupervised automatic optimal control. Based on the analysis presented before we believe that the criteria for the design of automatic optimal control could be formulated as follows (a full account of the theoretical background of these criteria cannot be given within the scope of this chapter, they are partly based on the references indicated and supplemented by personal judgements):

1. All climate factors: ideal detachment between long and short term control can be achieved if the actual value of a unit of

photosynthesis can be assessed at any time. In this case the control can be based upon maximisation of the economic value of crop photosynthesis minus associated costs (Challa, 1990). This issue will be discussed in larger detail below; it will appear that this criterion is a special case of the more general but more difficult to handle criterion of maximising the economic yield of the crop minus the associated costs.

2. CO_2: acceptable maximum concentration to avoid risks (Hand, 1990); technical boundary conditions: availability, both instantaneously and over a whole day (Nederhoff, 1990).
3. Temperature: maximum and minimum value as related to other conditions (stress); average daily temperature and average day and night temperature (Cockshull, 1988).
4. Water vapour pressure of the air: prevention of condensation on plant parts during a given maximum period of time; minimum and maximum turgor (the water pressure within the cells); minimum transpiration rate; minimum and maximum relative humidity (pollination); transpiration integral by day, by night and over 24 h (Hand, 1988).

The criteria 2 - 4 have the form of constraint conditions, either instantaneously, or integrated over a period of time. They arise because of effects not described by the crop models, and are still largely based upon experience. It should be noticed that for criteria based on process rates often no direct observation is available, which complicates the determination of boundary values as well as the control. Nevertheless we believe that the introduction of process oriented criteria will lead to a greater universality and flexibility. Models will be required to relate measurable variables to internal plant phenomena.

It should be noticed that the boundary conditions formulated might give rise to conflicting solutions. The system therefore should, in some way, be informed about priorities and, related to this, about the characteristics and uncertainties with respect to these boundary conditions.

A. Framework for Optimal Climate Control

Although the basis for diurnal automatic control is a short period of, say, 24 hours, as argued before, the design of an economically optimal control scheme cannot be done without due regard to the long term. The framework for model-based optimality is sketched as follows. First there is the dynamic behaviour of the greenhouse climate factors in response to control actions. Linked to this there is the immediate response of the photosynthetic rate with respect to the climate factors and external light. Also crop transpiration is of special interest because of its effect on air humidity and on the energy balance of the greenhouse/crop system. And, second, there is the response of the crop in the processing of photosynthetically assimilated material.

Model representation

Mathematically, this dynamic behaviour can be formulated in differential equation form as

$$\frac{dx_c}{dt} = f(x_c, x_p, u_e, u_c) \tag{1a}$$

$$\frac{dx_p}{dt} = g(x_c, x_p, u_e) \tag{1b}$$

with

x_c : the indoor climate state variables, e.g. temperature, CO_2-concentration and relative humidity,

x_p : the plant state variables, e.g. biomass, leaf area index, assimilate distribution over root and shoot, leaves and fruits,

u_e : the external inputs, e.g. solar radiation, outdoor temperature, outdoor humidity, outdoor CO_2, wind speed

u_c : the control inputs, e.g. heat supply, ventilation, CO_2-input flux.

Photosynthesis and respiration are not directly visible in this generic representation, but appear as important terms within both equations. This is shown in Figure 2. These terms respond in an immediate fashion to indoor climate and radiation, and are therefore controlled by the greenhouse climate states. The control variables have no direct influence upon the crop status, but act through the dependency of photosynthesis, respiration and transpiration upon the climate state variables.

The separation of the state variables in this way is useful because the vector function f in Eqn. (1a) has relatively fast dynamics (at least in greenhouses without considerable heat storage), while the vector function g in Eqn. (1b) has slow dynamics. This means that the whole system has the characteristics of stiff differential equations.

Goal function

The economically optimal control of the system of Eqns. (1) requires the explicit formulation of a goal function. As was indicated before, the economical aspect can be seen as the maximisation of the difference between crop yield and associated costs. The other factors discussed in the previous section arise because of phenomena which are not covered in the models, but which are nevertheless important. As noted these factors will thus have to be introduced as constraint conditions. A general formulation of the optimal steering problem is: find the control time pattern $\boldsymbol{u}_c$ such that this maximises the goal

$$J = \phi(\boldsymbol{x}_p(t_f)) - \int_0^{t_f} L(\boldsymbol{x}_c, \boldsymbol{x}_p, \boldsymbol{u}_{\bullet}, \boldsymbol{u}_c) \qquad (2)$$

under the constraint that the states obey the dynamic equations (1), and with further constraints on the states or integrated values of the states originating from the limitations 2 – 4 in the previous section. In the case of vegetative crops harvested at the end of the growing season, such as lettuce, the function ϕ represents the benefits obtained from selling the product at harvest time. The market price is an important parameter in this function. The function L represents the control costs made at any instant in time, which integrated over time yields the total costs of control. Important parameters here are prices to be paid for energy and CO_2-supply. In the case of generative crops, such as tomatoes, both costs and benefits are contained in the function L. The time t_f is the time over which the optimization is performed, usually equal to the season length, and possibly the result of a planning optimization.

Idealised solution

Given the greenhouse model and the crop state model, including the photosynthetic rate description, the task is to construct an optimal path generator that calculates the optimal path of both the climate factors $\mathbf{x}_{c*}$ and the steering controls $\mathbf{u}_{c*}$ over the optimisation horizon (see Figure 2). If the sequence of external

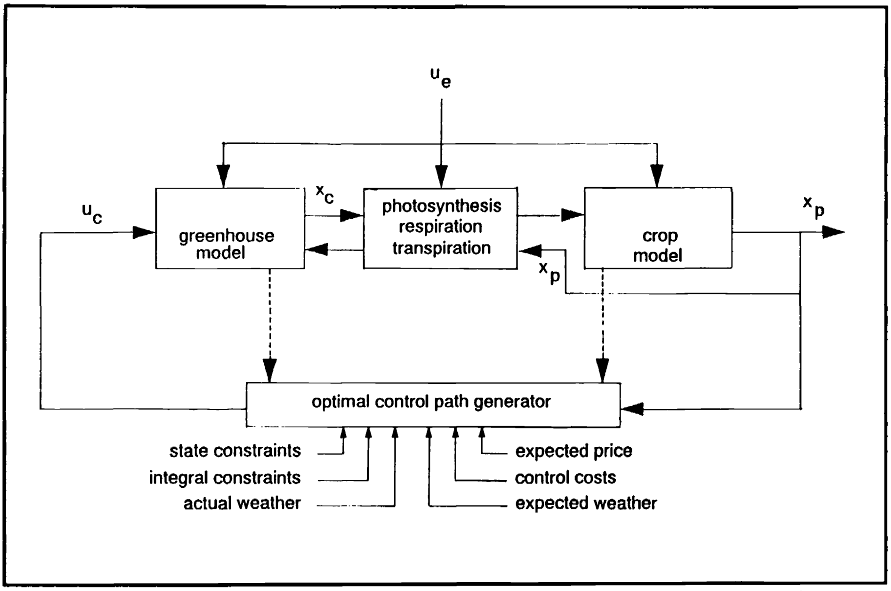


Figure 2. Schematic information flow diagram in optimal control of greenhouses.

weather u_e is completely known over the optimization horizon t_f, then the optimization problem can be solved in principle. In other words, a time path of control inputs u_c must be found that maximizes the goal function J, under constraints on the controls u_c, constraint rules on the calculated crop states x_p, and constraint rules, including integrated requirements, on the climate factors x_c. Several calculation methods are available to perform this task (e.g. Bryson and Ho, 1975) although the incorporation of integral constraints has not received much attention in the literature. An elegant procedure is to form the Hamiltonian function

$$H(x, u_c) = L(x, u_e, u_c) + \lambda^T f(x, u_e, u_c) \tag{3}$$

where λ is a vector of so called co-state variables, quite similar to the Lagrange multiplier in static optimization. This shows that the dynamic equation acts as an equality constraint. The derivative of the Hamiltonian to λ is just the systems dynamics equation, the derivative with respect to x yields the costate dynamics equation, which must be solved backward in time, while the derivative to the control inputs u_c leads to the stationarity condition:

Together with the boundary conditions related to the final condition, and the co-state equation this equations yields an expression of the

$$0 = \left(\frac{\partial H}{\partial u_c}\right)^T = \left(\frac{\partial L}{\partial u_c}\right)^T + \left(\frac{\partial f}{\partial u_c}\right)^T \cdot \lambda \qquad (4)$$

control input as a function of the external inputs, the states and the co-states (Lewis, 1986). The minimization problem is thus cast in the form of a two-point boundary value problem. The numerical solution is not obvious and may be difficult to obtain. Approximations can be found by taking the controls piece-wise continuous. Such a computation yields the optimal discrete sequence of the controls $\{u_c^*\}$, as well as the associated optimal sequence of the state variables. As on the seasonal scale the climate response is fast, equation (1a) can be treated as pseudo-static, as shown by singular perturbation theory (Van Henten and Bontsema, 1992). Also, the greenhouse plus conventional controllers can be viewed as immediately following the set-points of the controllers. In this case, the optimization does not calculate the control inputs, but rather the set-points for temperature, CO_2, and relative humidity. In both cases the stiffness is removed from the dynamic calculation. The idealized solutions have proved to be useful to gain insight in the strategies to be followed over the season (Seginer, 1986, Van Henten and Bontsema, 1991).

Practical solution

The direct implementation of the idealised solution is not possible because the weather is not known in advance. There are various possibilities to advance towards practical solutions, some of these still in their infancy:

a) An approach most directly related to the concept developed before is to directly assign an approximated economic value to each unit of photosynthesis (Challa, 1990, Tchamitchian et al., 1992). This approach is particularly feasible with generative crops. The assessment of an economic value to photosynthesis has been discussed in specific cases by Challa and Schapendonk (1986) and Heuvelink and Challa (1989), but has to be worked out within a wider scope. In single harvest crops there is a relatively long period where photosynthesis is not contributing directly to growth of the harvestable product, and hence a direct assignment of value to photosynthesis is difficult.

b) A more generic solution to the problem is to employ the Hamiltonian to first calculate the seasonal optimal path of the states and the co-states, using an expectation of the weather and the price of the produce. The assumptions made here express by the way in some sense the risk that a grower is prepared to take: he might be either optimistic or pessimistic. In the next step the results of the seasonal optimisation are used as a kind of strategy for the direct automatic control. With vegetative crops it appears that the co-state variables can be interpreted as the economic value of a change in crop state variables in response to changes in the climate factors. Given the practical observation made on a lettuce crop that the co-state pattern is not very sensitive to fluctuations in the weather, this offers the possibility to compute a sub-optimal control in feedback form. (van Henten and Bontsema, 1992). The control action is calculated to satisfy the stationarity condition, using the actual weather and the averaged co-state pattern calculated from the seasonal optimisation. Effectively, this is a comprehensive and generic way to circumvent the problem of assigning an economic value to a unit of photosynthesis. In the approaches tried out thus far, the response of the greenhouse to changes in set-points is still treated as instantaneous, although on the time scale of interest (in the order of minutes) the greenhouse dynamics will also play a part.

c) A comprehensive approach that does not rely on detachment of short and long term optimisation is the following. Once again, the first step is to calculate the optimal solution over the full planning horizon, by taking the actual weather condition, the short term weather forecasts, and averages or high or low assumptions for the longer term weather and prices, and giving due account to the constraint conditions. However, from the calculated control sequence, only the first one is actually applied, and the full calculation is repeated for the next time instant. The repetition of the calculation, supplemented with actual observations on greenhouse and crop states, effectively introduces a kind of feed-back, thus avoiding the problems arising from model and input errors.

The implementation of any of these schemes in practical situations requires that the optimal solution can be obtained within the interval time, which is in the order of a few minutes in greenhouse control. Especially in procedure c this may be a problem. An interesting option here is to calculate a large number of optimal solutions under various external conditions, and then use an artificial neural net to establish a link between control input and the external and state variables. This neural net is then used in stead of the full optimization, which offers the advantage of very fast computations and computational robustness. By observing the actual process, also a retraining of the net can be enforced as soon as the deviations between the model and the true system become too large, thus creating an adaptive capacity within the system. Retraining would also be necessary if the grower wishes to change the constraint conditions, or his expectations about weather and price. More research is needed before these ideas can be put to practice.

B. Information Resources

In this complex optimisation problem the information required comes from different resources: actual climatic conditions inside and outside the greenhouse are provided by sensors, detailed daily weather forecasts can be obtained nowadays commercially in electronic form, and qualitative and quantitative models describe the relations between environmental conditions, actuators and greenhouse and crop reactions. The grower should provide information on the crop, price forecasts and should be able to inform the system about his policy and personal assessment of the criteria used for control. One way of doing this, suggested by the framework above, is in the form of constraint rules.

The structure of the knowledge system that should handle the complex information requirements deserves much attention. In particular the interaction with the grower is an essential aspect, because he bears the final responsibility for the management of the production process, and he needs to feel confident with the solutions provided by the system. The system then should be able to inform the grower what the consequences of his decisions will be, as well as an explain facility to provide the grower insight in the control strategy.

IV. CONCLUSIONS

Climate control should be considered within the framework of the management of the nursery, where it is primarily a tool in operational management. Optimisation of climate control requires criteria that are formulated explicitly, because this process has to proceed automatically. The criteria, derived from the growers goals, have to be tuned in interaction with the grower. The main criterion for optimality is the crop yield, possibly simplified to just the rate of photosynthesis, expressed in economic terms, minus associated costs. Other criteria for control can probably be formulated as constraints.

In order to formulate these constraints some characteristics of crop responses have to be taken into account: a distinction can be made with respect to response time (<24 h and > 24 h) and response type (smooth, continuous versus strongly non-linear threshold reactions). For crop reactions with a response time > 24 h it is usually acceptable that climatic conditions vary within a 24 h period. This characteristic will then give rise to constraints on diurnal integrals, more than on instantaneous values.

Due to the integral constraints an optimisation horizon of at least 24 h is required. Optimisation has to be implemented at two levels: an optimal path over the whole optimisation period has to be generated, while short term control has to deal with the actual conditions, that may deviate from the prognoses, due to the weather conditions and model errors.

A sophisticated knowledge system based on quantitative and qualitative models and a high level user interface is required to deal adequately with the complex information exchange between grower and the proposed climate control system.

REFERENCES

Aikman, D.P. and G. Houter (1990). Influence of radiation and humidity on transpiration: implications for calcium levels in tomato leaves. *J. Hort. Sci.* 65(3), 245–253.

Anthony, R.N. (1965). Planning and control systems, a framework for analysis. Boston: *Harvard University.*

Bakker, J.C. (1991). ***Analysis of humidity effects on growth and production of glasshouse fruit vegetables.*** Diss. Wageningen Agric. Univ., Wageningen.

Boehlje, M.D. and V.R. Eidman (1984). *Farm Management.* John Wiley & Sons, New York.

Bryson, A.E. and Y.C. Ho (1975). *Applied Optimal Control* –

Optimisation, Estimation and Control; John Wiley & Sons, New York.

Challa, H. (1989). Modelling for crop growth control. *Acta Hort.* 248, 209-216.

Challa, H. (1990). *Crop growth models for greenhouse climate control.* In: R. Rabbinge, J, Goudriaan, H. van Keulen, F.W.T. Penning de Vries and H.H. van Laar (Eds.). Theoretical production ecology: reflections and prospects. Simulation monographs, 34, PUDOC, Wageningen. pp. 125-145

Challa, H., G.P.A. Bot, E.M. Nederhoff and N.J. van de Braak (1988). Greenhouse climate control in the nineties. *Acta Hort.* 230, 459-470.

Challa, H. and A.H.C.M. Schapendonk (1986). *Dynamic optimalization of CO_2 concentration in relation to climate control in greenhouses.* In: H.Z. Enoch en B.A. Kimball (Eds.). Carbon dioxide enrichment of greenhouse crops. CRC press, Inc., Boca Raton. pp. 147-160.

Cockshull, K.E. (1988). The integration of plant physiology with physical changes in the greenhouse climate. *Acta Hort.*, 229, 113-123.

Hand, D.W. (1988). Effects of atmospheric humidity on greenhouse crops. *Acta Hort.*, 229, 143-158.

Hand, D.W. (1990). CO_2 enrichment in greenhouses: problems of CO_2 acclimation and gaseous air pollutants. *Acta Hort.*, 268, 81-101.

Henten, E.J. van (1989). Model based design of optimal multivariable climate control systems. *Acta Hort.*, 248, 301-306.

Henten, E. J. van and Bontsema, J., 1991. *Optimal control of greenhouse climate.* In: Mathematical and control applications in agriculture and horticulture. Y. Hashimoto and W. Day (eds.). Proc. of the IFAC/ISHS workshop, 30 Sept. - 3 Oct., Matsuyama, Japan, Pergamon Press, Oxford, p. 27-32.

Henten, E. J. van and Bontsema, J., 1992. *Singular perturbation methods applied to a variational problem in greenhouse climate control.* In: Proceedings of the 31st CDC Conference, Tucson, Arizona, December 1992, 6 pp.

Heuvelink, E. and H. Challa (1989). Dynamic optimisation of artificial lighting in greenhouses. *Acta Hort.*, 260, 401-412.

Ho, L.C. and J.D. Hewitt (1986). *Fruit development.* In J.G.

Atherton and J. Rudich (Eds.). The tomato crop. Chapman and Hall, London. pp. 201–239.

Hofstede, G.J. (1992). *Modesty in modelling: on the applicability of interactive planning systems with a case study in pot plant cultivation.* Diss. Wageningen Agricultural University, Wageningen, the Netherlands.

Levitt, J. (1980). *Responses of plants to environmental stresses.* Academic Press, New York.

Lewis, F.L. (1986). *Optimal Control.* Wiley–Interscience, New York.

Nederhoff, E.M. (1990). Technical aspects, management and control of CO_2 enrichment in greenhouses. *Acta Hort.*, 268, 127–138.

Penning de Vries, F.T.W. (1975). *Use of assimilates in higher plants.* In J.B. Cooper (Ed.). Photosynthesis and productivity in different environments, Cambridge University Press. pp. 459–480.

Seginer, I., Angel, A., Gal, S. and Kantz, D. (1986). *Optimal CO_2 enrichment strategy for greenhouses. A simulation study.* J. of Agric. Eng. Res., **34**, 285–304.

Seginer, I., and Sher, A. (1992). *Optimal greenhouse temperature trajectories for a multi–state–variable tomato model.* Ch. 3–2, this book.

Tchamitchian, M., van Willigenburg, L.G., van Straten, G. (1992). *Short term dynamic optimal control of the greenhouse climate.* Wageningen Agricultural University, Dept. of Agricultural Engineering and Physics, MRS–report 92–3, 22 pp.

Chapter 6

OPTIMAL CONTROL FOR PLANT PRODUCTION IN GREENHOUSES

H.-J. Tantau

Institute for Horticultural Engineering,
University of Hannover
Hannover, Germany

I. INTRODUCTION

One of the goals of horticultural production in greenhouses is to increase the sustainable income of the grower. The investment costs for greenhouses as well as labour and energy cost are much higher compared with conventional plant production. This can be balanced out only with a better utilization of the yielding potential of plants, higher labour productivity and higher energy efficiency. Another goal must be the reduction of pollution and energy consumption.

Higher plant productivity and quality in combination with a reduction of pollution and energy use require a better control of the environment. Additionally to temperature, air humidity, CO_2-concentration and light intensity are controlled in commercial greenhouses. For reducing the pollution of ground water, the control of water and nutrition supply is getting more and more important. Closed irrigation systems including soilless culture such as NFT and Rockwool, can solve the pollution problem but require an improved control of nutrition, e.g. using ion-selective sensors.

As more and more microcomputers are used in commercial green-houses, it is possible to increase the accuracy of environmental control by highly sophisticated control algorithms.

Modern control strategies lead to dynamic control. More knowledge of plant physiology and physical processes can be incorporated in new, more intelligent control systems. The optimization of plant production in greenhouses is a complex problem for a number of reasons. The grower has several objectives, which often require at least partly conflicting reactions. Each climate factor is affecting several processes simultaneously. The balance of costs and economic returns has to be taken into consideration. The maximisation of return is subject to some assessment of risk and forecasting of environmental and market conditions. The solution of this problem opens a large field for application of computers in horticulture.

II. CONCEPTION OF PRODUCTION CONTROL

Research and development have been done to improve the control of plant production by more sophisticated control algorithms and new control strategies. As new tasks have been added, the complexity of the system is increasing. For the use in commercial greenhouses the system must be as flexible as possible in order to accomplish individual demands. This must be realised by a modular system. Furthermore, dividing the system into different control levels creates a hierarchical system. Fig. 1 shows a system with three main levels. Each level may be divided into several sublevels.

The top level is the level of tactical production planning with a planning horizon of about one year. For optimization simplified plant growth models ("planning models") are used in combination with mean values of climate and price data. The risk of production can be estimated using more extrem data of environmental and market conditions.

The middle level represents the operational management of plant production with a planning horizon of 24 h to several weeks. The optimal control problem is defined as to find daily temperature, humidity, CO_2 and light intensity setpoint trajectories that minimize the objective function.

On the bottom level the environmental control is realised using feedback–feedforward control.

In order to increase flexibility the system should be modular, that means, that each level will be divided into several modules or components which are principally independent to each other. In such a control system the modules must be able to communicate. This makes a network necessary. The principle requiries for a network are shown in fig. 2. From the bottom level (actor–sensor–level)

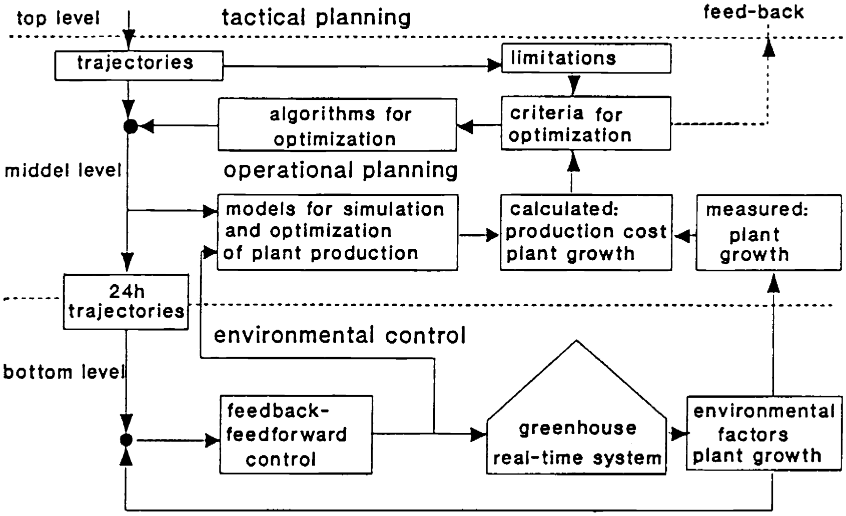


Fig. 1: Structure of production control.

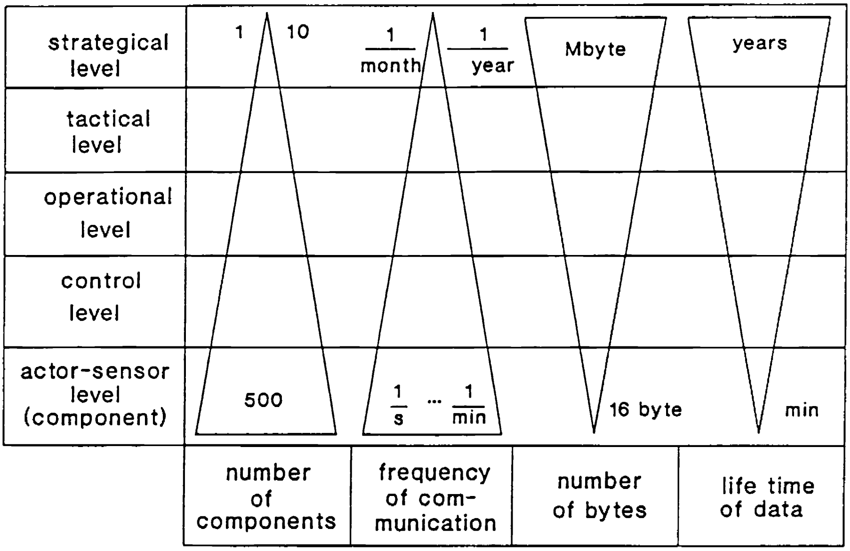


Fig. 2: Criteria for a hierarchical structure of communication and control

to the top level the number of components is decreasing as well as the frequency of communication; the number of bytes for one data transfer and the life time of data is increasing. Local area networks (LAN, e.g. Ethernet) can be used on the top level, while on the lower levels other networks are required which have been developed for industrial control applications.

In fig. 3 some examples are shown. For the application in horticulture the maximum length of the connection link is very important, especially for the control level (Fig. 2). The "Profibus" with the physical link RS 485 is a useable solution for field bus applications.

bussystem	control of bus	transfer rate [Kbit/s]	length [m]	update [ms]	controller
PDV-bus	central	312.5	1000 Koax	27.2	NEB 3000
Profibus	central decentral	1000	1000	40	80C51 80C31
Bitbus	central	375	300	20-40	RUPI
CANbus	central decentral	1000	30	6.4-10.8	82C200 82526
Abus	central decentral	500	30	4.6-6.8	Abus-IC

Fig. 3: Bus systems for industrial control applications (field bus).

III. CONTROL OF PLANT PRODUCTION (middle level, fig. 1)

For operation of commercial climate computers (bottom level) the eventuallity is given to use more than 350 parameter or setpoints for one compartment. With increasing complexity of the control system the number of parameter will increase too. Thus, operating such a system gets more difficult for the grower. The task of the middle level is to calculate the daily setpoints trajectories for the bottom level. The problem of how to control greenhouse climate is obviously an optimization problem. The reasons are

-the conflicting objectives,
-the fact that each climate factor is affecting several processes simultaneously and
-the required attention to the balance of costs and economic returns.

One main problem is caused by the fact, that the future development of outdoor climate and prices is unknown. If the planning horizon is less than one week (at this level) a weather forecast can help solving part of the problem. Due to investigations of Seginer (1991) precise knowledge of the immediate future (about one day) only is required to make correct control decisions. The effect of wrong assumptions about future prices may be much more important.

Plant growth and plant development must be controlled in order to achieve the crop production which is expected at top level. An optimal control of production should effect increasing plant productivity and e.g. decreasing energy input for maximizing economic net results (Schapendonk et al., 1984; Seginer et al., 1986; Day et al., 1991). Repeated optimization helps to compensate the decline that may be caused by deviations from the standard course of the climate (Rheinisch et al., 1989).Considering the corrections which are necessary due to the deviations from assumed standard outdoor climate the demand for a long term strategy arises.

Knowledge about the crop reactions on a change of the environment must be available for optimization. Two different response types are important. Some crop reactions show a smooth continous response to one or more climatic factors within the range normally encountered in greenhouses. Other reactions only become manifest if some boundary condition is exceeded. The destinction in smooth continous and strongly non-linear threshold reactions is important, because smooth continous reactions can be described by plant growth models, whereas threshold reactions can be summarized in a set of rules.

Several researchers are working in the field of modelling plant growth (Augustin et al., 1980; Challa et al., 1986; Liebig, 1989; Nederhoff et al., 1989). Several growth models are available, especially for different vegetables. Descriptive models are used normally for optimization.

One example for the use of such models is the optimized control of thermal screen (Bailey, 1988). By such a strategy the screen will be closed whenever the amount of required energy cost exceeds the monetary value of growth, which is lost due to light reduction. Temperature control by temperature integration is an other example (Bailey, 1985; Bailey, Seginer, 1989). Besides temperature integration the daily course of temperature is important for plant development. New control strategies -e.g. "diff" or "cool morning"- can influence the height of plants, which may be a question of quality.

For control and optimization of plant production a feedback from the plants would be favourable. This feedback can be used for feedback control and for an on-line parameter adaption of the growth models.

Several methods of measurements of plant growth and plant development have been investigated (Hack, 1989; Matsui, Eguchi, 1976, 1977, Hashimoto, 1989; Schoch, 1987):

-net CO_2-uptake (net photosynthesis)
 using the greenhouse as a cuvette;
-remote sensing of plant temperature as an indicator of stomatal opening
-transpiration rate using lysimeter or an electronic balance;
-growth rate of leaves with image processing;
-fresh weight with an electronic balance;
-height of the plants with mechanical sensor;
-stem diameter to detect the water status;
-plant development by image processing.

These methods are applicable in experimental greenhouses, especially for development of new control strategies. The frequency for measurements of plant growth will be in the range 1/h to 1/day (Fig. 2).This approach has been called "speaking plant-approach" (Hashimoto, 1989) or "reading face approach" (Takakura, 1991).

Present day greenhouse crop models contain many state variables. Optaining an optimal environmental control trajectory requires large computational resources, which will limitate the application for on-line optimization. That means, these models should be replaced by mor simple models for control purposes. The equivalent models may contain just a few properly selected state variables, which can be adapted by on-line parameter estimation.

For the optimization of plant production on the middle level numerical optimization techniques offer opportunities to derive optimal control from complex simulation models. The performance of different methods depends largely on the nature of the problem. If little is known about the surface or the solution space, random strategies mostly perform better than directed search methods. One problem is to find the global optimum and not local optima, especially when an on-line optimization is gained.

The implementation of on-line optimization will require the confidence of the grower in the system. Therefore the interface between the grower and the "intelligent" machine must be carefully designed. One possibility is the use of an expert system (s. fig.4).

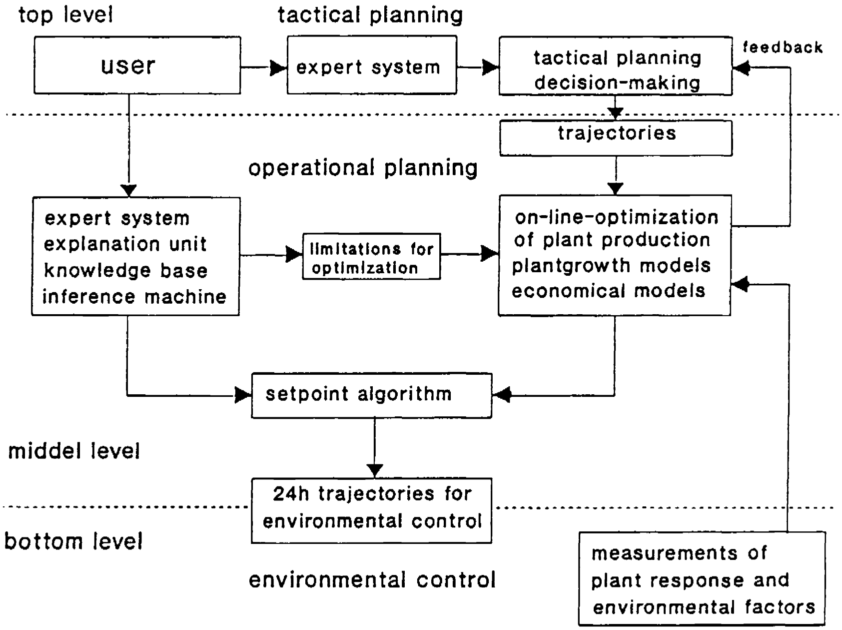


Fig. 4: Structure of the middle level (Fig.1) using on-line optimization and expert systems

The grower can interrogate aspects of the optimization routines to establish the factors leading to particular set point decisions and he can evaluate these factors in relation to his experience and crop management expertise. For this the explanation unit is very important, but unfortunately not included in most of the commercial expert systems.

Several researchers are working in the field of artificial intelligence, using expert systems for the control of plant production (Harazono et al., 1988; Jacobsen et al., 1987; Jones et al., 1988; Kozai, 1985; Kurata, 1988; Leung et al., 1985). The combination of heuristic and procedural knowledge is of high importance for optimal control strategies on the middle level.

IV. ENVIRONMENTAL CONTROL
(Bottom level, fig.1)

In the past most of research and development has been carried out to improve control on the bottom level. For environmental control commercial climate computers for climate control and substrate computers for control of water and nutrition supply (Takakura, 1989) are available.

From the engineering point of view exacte control of the climate in greenhouse is difficult, because of time delay and long time constant of the system (Veerwaaijen et al., 1985). The very thin covering material is necessary to get a high light transmissivity, but disturbances, e.g. solar radiation, outside air temperature and wind velocity, can so influence the climate inside a greenhouse very rapidly. Besides this, the heat load of a greenhouse can be reduced to 50% or even less within a few minutes by using thermal screens or other methods for energy saving.

Several investigations have been carried out to improve the control accuracy. One possibility is the use of adaptive control algorithms (Henten, 1989; Hooper, 1985; Tantau, 1985; Chotai, Young, 1991).

Normally a parameter optimized controller is used (e.g. PI or PID, PIP)(Young, 1991). However, several problems restrict the application of parameter–adaptive–controllers to real process: the necessity to provide a number of spezification parameter in advance, unsatisfactory control behaviour during the adaption phase, undesired effects like bursting of parameter estimates.

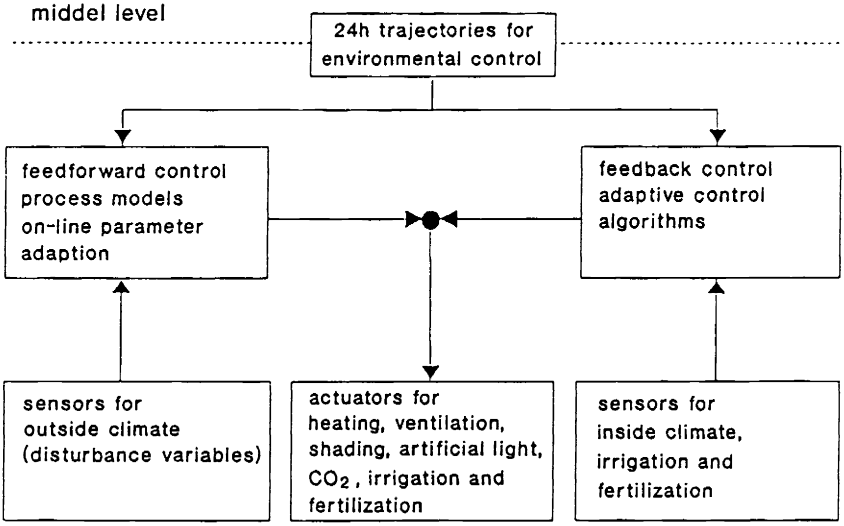


Fig. 5: Structure of feedback-feedforward control
(bottom level, fig.1)(Tantau, 1991)

Feedback control is improved by using feed-forward control additionally. (Tantau, 1985; Tantau, 1989). Fig. 5 shows the structure of the bottom level (Fig. 1) using feedforward-feedback control. Feedforward control includes the measurement of the disturbance variables (e.g. outside climate) and the calculation of the actuators change by means of process models.

Black box models or mathematical models based on heat balances may be used. The processes in the greenhouses are time and place dependent and must be described by partial differential equations (Bot, 1983; Bot, 1989). The system has distributed parameter. Ordinary differential equations, describing the greenhouse as a single compartment, working with single energy buffers, are usual for feedforward control. This method seems to give a reasonable result for the "average" temperature, humidity or CO_2-concentration, even when gradients in the greenhouse occur (Udink ten Cate, Vooren, 1983; Biemond, 1989).

The use of feedforward control has several advantages compared with feedback control (Tantau, 1990).

-The stability of the control loop will increase.

-The control system can react earlier on changes of outdoor weather conditions.

-Overshoots and undershoots will be avoided.

-Movements of the actuators are reduced.

Applying feedforward control needs an on-line process identification and an adaption of the model parameter. This requires models on a higher level of abstraction or the reduction of "tunable" model parameter. On-line parameter estimation is very important for the use of feedforward control in commercial climate computers. Wrong identification and "bursting" of parameter estimates must be prevented. Statistical methods and numerical optimization techniques are applicable for parameter estimation.

V. CONCLUSIONS

In the past most of research in the field of environmental control has been carried out on the bottom level (Fig. 1). The application of mathematics and control have led to an increase of control precision, an increase of stability of the control loops and a decrease of energy consumption. This is one base for an on-line optimization of plant production (middle level).

Another base is the knowledge of the plant response to a change of environmental condition. This must be expressed with plant growth models and rules for plant production. The lack of knowledge is still a serious restriction for the use of on-line optimization in commercial holdings. The use of expert systems is a possibility to make heuristic knowledge available for the control system and to develop an "user friendly" interface. For further research activities is the development of an explanation unit, which allows to handle "uncertain" knowledge, important. Another demand is a self-learning facility.

A problem, which must be solved is the standardization of the communication link between the different levels (Fig. 1), especially between the top level and the middle level.

REFERENCES

Augustin, P., and Schmidt, M., 1980. Model of basis for optimal control of growth factors in the culture of the greenhouse cucumber. Int. Symposium on System Analysis and Simulation, 5, 420-424.

Bailey, B.J., 1985. Wind dependent control of greenhouse temperature. Acta Hort. 174, 381-386.

Bailey, B.J., 1988. Improved control strategies for greenhouse thermal screens. Acta Hort. 230, 485-492.

Bailey, B.J., Seginer, I., 1989. Optimum control of greenhouse heating. Acta Hort., 245, 512-518.

Bakker, J.C., 1985. A CO_2 control algorithm based on simulated photosynthesis and ventilation rate. Acta Hort. 174, 387-392.

Biemond, F., 1989. A glasshouse climate model as part of a bio-economic model. Acta Hort. 260, 275-284.

Bot, G.P.A., 1983. Greenhouse climate: from physical processes to a dynamic model. Thesis, Agric. Univ. Wageningen, The Netherlands.

Bot, G.P.A., 1989. Greenhouse simulation models. Acta Hort. 245, 315-325.

Challa, H., and Schapendonk, A.H.C.M., 1986. Dynamic optimization of CO_2- concentration in relation to climate control in greenhouses. In: Carbon dioxide enrichment of greenhouse crops. Florida. Vol. I, 147-160.

Challa, H., and van Straten, G., 1991. Reflections about optimal climate control in greenhouse cultivation. Proceedings of the IFAC/ISHS Workshop, Matsuyama, Japan. 13-18.

Chotai, A., Young, P.C., 1991. Self-adaptive and self-tuning control of a Nutrient Film Technique (NFT) system. Proceedings of the IFAC/ISHS Workshop, Matsuyama, Japan. 33-40

Day, W.,et al, 1991. The path to dynamic optimization of carbon dioxide in the greenhouse. Proceedings of the IFAC/ISHS Workshop, Matsuyama, Japan. 23-26.

Gal, S., Angel, A., and Seginer, I., 1984. Optimal control of greenhouse climate: Methodology. European Journal of Operational Research, North Holland. 17, 45-56.

Hack, G., 1989. The use of image processing under greenhouse conditions for growth and climate control. Acta Hort. 245, 455-461.

Harazono, Y., Kamiya, H., Yabuke, K., 1988. A control logic based on an artificial intelligence technique and its application for controlling plant environment. Acta Hort.230, 209-214.

Hashimoto, Y., Fukuyama, T., Morimoto, T., 1982. Computer aided protected cultivation (CAPC) by monitoring "Speaking Plant". Abstracts 21st Int. Hortic. Congr. Hamburg, 1982.

Hashimoto, Y., Morimoto, T., 1985. Some speaking plant approach to the synthesis of control system in the greenhouse. Acta Hort. 174, 219-226.

Hashimoto, Y., 1989. Recent strategies of ptimal growth regulations by the speaking plant concept. Acta Hort. 260, 115-122.

Henten, E.J. van, 1989. Model based design of optimal multivariable control systems. Acta Hort. 248, 301- 306.

Henten, E.J. van, Bontsema, J.,1991. Optimal control of greenhouse climate. Proceedings of the IFAC/ISHS Workshop, Matsuyama, Japan. 27-32

Heuvelink, E., Challa, H., 1989. Dynamic optimization of artificial lighting in greenhouses. Acta Hort. 260, 401-412.

Hooper, A.W., Davis, P.F., 1985. Control of greenhouse air temperature with an adaptive control algorithm. Acta Hort. 174, 407-412.

Jacobsen, B.K., Jones, J.W., Jones, P., 1987. Tomato greenhouse environment controller: real-time, expert system supervisor. American Society of Agricultural Engineers, St. Joseph, Paper No. 87-5022.

Jones, P., Jacobsen, B.K., Jones, J.W., 1988. Applying expertsystem concepts to real-time greenhouse control. Acta Hort. 230, 201-208.

Kozai, T., 1985. Ideas of greenhouse climate control based on knowledge engineering techniques. Acta Horticulturae 174, 365-373.

Kurata, K., 1988. Greenhouse control by machine learning. Acta Hort. 230, 195-200

Leung, C.M., Wilson, D.R., Marquand, C.J., Tassou, S.A., 1985. A knowledge based system for the control of a greenhouse environment. Acta Hort. 174, 425-432.

Liebig, H.P., 1989. Growth and yield models as an aid for decision making in protected crop production control. Acta Hort. 260, 99-114.

Matsui, T., Eguchi, H., 1976/77. Computer control of plant growth by image processing. I. Mathematical representation of relations between growth and pattern area taken in inphotographs of plant. Envir. Control in Biol. 14, 1–7, 1976. II. Pattern recognition of growth in on-line system. Envir. Control in Biol. 15, 37–45, 1977.

Meyer, J., 1989. Evaluation of artificial light systems. Acta Hort. 245, 370–376.

Nederhoff, E.M., Schapendonk, A.H.C.M., 1985. Effects of environmental conditions on growth and production of cucumbers; comparison between empirical and simulation data. Acta Hortic. 174, 251–258.

Nederhoff, E.M., Gijzen, H., Vegter, J.G., Rijsdijh, A.A., 1989. Dynamic model for greenhouse crop photosynthesis. Validation by measurements and application for CO_2 optimization. Acta Hort. 260, 137–148.

Rheinisch, K., T mmer, C., Hopfgarten, S., 1984. Hierarchical on-line control algorithm for repetitive optimization with predicted environment and its application to water management problems. Syst. Anal. Model. Simul. 1, 4, 263–280.

Rheinisch, K., Arnold, E., Markert, A., Puta, H., 1989. Development of strategies for temperature and CO_2-control in the greenhouse production of cucumbers and tomatoes based on model-building and optimization. Acta Hort. 260, 67–76.

Schapendonk, A.H.C.M., Challa, H., Broekharst, P.W., Udink ten Cate, A.J. 1984. Dynamic climate control: an optimization study for earliness of cucumber production. Scienta Horticulturae 23, 137–150.

Schoch,H.-J., 1987, Influence du niveau d`alimentation hydrique sur les variations du diametre des tigres, du potentiel hydrique, de la resistance stomatique, de la transpiration et de la photosynthese de l`aubergine. Agric. and Forest Meteorology, 1987, 40, 89–104

Seginer, I., Angel, A., Gal, S., Kantz, 1986. Optimal CO_2-enrichment strategy for greenhouses: a simulation study. J. Agric. Engig. Res. 34, 285–304.

Seginer, I., 1991. Optimal greenhouse temperature trajectories for a multistate-variable tomato model. Proceedings of the IFAC/ISHS Workshop, Matsuyama, Japan. 73–80.

Takakura, T., 1989. Climate control to reduce energy inputs. Acta Hort. 245, 406–415.

Takakura, T., 1991. Greenhouse production in engineering aspects. Proceedings of the IFAC/ISHS Workshop, Matsuyama, Japan. 19–22.

Tantau, H.-J., 1985. Analysis and synthesis of climate control algorithms. Acta Hort. 174, 375-380.

Tantau, H.-J., 1985. Greenhouse climate control using mathematical models. Acta Hort. 174, 449-459.

Tantau, H.-J., 1989. Models for greenhouse climate control. Acta Hort. 245, 397-404.

Tantau,H.-J., 1990. Automatic control application in greenhouse. Reprints of 11th IFAC World Congress Pergamon Press, Oxford, Vol.11, 302-305.

Tantau, H.-J., 1991. Optimal control for plant production in greenhouses. Proceedings of the IFAC/ISHS Workshop, Matsuyama, Japan. 1-6.

Udink ten Cate, A.J., 1983. Modelling and (adaptive) control of greenhouse climates. Thesis, Agric. Univ. Wageningen.

Udink ten Cate, A.J., Vooren, J. van de, 1983. New models for greenhouse climate control. Acta Hort. 148.

Udink ten Cate, A.J., 1985. Modelling and simulation in greenhouse climate control. Acta Hort. 174, 461-468.

Veerwaaijen, P.W.T., Gieling, T.H., Meurs, W.T.M. van, 1985. Measurement and control of the climate in a greenhouse, where a heated concrete floor is used as low temperature energy source. Acta Hort. 174, 469-475.

Vogel, G., Heißner, A., Kläring, P., 1989. Ways and possibilities for automatic control. Acta Hort. 260, 25-66.

Vooren, J. van de, Steinbuch, F., Hendricks, P., 1985. A multi-level control system. Acta Hort. 174, 547-548.

Young, P.C., et al., 1991. Identification, estimation and control of glasshouse systems. Proceedings of the IFAC/ISHS Workshop, Matsuyama, Japan. 307-315.

Chapter 7

OPTIMAL GREENHOUSE TEMPERATURE TRAJECTORIES FOR A MULTI-STATE-VARIABLE TOMATO MODEL

Ido Seginer
Arnold Sher

Department of Agricultural Engineering
Technion, Haifa 32000, Israel

I. INTRODUCTION

Temperature control in greenhouses affects simultaneously several crop processes, most basic of which are photosynthesis and development. While it is normally desirable to increase photosynthesis rate, there may be situations where development must be retarded in order to meet timing requirements. Conflicts between photosynthesis and development requirements may arise at times. An optimal solution to the control problem must consider the biological and engineering properties of the crop-greenhouse system, as well as the weather and the economic environment. In principle, if a sufficiently accurate model of the system is available, and the exogenous variables (weather and prices) are known or assumed, an optimal solution can be obtained by the Pontryagin maximum principle (PMP; Pontryagin and others, 1962), or by dynamic programming (Bellman and Dreyfus, 1962). This approach involves, however, a prohibitive amount of computations when models with more than very few state variables are involved. The purpose of the present study is to explore computationally certain shortcuts, which, for the problem under consideration, seem to produce acceptable results.

Computational studies of optimal control for a greenhouse with a simple, single-state-variable crop, showed that for every set of outside conditions there is a unique control, independent of the state of the crop (Challa and van de Vooren, 1980; Challa and Schapendonk, 1986; Seginer and others, 1986; Critten, 1991). When, however, a two-state-variable (lettuce) model was considered, the optimal setpoints became also a function of the state of the crop, not just of the outside conditions (Marsh and Albright, 1991; Seginer and others, 1991). Seginer and McClendon (in preparation), who trained a neural network (NN) with optimal control results for the same two-state-variable crop model, found that, indeed, information about the state of the crop improved considerably the performance of the NN as a control-generating device.

The extension to general multi-state-variable models is not simple. Two main alternatives exit: Searching for the optimal trajectory over state-space, as with the PMP, or searching over control-space, for instance by non-linear programming methods (van Henten and Bontsema, 1991). A very simple version of the latter approach was successfully applied by Swaney and others (1983), to an irrigation problem involving a multi-state-variable soybean model. Swaney used an optimization method similar to that of Marsh and Albright (1991), which may be classified as a version of model predictive control (MPC; Garcia and others, 1989), and which will be called here sequential control search (SCS). The SCS cannot guarantee an optimal solution, since its control decisions are almost independent of future developments. Nevertheless, this rather primitive method gave surprisingly good results in the aforementioned studies. In particular, Seginer and McClendon (in preparation) showed that the control trajectory obtained by SCS was identical with the one obtained by the PMP.

The SCS apparently produces good results for multi-state-variable systems in which current control decisions are largely independent on future events. It does not, however, provide information about the relative importance of the various state variables of the system. Since this is expressed by the co-state variables, a modified PMP method was devised to supplement the information obtained by the SCS (Seginer and McClendon, in preparation). In the present study both methods were applied to a multi-state-variable tomato model (TOMGRO; Jones and others, 1991).

Most of the computations were designed to explore the behavior of a typical control trajectory and its response to different input information. Additional computations were carried out to search for the best transplanting and termination dates, to compare optimal trajectories with constant-temperature trajectories, and to study the usefulness of neural nets as control generating devices. The emphasis was on the qualitative behavior of the system.

II. METHODS

The problem was formulated as follows: A tomato crop was assumed to be growing in a capacity-less greenhouse. The crop model had over 50 state-variables, including 10 classes of leaves, stems and fruit, while the greenhouse model had none. The available control equipment consisted of a ventilation fan and a convective heater. An upper limit was placed on the capacity of the fan, while the capacity of the heater was unlimited. Weather conditions and prices were assumed to be known in advance for the whole season (deterministic problem).

An optimal daytime setpoint temperature was to be determined for each day separately. The nighttime setpoint was fixed in advance and was constant throughout the whole season. The temperature in the greenhouse was maintained at the setpoint by heating or ventilation. Ventilation was not applied during the night and sometimes was incapable of maintaining the required setpoint during daytime. The energy balance of the greenhouse did not consider evaporation explicitly. Heat was exchanged through the greenhouse glazing by convection, conduction, and radiation, all lumped into a single overall transfer coefficient. In addition, sensible heat and CO_2 were exchanged through ventilation and (constant) infiltration.

A suitable seasonal performance criterion for the problem (based on a unit floor area) was formulated:

$$J = \int_0^{t_f} k\{t\}f\{t\}dt - c_r t_f - c_h\int_0^{t_f} F\{t\}dt - c_v\int_0^{t_f} Q\{t\}dt, \qquad [1]$$

where the terms on the right are, respectively, the income from selling marketable fruit, cost of rent, cost of heating and cost of ventilation. Here t is time and t_f is termination time; f is the rate at which marketable fruit is produced; F is heating flux and Q is ventilation flux; k, c_r, c_h and c_v are the unit prices of fruit, rent, heat and ventilation. Note that the market price of tomato may vary with time.

As already mentioned, two methods were used to obtain the required daytime setpoints: the SCS and a modified maximum principle method. Both are described by Seginer and McClendon (in preparation), and the following is a brief abstract of that description.

The SCS assumes a reasonable sequence of future disturbances (weather), as well as a reasonable future control policy. At decision time, a number of control options for the ensuing day are put to the test, by running corresponding seasonal simulations of the system. The control decision which yields the best value of the seasonal performance criterion is implemented.

The modified maximum principle method utilizes a reduced form of the Hamiltonian and a correspondingly reduced set of co-state equations. In this study the "reduced maximum principle" (RMP) utilized only two of the state variables, namely number of plant nodes, N, and total dry weight, W. The first is a measure of plant development, and responds mainly to temperature. The second is a measure of assimilation and responds mainly, but not solely, to solar radiation and CO_2 concentration. The rates of change of N and W are functions of the state of the crop, the weather conditions and the control fluxes F and Q. The form of the reduced Hamiltonian is

$$H\{t\} = k\{t\}f\{t\} - c_r - c_h F\{t\} - c_v Q\{t\} + p_N\{t\}(dN/dt) + p_W\{t\}(dW/dt) \qquad [2]$$

where $p_N\{t\}$ and $p_W\{t\}$ are the co-state variables of N and W.

The numerical solution started with initial values of the state variables corresponding to transplanted seedlings. A guess of the initial co-state values $p_N\{0\}$ and $p_W\{0\}$ made it possible to search for the control fluxes which maximize the Hamiltonian (Eq. [2]). The state and co-state variables were then advanced by the state equations (the complete TOMGRO model) and the co-state equations for p_N and p_W. The latter step required numerical evaluations of $\partial H/\partial N$ and $\partial H/\partial W$. The whole cycle was repeated until termination time has been reached, at which point the performance criterion (Eq. [1]) could be evaluated. An outer computational loop, searching for the pair $[p_N\{0\}, p_W\{0\}]$ which maximizes the performance criterion, completed the program. Note that the performance criterion is still based on a complete simulation of the system behavior. The approximation is restricted to the control decision mechanism.

The neural net computation were carried out with a commercially available program (NeuroShell™, Ward System Group, Inc.). It was used first to train the NN with optimal results obtained by the SCS method, and later to test its performance as a control generating rule over a set of validation data, which were new to the NN.

III. RESULTS

The results can be divided into three unequal parts. The first, and largest, explores computationally some of the intrinsic properties of the system, such as sensitivity to assumptions about the future and relative importance of state variables. The second part, reported in the penultimate sub-section, gives examples of applications: finding best transplanting and termination dates, and comparing optimal control with constant-setpoint control. The final subsection shows the performance of trained NNs as control generating devices.

REFERENCE CASE

The somewhat arbitrary reference case considered a tomato crop transplanted on 1 January 1963 and grown for the next 240 days under Ohio (USA) weather. All prices, including the price of tomato, were fixed throughout the season as follows: $c_h = 0.02$ \$/MJ; $c_v = 2\times10^{-6}$ \$/m^3; $k = 2$ \$/kg. Expenditure for rent, which is proportional to the length of the growing season, was not charged, but can be properly corrected for whenever t_f is given in advance. A coarse setpoint scale was considered: 10 to 35°C with 5K intervals.

The reference solution to this problem was obtained with the SCS. Future temperature setpoints were taken to be constant at 30°C, since this was the temperature which produced the highest yield. It was also found to be the economically best constant setpoint for the reference situation.

The mean daytime outside temperature and the total solar radiation for each day are shown at the top of Figure 1. Below these, the daytime temperature setpoints for the reference case, as obtained by the SCS, are shown (Figure 1[a]). The results show a transition from low setpoints (mostly 15°C) during the first 50 days, to high setpoints (mostly 30°C) during the rest of the season. The excursions of the setpoints on individual days (upward from 15°C and downward from 30°C), correlate positively with exceptional solar radiation levels (not shown).

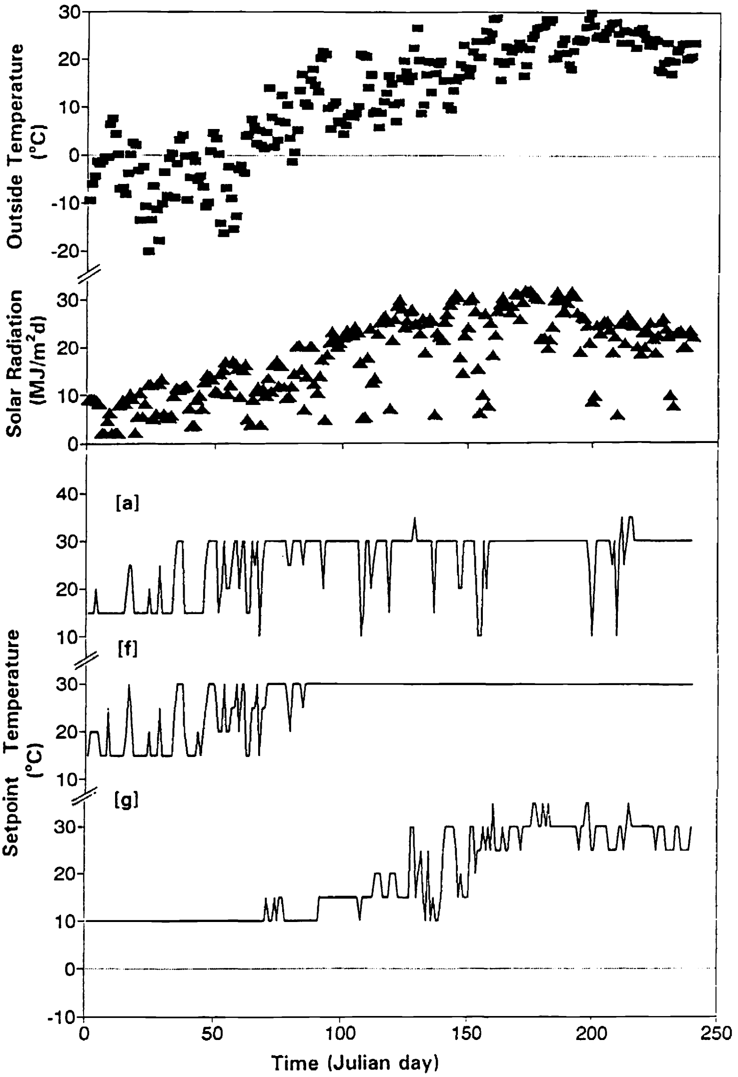


Figure 1. Weather conditions and calculated optimal daytime setpoints for greenhouse tomato under 1963 Ohio weather. Top to bottom: Mean daytime outside temperature; Total daily solar radiation; [a] Optimal daytime setpoints with SCS; [f] RMP setpoints obtained with p_N; [g] RMP setpoints obtained with p_W.

The total marketable yield, total heating energy and total ventilation volume, as well as the resulting performance criterion, are shown in Table 1A, Line [a] (Columns 3 to 6). Deviations from the relevant reference performance criterion are also shown in the table (Column 7). It should be mentioned that the results do not reflect the actual performance of spring tomato crop in Ohio. There are important differences between the true and simulated cases in terms of variety, schedule, CO_2 enrichment (none in the simulation) and prices.

The results of a variation on the reference case, with a finer setpoint scale (10°C to 36°C at 2K intervals) are presented in Table 1A, Line [b]. There is an improvement of over 0.4 $\$/m^2$ in the performance of the system, but there is no qualitative change of the setpoint sequence (not shown). As a result, and in order to economize on computer time, the reference case was **not** replaced by case [b].

SENSITIVITY TO ASSUMPTIONS ABOUT THE FUTURE

The SCS makes at each decision time assumptions about the future values of the control actions, the weather and the prices. The sensitivity of the solution to the assumed future control sequence was examined first, by setting it to a constant 10°C, rather than to the 30°C level of the reference case. The resulting setpoint sequence (not shown) and the performance criterion (Table 1B, Line [c]) are only marginally different from those of the reference case.

Along the same vein, instead of an arbitrary initial setpoint sequence, the solution of the **reference** case could be used as the initial sequence in a second iteration. As Table 1B, Line [d] shows, the improvement obtained in this manner was also marginal.

Having shown that the solution is hardly sensitive to what the control might be in the future, the effect of assumed future weather was explored next. Run [e] was started with the weather sequence of 1962, which was replaced day-by-simulation-day with that of 1963 (standard sequence). This process is an emulation of a real on-line operation with a perfect 24 hour forecast. The results in Table 1B Line [e] show again that precise knowledge of only the immediate future is required to make correct control decisions.

It should be clear that ridiculous assumptions about the control or weather of the far future may lead to a wrong solution. For example, if future weather is absurdly extreme (say no solar radiation), the current SCS decision would be to cut the losses to a

Table 1 Summary of results

Run	1 transp date Julian day	2 season length days	3 total yield $\mathrm{kg/m^2}$	4 heating energy $\mathrm{MJ/m^2}$	5 vent volume $\mathrm{k(m^3)/m^2}$	6 perform criterion $\mathrm{\$/m^2}$	7 relative loss* $\mathrm{\$/m^2}$
A reference							
[a] reference	1	240	19.00	1019	219	17.173	0.000
[b] variation	1	240	19.02	1000	211	17.606	-0.433
B sensitivity to future							
[c] initially 10°C	1	240	18.91	1012	224	17.133	0.040
[d] 2nd iteration	1	240	19.10	1029	217	17.196	-0.023
[e] 1962 weather	1	240	18.97	1018	216	17.152	0.021
C RMP versus SCS							
[f] p_N	1	240	18.99	1046	187	16.682	0.491
[g] p_W	1	240	12.99	790	333	9.507	7.666
[h] p_N and p_W	1	240	18.94	1038	186	16.738	0.435
D Effect of state (uniform weather)							
[i] optimal		240	16.68	911	63	15.018	0.000
[j] constant 30°C		240	17.46	1061	36	13.611	1.407
E Various price sequences; standard timing							
[k] const price, opt.	1	240	19.00	1019	219	17.173	0.000
[l] " " , 30°C	1	240	20.41	1264	186	15.160	2.013
[m] high to low, opt.	1	240	19.49	1208	265	24.113	0.000
[n] " " , 30°C	1	240	20.41	1264	186	22.883	1.230
[o] low to high, opt.	1	240	12.22	824	305	19.312	0.000
[p] " " , 30°C	1	240	20.41	1264	186	7.438	11.874
F Various price sequences; best timing							
[q] const price, opt.	71	235	22.22	352	279	36.832	0.000
[r] " " , 30°C	71	235	22.31	389	239	36.366	0.466
[s] high to low, opt.	5	175	12.11	1146	102	24.612	0.000
[t] " " , 30°C	5	175	12.04	1164	87	23.924	0.688
[u] low to high, opt.	101	265	21.08	700	270	66.575	0.000
[v] " " , 30°C	101	265	21.58	900	233	64.638	1.937

* negative numbers in Col. 7 represent gain relative to relevant reference.

minimum by never heating above the required minimum (10°C).

The effect of wrong assumptions about future prices (of tomato) will be considered in the penultimate subsection.

COMPARISON OF THE SCS WITH THE RMP

Based on previous experience (Seginer and McClendon, in preparation) and the tests just mentioned, it is likely that the SCS solution is in fact close to the global optimal solution. The next question, of how well does the RMP method do in comparison, could now be raised. Three runs were conducted to explore this question: In the first, [f], only p_N was retained in the Hamiltonian (Eq. [2]), in the second, [g], - only p_W, and in the third, [h], - both p_N and p_W.

The results, Table 1C, show the expediency of retaining p_N rather than p_W in the Hamiltonian. Compared to the reference solution, retaining just p_N (out of the 50 odd co-state variables of the system) resulted in a loss of less than 0.5 $\$/m^2$ per season. Retaining only p_W resulted in a severe loss - over 7.6 $\$/m^2$. Adding p_W as a second co-state variable to p_N, resulted in a negligible improvement. Other choices of co-state variables were not tried.

Figures 1[f] and 1[g] show the setpoints obtained with p_N and p_W ,respectively. Sequence [f] is similar to the reference solution [a], except that it clings to 30°C from about day 90 and on. Sequence [g] is quite different from [a] and [f], particularly in that the transition from low to high temperatures is postponed by roughly 80 days. This proved to be a bad strategy.

Figure 2 shows the trajectories $p_N\{t\}$ and $p_W\{t\}$ for Run [h]. The value of p_N is practically constant with time, while $p_W\{t\}$ first rises slightly and then dips considerably toward the end. This may be interpreted as indicating that promoting plant development is equally important throughout the season, while the production of dry matter becomes less important toward its end. This behavior may be an artifact of the truncated Hamiltonian, but probably not, since it makes sense: During the first part of the season promotion of development is presumably required in order to initiate production as early as possible, while toward the end of the season its main purpose is to ripen already formed green fruit. In contrast, there is not much point in creating new dry matter toward the end of the season, when most of it may not mature in time for marketing.

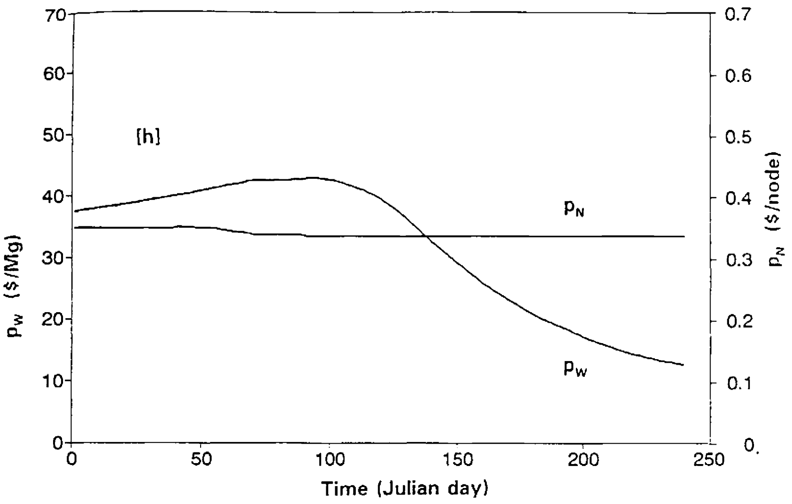


Figure 2. Evolution of the co-state variables p_N and p_W with time along the season.

EFFECT OF THE STATE OF THE CROP

All the results to this point were for essentially the same problem, and they all show how the chosen setpoints increase from the beginning of the season to its end. It is not clear, however, if this change is mainly related to the transition from winter to summer or to the changing state of the crop. To explore this question, an artificially uniform weather sequence was created, repeating a daily cycle typical of April (solar radiation 20.4 MJ/(m^2d) and temperature range 1.7°C to 10.6°C). The setpoint results as obtained by the SCS are shown in Figure 3. The season starts and ends with high (30°C) setpoints (presumably promoting earliness in the beginning of the season and fruit maturity toward the end), while around day 175 there is a pronounced minimum (15°C). Line [i] of Table 1D summarizes the results for this case in comparison with the results of a constant 30°C setpoint on Line [j]. A loss of about 1.4 $/$m^2$ is predicted as a result of maintaining a constant 30°C. This shows that the optimal setpoint trajectory depends to a measurable degree on the changing state of the crop.

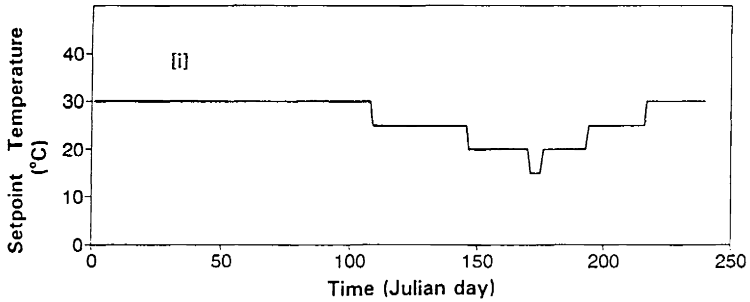


Figure 3. Optimal setpoints for a uniform weather sequence.

BEST TIMING OF GROWING SEASON

Given a particular climate and economic environment, one expects that there should be an optimal transplanting time and an optimal termination time. To explore this problem, three artificial tomato-price sequences were considered:

(1) Constant price at 2 $/kg throughout the year (as was assumed for the reference case).

(2) Price of 4 $/kg for the first 180 days of the year and zero for the rest of the year (high to low in Tables 1E and 1F).

(3) Zero value for the first 180 days and 4 $/kg for the rest of the year (low to high).

First, the optimal solution was calculated for the standard timing: transplanting on Julian day 1 and growing for 240 days. The results for the three price sequences are compared in Table 1E, lines [k] (which is a copy of [a]), [m] and [o], with corresponding results for a constant 30°C setpoint, on lines [l], [n] and [p]. The optimal solutions for the three price sequences show an economic performance of the same order. The optimal solution is always significantly better than its constant 30°C counterpart, and remarkably better for price sequence (3). Note that the **yield** is always higher for the constant 30°C regime.

Figure 4 shows the optimal setpoints and the resulting accumulation of marketable fruit for price sequence (3), Run [o], in comparison with the results for the reference solution [k] (≡[a]). It is clear that the control attempted to minimize the expenditure on heating during the first 110 days, and by doing so it effectively delayed the development of the crop. Marketable fruit only starts to appear on day 140, about 35 days later than for the constant-price case [k]. Marketing starts on Julian day 180 (vertical segment in the figure), when 4 $/kg can finally be fetched for the fruit (note that the **rate** of production has just reached its maximum about day 180). This suggests that a better performance could be obtained by postponing the transplanting date, and by moving the termination date, perhaps to the end of the year, when the price of tomato drops back to zero.

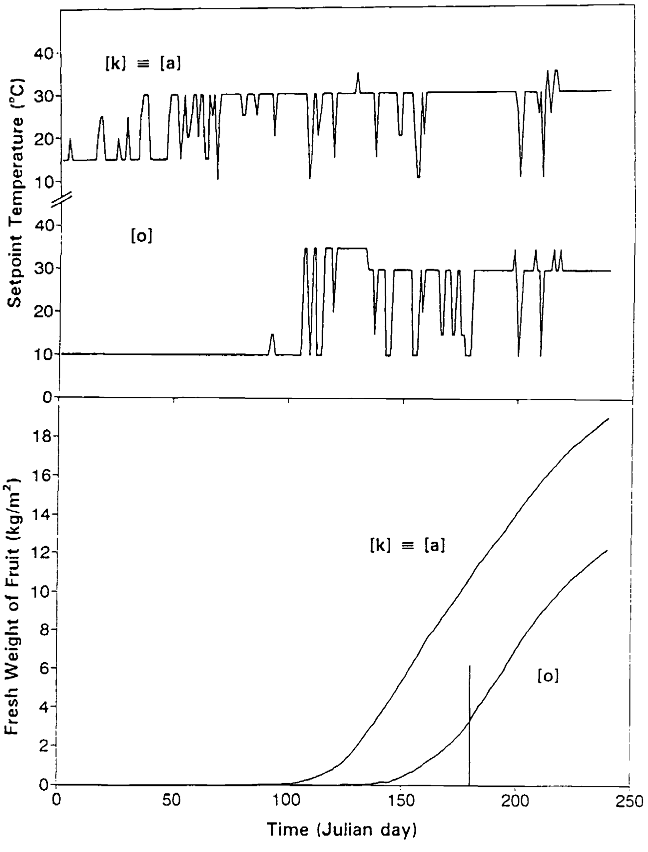


Figure 4. Optimal setpoints and resulting accumulation of mature fruit for two price sequences. [k]≡[a] uniform price at 2 $/kg; [o] zero price up to julian day 180 and 4 $/kg from then on.

The best transplanting and termination dates (±2.5 days) for the three price sequences and a constant 30°C regime (Lines [r], [t] and [v] in Table 1F) were obtained by trial and error. The improvement over the standard timing is relatively small for price sequence (2), but otherwise it is very impressive. The improvement for sequence (1) (Line [r] compared to Line [l]) is due to shifting of the production period into the summer, resulting in a higher yield and lower expenses for energy. The small improvement for sequence (2) (Line [t] compared to Line [n]) is mainly due to termination on day 180, when the price of tomato drops to zero. As a result, the greenhouse is only occupied for 175 days. The improvement for sequence (3) (Line [v] compared to Line [p]) is dramatic and is a consequence of shifting and expanding the growing season, as suggested in the previous paragraph. Note that finding the optimal timing requires information about the cost of rent (here c_r=0).

The optimal solutions for the new timing are summarized in Table 1F, Lines [q], [s] and [u]. The improvement over the constant temperature regime, Lines [r], [t] and [v], is now by an order of magnitude smaller than observed for the standard timing. This suggests that while proper timing produces much better results than arbitrary timing, the gain by optimization is less pronounced when the proper timing is selected. In other words, adjusting the temperature setpoint to the day to day variation of weather has only a minor effect on the performance. Note that this may not be the case with CO_2 control.

NEURAL-NET RESULTS

The reference SCS solution (1963) was used to train several neural nets. The nets differed in terms of the inputs they accepted and the number of nodes in the (single) hidden layer (not to be confused with plant nodes, where leaves emerge from the stem). The inputs were of two kinds:

State inputs :	N	number of plant nodes,
	L	dry weight of leaves,
	M	fresh mass of mature fruit,
	D	time from transplanting,
Weather inputs:	T	daytime mean outside temperature,
	S	daily solar radiation.

All six data elements were readily available in the outputs of the SCS program. The weather inputs represent a perfect weather forecast for the ensuing day. Note that, strictly, D is not a state variable of the system; it is, however, readily available in practice and is, presumably, well correlated with the state of the crop.

Three NN configurations were tried:

NN1, trained with D alone as input, and with 2 hidden nodes.
NN3, trained with D,T and S as input, and with 4 hidden nodes.
NN6, trained with D,T,S,N,L and M as input, and with 5 hidden nodes.

The NNs trained in this manner are only expected to produce useful setpoints for Ohio weather, for transplanting on 1 January and for tomato price of k = 2 $/kg throughout the year (reference case).

Figure 5 compares the sequence of optimal setpoints for 1963 with the values generated with NN6 and NN1 (trained with the data of the same year). The graph shows that NN6 somewhat underestimates the setpoints during the first two months and overestimates them later on. In most cases the peaks and troughs are well determined. Note that the NN-generated setpoints are not restricted to multiples of 5K. As to be expected, NN1 is only capable of fitting a smooth curve through the optimal trajectory.

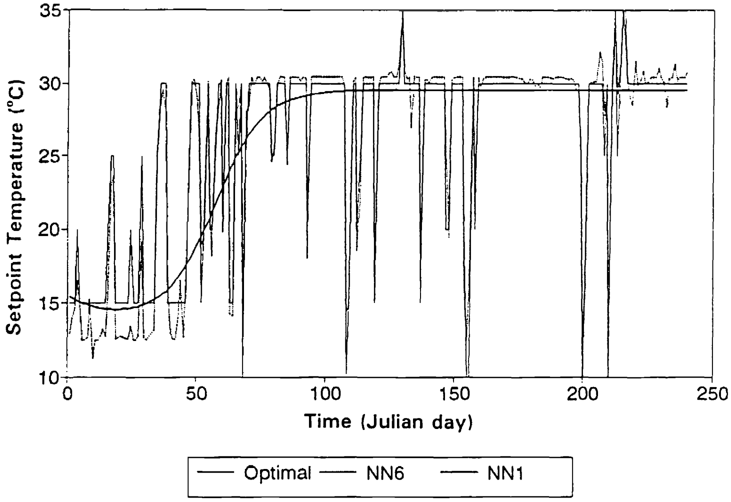


Figure 5. Comparison of optimal setpoints (fine line) with setpoints produced by neural nets NN6 (dotted line) and NN1 (heavy solid line).

The training of NN3, which, like NN6, had weather input, produced a fit (not shown) only slightly inferior to that of NN6. The main difference was that the 35°C peaks were not simulated with NN3.

The trained NNs were used to generate temperature setpoints for 15 years: 1964 to 1978 (N,L and M were taken from the optimal solutions). These were then used as input to the simulation model, resulting in performance criterion values which were compared with the optimal results for those years.

The performance criteria obtained with the NN-generated setpoints for 1963 (arrow), as well as for each of the 15 test years, are shown in Figure 6. The SCS generated setpoint sequence always produced the best performance criterion. The success of the NNs clearly increases as the number of inputs increases.

Over the 15 years, the performance criteria obtained with NN6 are, on the average, only 0.45 $/m^2 lower than optimal. The loss increases to 0.71 $/m^2 per year with NN3 and to 1.45 $/m^2 per year with NN1.

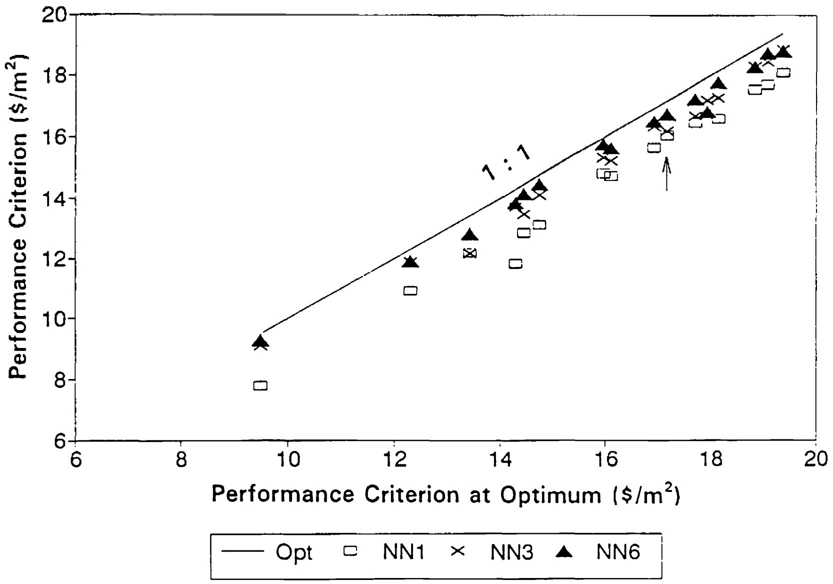


Figure 6. Performance criteria obtained with NN1, NN3 and NN6, compared with the optimal performance over 15 years. The NNs were trained with the optimal solution for 1963 (arrow).

IV. DISCUSSION AND CONCLUSIONS

Optimal control of the greenhouse climate is often viewed as having a hierarchical structure, where the control rule (law, policy) is obtained off-line, perhaps once a year. This rule is then invoked on-line at frequent intervals, to generate specific control decisions. The present study deals mainly with the first stage of developing a control rule, namely finding optimal solutions to historic control problems. It also touches on the second stage, in which these solutions are used to train NNs, so that they can serve as on-line control generators.

This study is merely a preliminary exploration into the properties of multi-state-variable crop models from the point of view of greenhouse environmental control. Nevertheless, some tentative conclusions may be drawn from this limited evidence.

1. The SCS apparently could find optimal solutions under wrong (but not absurdly wrong) control and weather assumptions beyond the immediate future (Table 1B). This indicates that the control decisions of this and similar problems may be based on the current state of the system and on simple, short term, weather forecasts. Any simplified rule, such as a neural net, designed to emulate the optimal control process, will also only require current input.

If the system includes an important operational storage, such as soil water capacity, or temperature integration capability of the crop, decisions may become dependent on the assumptions made about the future. The characteristic of an operational storage is that it is limited. As a result, its inclusion in a system leads to an optimization problem with constraints on the state variable representing the storage (e.g., Bailey and Seginer, 1989). Intuitively, if the characteristic time of the operational storage is of the order of several days, a weather forecast for that duration will be required. In addition, the optimal control sequence for that whole period will have to be obtained at each decision time. Israeli and Lambert (1986) applied a similar approach to irrigation scheduling.

2. Unlike assumptions about future control and weather, assumptions about future prices have had a significant effect on the control decisions in this study (Table 1E; Figure 4). The dramatic effect is, however, the result of comparing very different price sequences. The price difference from one season to the next is in reality not nearly as drastic. Therefore, one may expect to manage

quite well with last year's prices as a guide.

Figure 4 shows that in Run [o] heating started on Julian day 105, 75 days in advance of the price step. This is more than the time required for fruits to mature (30 to 50 days; not shown). The extra time was required to reach day 180 not just with mature fruit but with the maximum possible rate of production. This suggests that an initially unexpected steep rise in tomato price should be known almost two months in advance in order to take full advantage of it. On the other hand, if prices drop suddenly and are expected not to rise again (Run [m]), heating can be stopped on a short notice.

3. Experimentation with the RMP method (Table 1C, Figure 2) showed that a properly chosen single state variable is capable of producing a satisfactory control trajectory. The choice of state variables for inclusion in the reduced Hamiltonian differs between problems. The single state-variable formulation of Seginer and others (1986) resulted in a constant value for the co-state variable, independent of the state of the crop. (This constant value may be regarded as an expression of the intensity of cultivation (Seginer, 1989)). The study of a two state-variable lettuce model (Seginer and others, 1991), showed total dry matter to be dominant, its importance growing steadily towards harvest time. In the present study N (and not dry matter, W) turned out to be a very dominant state variable (Table 1C). The value of its co-state variable, p_N, is, however, roughly constant throughout the season (Figure 2), which means that, to a first approximation, the control is independent of the state of the crop. The change of control along the season (Figure 1[f]) is, therefore, due to weather changes along the season (similar to the single-state-variable crop mentioned above). Note that the behavior of co-state variables may depend not only on the basic system equations, but also on the control available in the greenhouse, on the general nature of the disturbances (weather and prices) and on the performance criterion.

4. Optimal timing of the season is a critical factor, because it considerably affects the performance criterion, even when the price of produce is constant throughout the year (Run [q] compared with Run [a]=[k]). In practice, however, the optimal timing requires only minor adjustments from one year to the next. Since TOMGRO tends to produce maximum yield at a constant 30°C daytime setpoint, optimal timing tends to match the growing season with the period when pushing for maximum yield is justified. As a result, when

timing is optimal, further gain from day by day optimization is not large (Table 1F). Day by day optimization may be more critical whenever CO_2 enrichment is considered.

5. Neural nets may be trained to serve as black-box control rules (strategies). They have been demonstrated previously (Seginer and McClendon, in preparation) and now, again, to emulate usefully optimal decisions. In the present study, neural nets, trained with just a single (whole season) optimal trajectory (1963), generated successfully optimal day by day control decisions for other years. Once trained, the on-line computing time required by the NN is minimal.

The training process revealed the following interesting point. The present version of TOMGRO aborts fruits excessively when temperatures higher than 30°C persist. As a result, the loss in terms of the performance criterion is asymmetric around 30°C: more loss is suffered by too high temperatures than by too low ones. This would suggest that the training error should not be evaluated by deviations from the desired output of the NN (i.e. temperature setpoints), but rather by the loss of performance criterion. Unfortunately, Neuroshell does not have the capability of alternative error definitions.

In summary, the SCS seems to produce close to optimal setpoint trajectories. This will probably hold true for similar problems, provided that the system contains no important operational storage, and as long as the uncertainty about the future is not substantial. Further, despite the multi-state-variable structure of present-day crop models, the RMP results suggest that they may be replaced by much simpler models for control purposes. The equivalent models may contain just a few properly selected state variables, preferably measurable under commercial conditions. They may be explicitly determined or implicitly included in a control rule trained with optimal solutions based on the complete model. In the latter approach, the argument list of the control rule would include the selected state variables together with parameters describing the impending weather and the expected prices.

REFERENCES

Bailey, B.J. and Seginer, I. (1989) Optimum control of greenhouse heating. Acta Hort. 245:512–518

Bellman, R.E. and Dreyfus, S.E. (1962) Applied Dynamic Programming. Princeton University Press.

Challa, H. and van de Vooren, J. (1980) A strategy for climate control in greenhouses in early winter production. Acta Hort. 106:159–164.

Challa, H. and Schapendonk, A.H.C.M. (1986) Dynamic optimalization of CO_2 concentration in relation to climate control in greenhouses. Chapter 9 in: Carbon dioxide enrichment of greenhouse crops, H.Z. Enoch and B.A. Kimball, Editors, Vol. I:147–160, CRC Press.

Critten, D. (1991) Optimization of CO_2 concentration in greenhouses: A model analysis for the lettuce crop. J. Agric. Eng. Res. in press.

Garcia, C.E., Prett, D.M. and Morari, M. (1989) Model Predictive Control: Theory and practice – a survey. Automatica 25(3):335–348.

Israeli, I. and Lambert, J. (1986) Irrigation scheduling of movable irrigation systems to maximize return. ASAE Paper No. 86-2073, 33 p.

Jones, J.W., Dayan, E., Allen, L.H., van Keulen, H. and Challa H. (1991) A dynamic tomato growth and yield model (TOMGRO). Trans. ASAE 34(2):663–672.

Marsh, L.S. and Albright, L.D. (1991) Economically optimum day temperatures for greenhouse hydroponic lettuce production. Part I: A computer model. Trans. ASAE 34(2):550–556

Marsh, L.S. and Albright, L.D. (1991) Economically optimum day temperatures for greenhouse hydroponic lettuce production. Part II: Results and simulations. Trans. ASAE 34(2):557–562.

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V. and Mishchenko, E.F. (1962) The mathematical theory of optimal processes. John Wiley Interscience. 360 pp.

Seginer, I., Angel, A., Gal, S. and Kantz D. (1986). Optimal CO_2 enrichment strategy for greenhouses. A simulation study. J. Agric. Eng. Res. 34:285–304.

Seginer, I. (1989) Optimal greenhouse production under economic constraints. Agricultural Systems 29:67–80.

Seginer, I., Shina, G., Albright, L.D. and Marsh, L.S. (1991) Optimal temperature setpoints for greenhouse lettuce. J. Agric.

Eng. Res. In press.

Seginer, I. and McClendon, R. (1992) Methods for the optimal control of the greenhouse environment. In press.

Swaney, D.P., Mishoe, J.W., Jones, J.W. and Boggess, W.G. (1983) Using crop models for management: Impact of weather characteristics on irrigation decisions in soybeans. Transactions of the ASAE 26(6):1808–1814.

van Henten, E.J. and Bontsema, J. (1991) Optimal control of greenhouse climate. in: IFAC Workshop on Mathematical and Control Applications in Agriculture and Horticulture. Y. Hashimoto and W. Day, Eds. IFAC Workshop Series, No. 1, pp. 27–32.

NOTATION

symbols		SI units
c_h	unit price of heating energy	\$/J
c_r	unit price of rent	$\$/(m^2[soil]s)$
c_v	unit price of ventilation	$\$/(m^3)$
D	time from transplanting	s
F	heat flux from heater	$W/m^2[soil]$
f	rate at which marketable fruit is produced	$kg/(m^2[soil]s)$
H	Hamiltonian	$\$/(m^2[soil]s)$
J	seasonal performance criterion	$\$/m^2[soil]$
k	unit market price for tomato	\$/kg[f.w.]
L	dry weight of leaves	$kg[d.w.]/m^2[soil]$
M	fresh mass of mature fruit	$kg[f.w.]/m^2[soil]$
N	number of plant nodes	$[node]/m^2[soil]$
p_N	co-state variable of N	\$/[node]
p_W	co-state variable for W	\$/kg[d.m.]
Q	ventilation rate	$m^3/(m^2[soil]s)$
S	solar radiation received during one day	$J/m^2[soil]$
T	daytime mean outside temperature	K
t	time	s
t_f	termination time	s
W	total dry matter	$kg[d.m.]/m^2$

SECTION 3

Computer Applications and Artificial Intelligence

Chapter 8

COMPUTER INTEGRATED SYSTEM FOR THE CULTIVATING PROCESS IN AGRICULTURE AND HORTICULTURE

- Approach to "Intelligent Plant Factory" -

Y. Hashimoto

Department of Biomechanical Systems
Ehime University
Tarumi, Matsuyama, 790, Japan

I. INTRODUCTION

Computers have been introduced to almost all industries. Computer network composed of both workstation and several usual personal computers may amplify the probable ability of the computer. Now, we could see the highly performed systems with computer network, which are called as "computer integrated system". From the view of greenhouse, I would like to discuss the application of the computer integrated system.

It seems very difficult to control the environment factors strictly in solar greenhouse from the engineering point of view, because of the great disturbance from the time-variable sunlight. As for the environmental (namely climate) control of greenhouse, great progress in physics model and control system has been made for these ten years (Bot, 1983; Udink ten Cate, 1983; Takakura, 1991). In the environment control of greenhouse, it has been proved valid that optimal control is much more important than strictly constant control based on the setpoint (Tantau, 1990). Furthermore, wide applications of computers are introduced to the management of greenhouse production as well as the environment control (Cros and Martin-Clouaire, 1990).

On the contrary, under these controlled environment in the greenhouse, horticultural crops show the different behavior from the traditional one which have been acquired through a long history of agriculture and horticulture, cultivated in the field. Especially, it might be interesting problem that the crop response is closely related to the control strategies in greenhouse (Challa and Van Straten, 1991). Optimal conditions for the horticultural crops should be examined both from the horticultural operation (Vogel et al., 1989) based on physiological plant ecology as well as plant and cell physiology (Nonami et al., 1991a, 1991b) and from the management strategies for economic yield, resulting the decision of the strategies for the setpoints of the control in the greenhouse (Reinisch et al., 1989).

Recently, plant growth factory becomes prosperous. There are two types in the plant factory system. These are perfectly controlled system with artificial light and solar system. The former control system is similar to that in the growth chamber. On the other hand, the latter one is similar to that in the greenhouse. Comparing with the greenhouse, plant factory is not so familiar to our readers. Therefore, we would like to describe it in more detail. When plant is limited for producing vegetable only, it is sometimes called vegetable factory. If we are asked to find the origin of the plant factory, we should go back to thirty years ago when the first plant factory in the world was born in Denmark. Therefore the history is not so long. In the original system, the cultivation stage is treated as a manufacturing process. Several cultivating processes are combined into a batch process and controlled as assembling line. This made a great change in the concept of agriculture, and had an impact on the image of the traditional cultivation. Since then, several plant factories have been developed in the world. Now, the plant factory noticed in Japan (Takatsuji, 1991) has environmental control system like a large growth chamber. In this system, the emphasis is laid on the process control rather than on manufacturing process. It is the system with standardized cultivating process, with which the optimum production can be realized in the shortest time. In detail, the system can make plants growing three to five times faster than in the field cultivation by supplying the twenty four daily illumination hours with artificial lighting system, controlling the CO_2 concentration in several times of that in the natural environment, and regulating the temperature, pH, EC in the optimal state. This system is the factory with process control. Moreover, in the factory, computer, all the sensor possible for use and some robots are invested to reduce the personnel expenses and to get the maximum harvest in the shortest time. In other words, this system seems to be the factory automation in agriculture and horticulture. Now the concept and system of the plant factory are going to be introduced to a seed bed and a nursery of young plant. Although it is hardly to say that system for seed bed and nursery are not different from cultivating systems for various crops, from the process control or the system automation point of view, they may be much the same.

In such a symbolized system as plant factory including advanced greenhouse, several types of computers and its networks as well as almost all kinds of high technologies are introduced. Therefore, it might be necessary for us to examine these computer integrated system, which we proposed as "computer integrated system" in our previous paper (Hashimoto, 1991).

II. PROCESS CONTROL FOR GREENHOUSE AND PLANT FACTORY

A. Environmental Control

Environmental control in the agricultural field has been developed by greenhouse engineers. Now a lot of controlling technologies are introduced to the advanced system of greenhouse and plant factory from process industries. From the aspect of computer control of the system, the routine of the process control is established based on micro-computer (micro-processor). Control algorithm is also ready in PID algorithm based on feedback control system. Computer calculates the deviation between the set-point for the environment and the data monitored from the change of the environment in each sampling interval, and decides the manipulating signal by operating PID action on the deviation. Environmental factors including the light, temperature, relative humidity and CO_2 concentration can be divided into several sub-systems when those factors are controlled respectively. Then, several minutes of sampling interval may be sufficient enough to control the environmental factors. Furthermore, only very short time is needed for processing sampled data in PID control algorithm. Therefore, it is not difficult to combine all the sub-systems by one micro-computer based on a time sharing method. Nutrient control, however, has a big dead time. Therefore, it might be necessary that the system should be examined based on feedforward control algorithm as well as feedback one. Furthermore, EC and pH could not identify the state of nutrient solution. Therefore the ion sensors are introduced to the system (Hashimoto et al., 1988b). New method for the nutrient supply is also proposed (Kurata, 1991). As for the control of light, the intensity and spectrum of the light have strong effect on plant growth. It seems interesting that the effect of frequency involved in the on-off interval of fluorescent lamp is not meaningless (Hashimoto et al., 1987, 1988c). These technology of environmental control is now introduced to the process of micropropagation of tissue culture and nursery system (Kozai, 1991).

B. Adaptive, Optimal, and Fuzzy Control

To control the environment more flexibly, it is necessary to change the set point of the environment and the system parameters disturbed by the sudden changes in the outside. This belongs to one of the adaptive control problems. To reduce the cost to the minimum in non-biological system, the optimal strategy is needed. This can be solved with the optimal control theory. The optimal control is much more effective for biological system (Hashimoto, 1989a). Though the advanced control algorithm is rather complicated , it could be improved partly from the fundamental PID algorithm. Further, nutrient control is also examined as the example of fuzzy control(Morimoto et al., 1991a ; Morimoto and Hashimoto, 1991b).

For the development of the better control system, the system identification and the parameter estimation are inevitable (Young and Chotai, 1991). From this point of view, control system design is proposed for greenhouse system (Young and Chotai, 1991). On the other hand, fuzzy logic and AI are also necessary for the realization of better control system (Morimoto and Hashimoto, 1991b).

III. CONTROL FOR MECHANIZATION

A. Greenhouse Automation

Human labor has been used necessarily in seeding, transplanting, harvesting and spreading farm chemicals, even in the process control system. Some robots and automated mechanizations begin to take place of these human labor in the agricultural and horticultural field. In England, we can find an interesting example in commercial grower, where several types of automated mechanization substitute farmers as a pioneer of greenhouse automation.

B. Robot for Micropropagation

Furthermore, clean and micro-robot are used in the LSI (Large Scale Integrated circuit of semi-conductor) factory, which could get over the unable barrier by human labor. Circumstance of micro-propagation is closely similar to that of LSI industry. The robots for micropropagation and transplanting in tissue culture are developed (Fujiwara, 1991; Kinase and Watake, 1991; Simonton, 1991), which will greatly improve nursery industry in agriculture and horticulture.

C. Plant Factory Automation

In the near future, robots, containing intelligent robot and intelligent tele robot, will substitute the human labor in plant factory including tissue culture. With these robots and mechanization, the plant factory may be called "Plant Factory Automation" just like FA (Factory Automation) in the industrial manufacture (Hashimoto, 1989b).

IV. OPTIMIZATION BASED ON PLANT RESPONSES [SPEAKING PLANT APPROACH TO CULTIVATING PROCESS CONTROL]

In order to promote the process optimization of plant factory, beyond the mere process control without biological elements, the identification of the system in the relationship between the environmental physics and plant responses to the environmental stresses and the control concomitant with the advanced algorithm should be indispensable. In another word, it is important to improve the environmental control from the fundamental level without biological element to the advanced level including biological elements.

Optimal condition should be obtained based on the deep understanding of physiological plant ecology as well as the mathematical model of the environmental physics. Describing in more detail, the optimal cultivating condition could not be identified without measuring the plant responses, while only environmental variations are measured in the process control .

From the measurement point of view, there are many kinds of data obtained from the plant responses, from long-term responses about growth to short-term responses about initial process of photosynthesis. In biological process of cultivating crop, knowledge processing is still important rather than mathematical processing. The knowledge processing means the expert system based on the knowledge base. Therefore, in the advanced control including biological elements, its success may be at the mercy of measurement , and knowledge processing as well as system theory.

This is the concept that should be defined as the speaking plant approach (SPA) to the process (or environmental) control of plant factory or greenhouse (Hashimoto, 1979, 1985a, 1989a).

A. Approach from Cultivating Technology and Plant Responses

Increasing the quality and quantity of farm products means cultivating the variety with good inheritance and supplying an optimal environment condition. The optimal environmental condition varies with plant variety and also varies with growing stage. Furthermore, as a result of the evolution for long years, plant growth is related to seasonal change affected by the altitude in the earth. Plant have the adaptability to daily environmental change existing in anywhere. All those show that the optimal environmental condition could not be simple.

In general, agricultural products with high added-value could not be yielded with the environment only promoting the speed of the growth except when only the bio-mass is the harvesting object. Optimal environmental condition does not always mean the environment promoting plant growth, sometimes, the environmental stress that retards the growth but improves the quality is also concluded in the optimal condition. Environmental stress is necessary to the plant according to the growing period. This kind of cultivation can only be realized based on the knowledge of the expert who is praised to an excellent farmer. The knowledge of the expert has become possible to be input into the AI system (computer with artificial intelligence), but the plant stress such as water stress by withering can not be identified by the untrained people or the computer. This should be analyzed scientifically.

Responses of plants to environment are dealt scientifically in physiological ecology (Hashimoto and Nonami, 1990a). Growth is promoted by photosynthesis. The quantity of photosynthesis is evaluated with photosynthetic rate. It is mainly affected by quantum flux density and CO_2 concentration. Photosynthesis is usually measured in net CO_2 uptake with the light intensity (Parkinson and Day, 1990). Photosynthesis is carried on the existence of CO_2 in the air surrounding the plant. Plant absorbs CO_2 from the stomata distributing in the leaf. Behavior of stomata (Omasa et al., 1985) is not negligible for plant growth, because the photosynthesis will stop when the stomata are closed. Stoma is controlled by the water status in the leaf, which can be quantified by water potential (Boyer and Nonami, 1990). Quantitative analysis of stomata is carried out with transpiration and stomata conductance. Photosynthesis can not go smoothly even when the stomata are open, and the quantum flux density and CO_2 are sufficient if the photosynthate is accumulated in the leaf. Therefore, translocation in the leaf should be grasped. In Duke University, dynamics of translocation is measured by using ^{11}C, and the measurement system is a large one integrated by the computer (Strain et al., 1990). Its application to the problem discussed here would be difficult at present.

The realization of optimal environment control should be done based on the premise of grasping plant responses to the environmental stress in relation to the academic knowledge of physiological ecology, and making full use of the knowledge by farmers engaged in the actual cultivation.

B. Approach from Information and Knowledge Processing

Now, almost all kinds of data can be processed in real time by the aid of computer system. Information processing, including analog electric signal, could be easily done using computer–soft for digital signal processing. Especially, application of image processing is owing to Fourier transform , one of the most popular algorithm among digital signal processing. Image processing technique has elucidated plant responses in two dimension (Eguchi, 1990; Hashimoto et al., 1982, 1984c: Hashimoto, 1990b; Omasa et al., 1982, 1990). NMR–CT used in medical diagnosis begins to be used in physiological plant ecology (Kramer et al., 1990). This may never be considered if without fast Fourier transform. Sensor with processing ability (Yamasaki, 1990) is also expected to get plant responses much more.

Computer processing the human's knowledge and being capable to make decision in stead of digital data comes into practical use. Because this system can make decision according to the input data, it is called AI (artificial intelligence) or knowledge processing. Fuzzy and neural networks is also found effective (Hirafuji, 1991; Morimoto and Hashimoto, 1991b). These processing should be introduced into speaking plant approach.

C. Approach from System Identification and Control Engineering

Though many papers have been published about identification, modelling and control of non biological system, there are few papers about plant growth. Growth of plants or crops is closely related to photosynthesis as affected by many environmental stresses surrounding them, in which CO_2 concentration and light intensity could be thought the most important factors. Therefore, CO_2 uptake responded to environmental stresses is the first target of identification and modelling of the environmental control system. This has been partly identified based on both spectral analysis and method of least squares.

(1) Spectral Analysis

Spectral analysis is applied to photosynthesis of sunflower as affected by light intensity (Hashimoto et al., 1984a). The method can be described in brief as follows:
The light intensity is described by the signal x(t) and the net CO_2 uptake by y(t) in the time domain. In the case, we suppose that the Fourier transform of the light intensity is X(f) and the Fourier transform of the net CO_2 uptake is Y(f) in the frequency domain. Further, we assume that the relation between input and output is described by linear filtering based on g(t) or G(f) as in Fig.1.

In general, the time series of x(t) may be random, or stochastic under natural condition. Spectral analysis may be considered one of the most effective identification methods in a system where the input and the output are random variables. Actually, we may compute any kind of power spectrum by using FFT (Fast Fourier Transform) algorithm as shown in Fig.2.

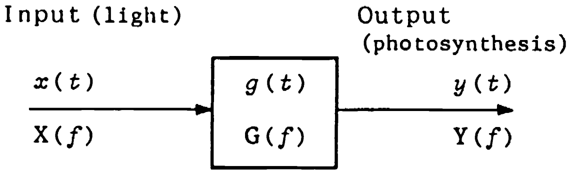


Fig.1. Input–output relation of the system. In the time domain, input is x(t), output is y(t) and g(t) of black box is impulse response. In the frequency domain, input is X(f), output is Y(f) and G(f) of black box is frequency response function.

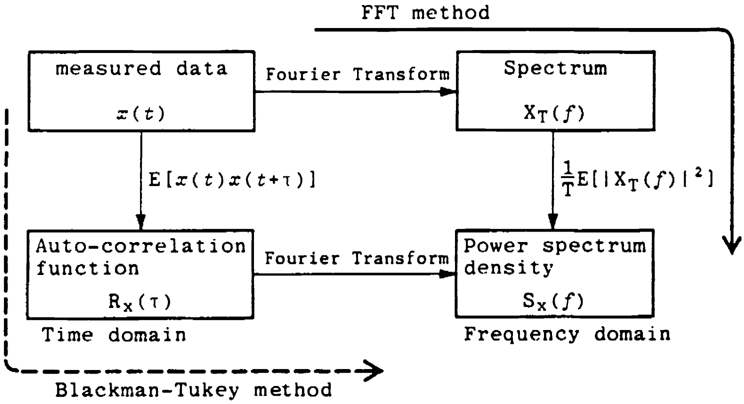


Fig.2. Flow chart for the estimation of power spectrum density. There are two methods. One is FFT (Fast Fourier transform) method and another is Blackman–Tukey's method. FFT method is indicated by the right arrow of solid line and Blackman–Tukey's method, by the left arrow of dotted line.

Finally, we can obtain G(f) from the equation given in

$$Sxy(f) = G(f) \cdot Sx(f)$$

Thus, g(t) can be calculated from G(f) by operating inverse Fourier transform, and the system is identified.

(2) Method of Least Squares based on ARMA

On the other hand, method of least squares may be explained as follows(Hashimoto and Morimoto, 1985; Torii et al., 1991). This method is based on ARMA(Auto Regressive Moving Average). Let us suppose that the input signal is described by u(t), the output signal by y(t) and the noise by v(t) respectively. We assume that v(t) and u(t) are statistically independent. We obtain the following input-output relation from Fig.3.

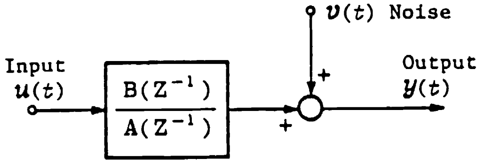


Fig.3. System configuration required for the identification.

$$y(t) = B(z^{-1})\cdot u(t)\,/A(z^{-1}) + v(t),$$

We assume that $A(z^{-1})$ and $B(z^{-1})$ are polynomials in the backward shift operator z^{-1}.

That is:

$$A(z^{-1}) = 1 + a_1z^{-1} + ... + a_nz^{-n}, \qquad B(z^{-1}) = b_0 + b_1z^{-1} + ... + b_nz^{-n}$$

Then, we obtain y(t) as follows.

$$y(t) + a_1y(t-1) + ..+ a_ny(t-n) = b_0u(t) + b_1u(t-1) + ..+ b_nu(t-n) + e(t)$$

$$\text{where} \quad e(t) = v(t) + a_1v(t-1) + ... + a_nv(t-n)$$

Suppose that each value of $a_1, a_2, ..., a_n$ is zero, thus we obtain

$$y(t) = b_0u(t) + b_1u(t-1) + ... + b_nu(t-n) + v(t)$$

This equation shows y(t) given by impulse response of u(t). As y(t) may be obtained based on time series, then y(t) is described as

$$\{ y(t),\ t = 0, 1, 2, ...N \}$$

Or, we can describe Y(t) in vector as

$$Y = [\ y(0),\ y(1),\ \ldots,\ y(N)]^T$$

Similarly, U(t) in matrix, b and V(t) in vector as

$$U = \left|\begin{array}{lll} u(0) & u(-1)\ldots\ldots & u(-n) \\ u(1) & u(0)\ \ldots\ldots & u(1-n) \\ & \ldots\ldots. & \\ u(N) & u(N-1)\ \ldots & U(N-n) \end{array}\right|$$

$$b = [\ b_0,\ b_1,\ldots,\ b_n\ ]^T, \qquad V = [\ v(0),\ v(1),\ldots,\ v(N)\ ]^T$$

Therefore, we can obtain the following equation.

$$Y = Ub + V$$

We may denote the optimal least squares by $\hat{Y} = U\,\hat{b}$.
Then,

$$\partial J/\partial\hat{b} = 0\ ,\ \text{where}\ J = \|Y - \hat{Y}\|^2$$

Thus, $\hat{b} = (U^TU)^{-1}U^T\hat{Y}$

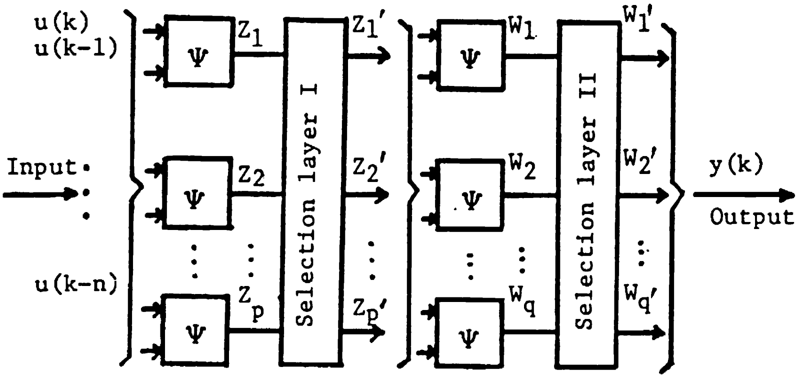
Therefore we can predict any output by using $\hat{b}$ and the system is also identified.

(3) GMDH(Group Method of Data Handling)

Dynamic model of CO_2 uptake is also identified based on GMDH (Hashimoto and Yi, 1988a). It is reported that GMDH is developed by Ivakhnenko based on the pioneering work of Wiener and Gabor for nonlinear system representations and the work of Rosenblatt for the perception concept (Toldalagi and Mehra, 1982). The method is effective for identifying the nonlinear systems of unknown structure. A schematic diagram of GMDH is shown in Fig.4.

The algorithm can be described in brief as follows:
Let y(k) denote the system output, u(k) denote the system input. Choose u(k−i) and u(k−j) arbitrarily from u(k), u(k−1),..., u(k−n),

where, n is the input variable number, and combine the two variables with a quadratic polynomial Ψ, then n! / (2(n−2)!) equations, which are called partial description, can be obtained.



Ψ :Quadratic polynomial in two variables

Fig.4. The GMDH model structure.

Denote those partial descriptions with Z_1, Z_2, ..., Z_p, and select them using a selection criterion in selection layer I. Suppose that p' equations are passed through selection layer I, and let $Z_{1'}$, $Z_{2'}$,..., $Z_{p'}$ denote these equations. Combine Z_i and Z_j with the same quadratic polynomial Ψ, then, p'! / (2(p'−2)!) partial descriptions are obtained. Denote them as W_1, W_2,...., W_q, and make a selection with the criterion in the selection layer II. Combine W_i, W_j with the same quadratic polynomial Ψ from W_1, W_2,..., W_q passed through selection layer II and select the partial descriptions again as described above. Repeat those steps until they are passed through a satisfactory criterion, then, the optimal representation of the input−output relation can be obtained.

(4) Applications of Artificial Neural Network to Identification

On the other hand, applications of neural network to the identification are examined (Morimoto and Hashimoto, 1991a ; 1991b). The method is explained as follows:
Suppose that the input−output relation are described by ARMA. Thus,

$$y(t) = - a_1y(t-1) - \dots - a_ny(t-n) + b_0u(t) + b_1u(t-1) + \dots + b_nu(t-n)$$

From this equation, it seems evident that the output y(t) is obtained from the weighted sums of the n th historical output data {y(t−1),..., y(t−n)} and (n+1) th historical input data {u(t), U(t−1),..., U(t−n)}. The structure of the neural network is built based on the equation. That is composed of three layers as shown in Fig.5.

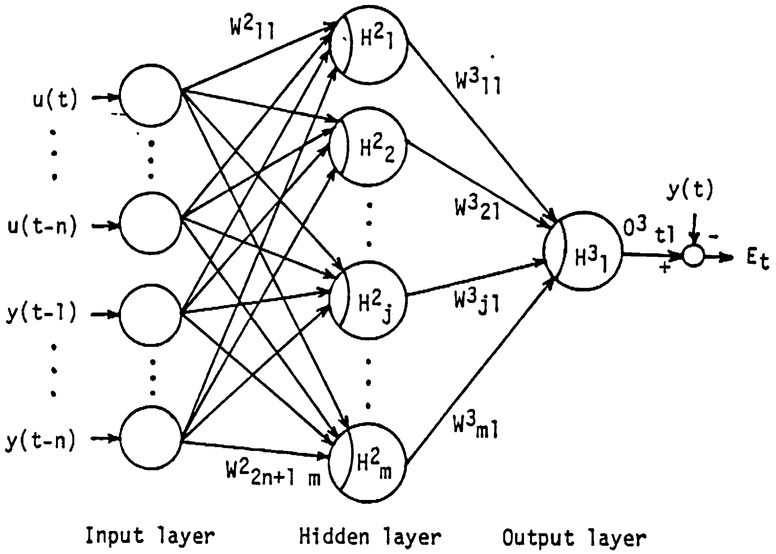


Fig.5. Structure of neural network.

These are input layer, hidden layer and output layer. The input layer is composed of (2n+1) data given as { y(t−1),..., y(t−n)} and { u(t), U(t−1),..., U(t−n) }. Therefore the input layer is thought to have neurons composed of (2n+1) data.

It may be noted that the neural network can acquire the complex dynamics such as non-linearity and time-varying parameters by its own high learning capability.

The learning procedure used here is based on the error back propagation, where the optimal weights(W_{ij}) and biases(H_i) are obtained by computing the squared error between the observed and the calculated values to its minimum value.

V. COMPUTER SUPPORT SYSTEM FOR CONTROL, HORTICULTURAL OPERATION AND MANAGEMENT

Application of computer becomes more broad (Day, 1991). Computer is introduced to decision support system in horticultural operation and management as well as to real time control system. Furthermore AI (artificial intelligence) is now used for the expert and diagnosis system.

The tomato cultivating support system, based on the program of horticultural operation, was designed and examined (Hatou et al.,, 1990, 1991b), system of which had the computer network composed of both usual micro-computer for environmental control of crops and the special computer for the artificial intelligence. The support system is found to be effective for large scale plant growth factory.

On the other hand, "ES (expert system)" is more biological and use of AI computer is inevitable. Expert system for disease and pest diagnosis for tomato was examined (Hoshi and Kozai, 1988). Expert system for control system as well as adjustment of set-point based on horticultural operation, is also examined (Hatou et al., 1991a). In the important stage through the whole cultivating term, the status of the crop is able to be discriminated based on the AI computer. Thus, the adequate set-point of the environmental control could be decided based on the status of the crop.

As the next stage, expert system for the management of plant growth factory based on some strategies may be expected, which is so called CIM in agriculture, in other words, the high levelled decision system from the point of management view. This problem should be examined in "computer integrated agricultural production" rather than in the support system (Hashimoto and Hatou, 1992).

VI. COMPUTER INTEGRATED SYSTEM

In industrial manufacture, application of robot has graded up the automation from process automation (PA) to factory automation (FA). Furthermore, the computer has been introduced into the factory management as called computer integrated manufacture (CIM). Referring the circumstance in the industries, computer integrated system in agriculture and horticulture may be proposed as shown in Fig.6 (Hashimoto, 1991) .

Fig.6 shows the recent computer integrated system. The system has the computer network composed of host computer, usual micro-computer for process control and the special computer for knowledge processing based on the artificial intelligence.

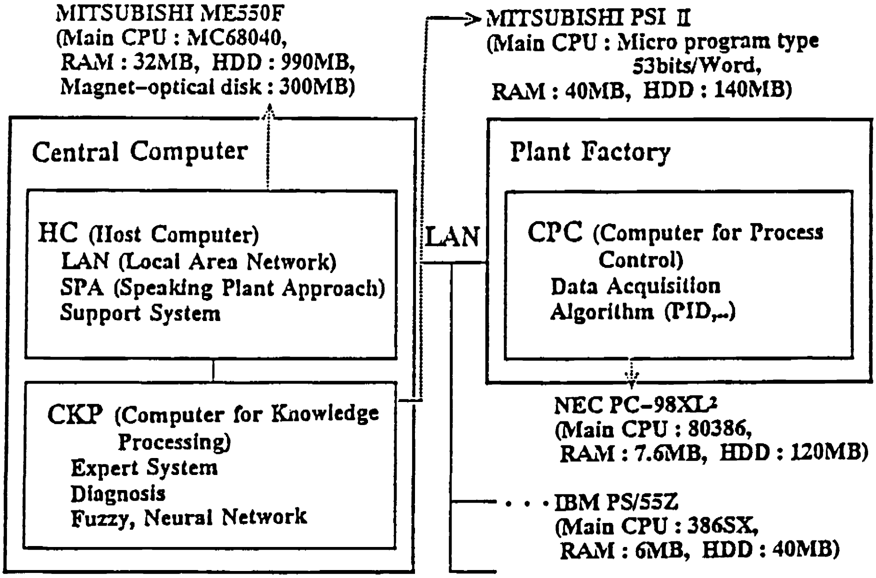


Fig. 6. The computer integrated system.

Necessary menu is chosen through the conversation between operator and computer as shown in Fig.7. Six main menus are prepared , these are "LAN", "Process Control", "SPA(speaking plant approach)", "Support System", "Knowledge Processing" and "Factory Automation" as shown in the left part of the figure. Each menu is used for the purpose described in the right part of the figure. The languages used in the main menu are "BASIC", "C" and "PROLOG".

Menu for "LAN" deals with computer communication with proper protocol. The system is programmed with about several hundred steps.

Menu for "Process Control" is always used for monitoring and real time control, depending on mainly mathematical and physical method.

On the other hand, menu for "SPA" is more complicated for the purpose of identification and optimization of total and sub-system based on physiological plant ecology.

Menu for "Support System" is prepared for the decision support of horticultural operation and management.

Menu for "Knowledge Processing" is developed for expert

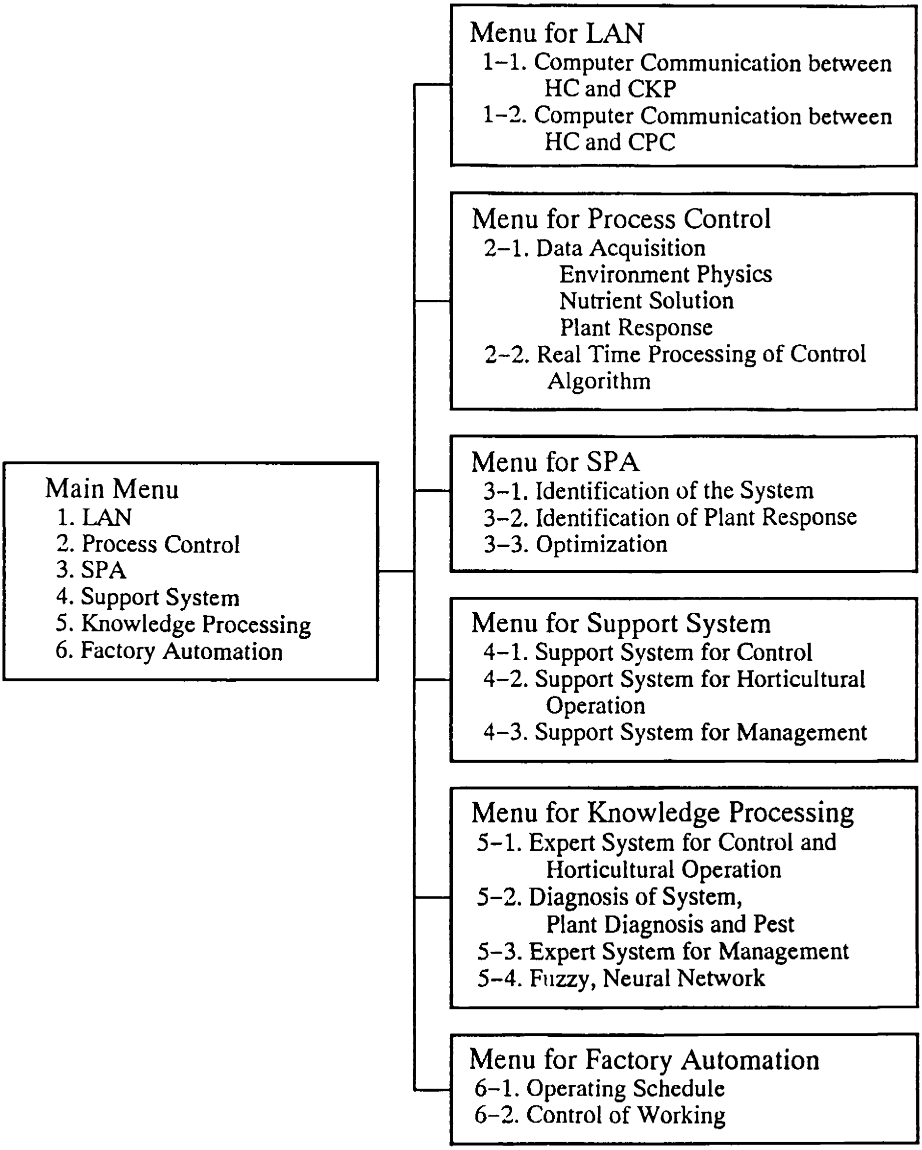


Fig.7. The menu of software for the computer integrated system.

system and diagnosis system based on the artificial intelligence. Thus, plant factory is elevated to the level of "Intelligent System".

Finally, menu for "Factory Automation" is ready for robot and mechanization system in primitive control.

VII. CONCLUSION

Plant growth factory has been applied both to cultivating process and to micropropagation process. The value of them increases enormously, if computer integrated system could be introduced into these system. As PA (process automation), FA (factory automation), and CIM (computer integrated manufacture) in chemical process industry have much improved the productivity, greenhouse and plant growth factory which are considered to be closely similar to them among agriculture and horticulture, are also expected to improve the productivity.

Computer integrated system, composed of " computer for communication and network (LAN)", "computer for environmental control", "computer for nutrient control", "computer for factory automation", "computer for identification and optimization", "computer for support system of horticultural operation and management", and , "computer for knowledge processing in expert and diagnosis system", is inevitably necessary for the industrialization of agricultural production. It seems evident that computer integrated system proposed in this paper should be the most reasonable system and important approach to " Intelligent Plant Factory".

REFERENCES

Bot, G.P.A. (1983). Greenhouse climate: from physical processes to a dynamic model. pp.240, Agricultural University, Wageningen, The Netherlands.

Boyer, J.S. and Nonami, H. (1990). Water Potential and Its Components in Growing Tissues. In Measurement Techniques in Plant Science (Hashimoto, Y., Kramer, P.J., Nonami, H. and Strain, B.R.(eds)) Academic Press, San Diego, pp.101-112

Challa, H. and Van Straten, G. (1991). Reflections about Optimal Control in Greenhouse Cultivation. In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) Pergamon Press, Oxford, pp.13–18

Cros,M.-J., and Martin-Clouaire, R. (1990). Management of Agricultural Production Systems. Proc. 3rd International Congress for Computer Technology, 263–287

Day,W. (1991). Computer Applications in Agriculture and Horticulture. In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) Pergamon Press, Oxford, pp.247–251

Eguchi, H. (1990). Digital Processing of Plant Images Selected by Spectral Characteristics of Reflectance for Evaluation of Growth. In Measurement Techniques in Plant Science (Hashimoto, Y., Kramer, P.J., Nonami, H. and Strain, B.R. (eds)) Academic Press, San Diego, pp.361–372

Fujiwara, H. (1991). Discriminating robot system for carnation seedlings with fuzzy logic. In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) Pergamon Press, Oxford, pp.231–235

Hashimoto, Y. (1979). Computer control of short term plant growth by monitoring leaf temperature. Acta Horticulturae 106, 139–146

Hashimoto, Y., Morimoto, T. and Funada, S. (1982). Image processing of Plant Information in the Relation between Leaf Temperature and Stomatal Aperture. Proc. IMEKO 9th World Congress Vol.3:313–320, North Holland Publ.Co., Amsterdam

Hashimoto, Y., Strain, B.R. and Ino, T. (1984a). Dynamic behavior of CO_2 uptake as affected by light. Oecologia 63, 159–165

Hashimoto, Y., Ino, T., Kramer, P.J., Naylor, A.W. and Strain, B.R. (1984b). Dynamic analysis of water stress of sunflower leaves by means of a thermal image processing system. Plant Physiology 76:266–269

Hashimoto, Y., Morimoto, T. and Fukuyama, T. (1985a). Some speaking plant approach to the synthesis of control system in the greenhouse. Acta Horticulturae 174, 219–226.

Hashimoto, Y. and Morimoto, T. (1985b). Identification of water relation and CO_2 uptake in physiological ecological processes in a controlled environment. Proc. IFAC 7th Symposium on Identification and System Parameter Estimation: 1677–1681, Pergamon Press, Oxford.

Hashimoto, Y., Yi, Y., Nyunoya, F., Anzai, Y., Yamazaki, H., Nakayama, S. and Ikeda, A. (1987). Vegetable growth as affected by on-off light intensity developed for vegetable factory. Acta Horticulturae 229, 259–264

Hashimoto, Y. and Yi, Y. (1988a). Dynamic model of CO_2 uptake based on system identification. Acta Horticulturae 248, 295–300

Hashimoto Y., Morimoto T., Fukuyama T., Watake H., Yamaguchi S., and Kikuchi, H. (1988b). Identification and control of hydroponic system using ion sensors. Acta Horticulturae 245, 490–497.

Hashimoto Y., and Yi, Y. (1988c). Control model of plant as affected by the pulsed light illumination. Acta Horticulturae 245, 448–454.

Hashimoto, Y. (1989a). Recent strategies of optimal growth regulation by the speaking plant concept. Acta Horticulturae 260, 115–121.

Hashimoto, Y. (1989b). Dynamical approach to plant factory', Proc. AGROTIQUE 89, 111–118.

Hashimoto Y., and Nonami, H. (1990a). Overview of current measurement techniques from aspects of plant science, In Measurement Techniques in Plant Science (Hashimoto, Y., Kramer, P.J., Nonami, H. and Strain, B.R. (eds)), Academic Press, San Diego, pp.7–24.

Hashimoto, Y. (1990b). Leaf Temperature based on Image Processing. In Measurement Techniques in Plant Science (Hashimoto, Y., Kramer, P.J., Nonami, H. and Strain, B.R. (eds)), Academic Press, San Diego, pp.372–386

Hashimoto, Y. (1991). Computer integrated plant growth factory for agriculture and horticulture, In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) , Pergamon Press, Oxford, pp.105–110.

Hashimoto, Y. and Hatou, K. (1992). Knowledge based computer integrated plant factory. Proc. IFAC Workshop on Expert System in Agriculture 9–12.

Hatou, K., Nishina, H. and Hashimoto, Y. (1990). Tomato Cultivating Support System Based on Computer Integrated Agricultural Production. Proc. 3rd International Congress for Computer Technology 100–107

Hatou, K., Kamio, Y. and Hashimoto, Y. (1991a). Computer Integrated Plant Factory based on AI. Proc. IFAC 9th Symposium on Identification and System Parameter Estimation, 616–621, Pergamon Press, Oxford.

Hatou, K., Nonami, H., Itou, M., Tanaka, I. and Hashimoto, Y. (1991b). Computer Support System for Tomato Cultivation in

Plant Factory. In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)), Pergamon Press, Oxford, pp.301-306

Hirafuji, M. (1991). A Time-continuous Quantitative Plant Growth Model Base on Neural Networks and Lotka-Volterra Equations In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) , Pergamon Press, Oxford, pp. 289-294

Hoshi, T. and Kozai, T. (1988). Disease and pest diagnosis for tomato. Proc. 2nd International Congress for Computer Technology 457-472.

Kinase, A. and Watake, H. (1991). Robot for masspropagation', In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)), Pergamon Press, Oxford, pp.225-230.

Kozai, T. (1991). Environmental Control in plant tissue culture and its Application for Micropropagation', In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) , Pergamon Press, Oxford, pp. 99-104.

Kramer, P.J., Siedow, J.N. and MacFall, J.S. (1990). Nuclear Magnetic Resonance Research on Plants. In Measurement Techniques in Plant Science(Hashimoto, Y., Kramer, P.J., Nonami, H. and Strain, B.R. (eds)), Academic Press, San Diego, pp403 `431

Kurata, K. (1991). Roots Dividing Method for Nutrient Supply in Soilless Culture, In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) , Pergamon Press, Oxford, pp. 137-140.

Morimoto, T., Cho, I. and Hashimoto, Y. (1991a). Identification of Hydroponics in an Advanced Control System of the Greenhouse, Proc. IFAC 9th Symposium on Identification and System Parameter Estimation 610-615.

Morimoto, T. and Hashimoto, H. (1991b). Application of Fuzzy Logic and Neural Network to the Process Control of Solution pH in Deep Hydroponic Culture, In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) , Pergamon Press, Oxford, pp.147-152.

Nonami, H., Hashimoto, Y. and Boyer, J.S. (1991a). Environmental Control for Plant Growth in Plant Factory Operation and Greenhouse Management from Physiological Viewpoint. In Mathematical and Control Applications in Agriculture and

Horticulture (Hashimoto, Y. and Day, W. (eds)) , Pergamon Press, Oxford, pp.125–130

Nonami, H., Fukuyama, T., Cho, I. H. and Hashimoto, Y. (1991b). Improvement of Tomato Fruit Production by Nutrient Supply Management in Greenhouses. In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W.(eds)) , Pergamon Press, Oxford, pp.413–418

Omasa, K., Aiga, I. and Hashimoto, Y. (1982). Image Instrumentation for Evaluating the Effects of Air Pollutants on Plants. Proc. IMEKO 9th World Congress Vol.3:303–312, North Holland Publ. Co., Amsterdam

Omasa, K., Hashimoto, Y., Kramer, P.J., Strain, B.R., Aiga, I. and Kondo, J. (1985). Direct Observation of Reversible and Irreversible Stomatal Responses of Attached Sunflower Leaves to SO2. Plant Physiology 79: 153–158

Omasa, K. and Shimazaki, K. (1990). Image Analysis of Chlorophyl Fluorescence in Leaves. In Measurement Techniques in Plant Science(Hashimoto, Y., Kramer, P.J., Nonami, H. and Strain, B.R. (eds)), Academic Press, San Diego, pp. 387–401

Parkinson, K.J. and Day, W. (1990). Design and Testing of Leaf Cuvettes for Use in Measuring Photosynthesis and Transpiration. In Measurement Techniques in Plant Science (Hashimoto, Y., Kramer, P.J., Nonami, H. and Strain, B.R. (eds)), Academic Press, San Diego, pp. 207–228

Reinisch, K., Arnold, E., Markert, A. and Puta, H. (1989). Development of Strategies for Temperature and CO_2 Control in the Greenhouse Production of Cucumbers and Tomatoes Based on Model building and Optimization. Acta Horticulturae 260, 67–76

Simonton, W. (1991). Automated plant handling and processing in a robotic workcell, In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) , Pergamon Press, Oxford, pp.213–218.

Strain, B.R., Goeschl, J.D., Fares, Y., Magnuson, C.E. and Jaeger, C.E. (1990). The Use of 11C in Physiological–Ecological Research, In Measurement Techniques in Plant Science (Hashimoto, Y., Kramer, P.J., Nonami, H. and Strain, B.R. (eds)), Academic Press, San Diego, pp. 265–276

Takakura, T. (1991). Greenhouse Production in Engineering Aspects, In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)) , Pergamon Press, Oxford, pp.19–21.

Takatsuji, M. (1991). Fundamental Study of Plant Factory. In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)), Pergamon Press, Oxford, pp.81-84

Tantau, H.-J. (1990). Automatic Control Application in Greenhouse. Proc. IFAC 11th World Congress, Vol.11 :302-305, Pergamon Press, Oxford.

Toldalagi, P.M. and Mehra, R.K. (1982). Application of nonlinear identification to long range weather prediction, Proc. IFAC 6th Symposium on Identification and System Parameter Estimation Vol.1, 512-520.

Torii, T., Ochiai, T., Okamoto, T. and Kitani, O. (1991). Identification of the water status of the plant-Identification of AR model, In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)), Pergamon Press, Oxford, pp.389-393.

Udink ten Cate, A.J. (1983). Modelling and (adaptive) control of greenhouse climate. pp159, Agricultural University, Wageningen, The Netherlands.

Vogel, G., Heissner, A. and Klaering, P. (199). Ways and Possibilities for Automatic Control of Plant Growth and Yield/Formation Processes and Computer Aided Soil and Crop Management in Vegetable Production. Acta Horticulturae 260: 25-66

Yamasaki, H. (1990). From Sensor Devices to Intelligent Sensing Systems. In Measurement Techniques in Plant Science (Hashimoto, Y., Kramer, P.J., Nonami, H. and Strain, B.R. (eds)), Academic Press, San Diego, pp.25-42

Young, P.C., Tych, W. and Chotai, A. (1991). Identification, Estimation and Control of Glasshouse Systems, In Mathematical and Control Applications in Agriculture and Horticulture (Hashimoto, Y. and Day, W. (eds)), Pergamon Press, Oxford, pp.307-315

Chapter 9

USING EMPIRICAL KNOWLEDGE FOR THE DETERMINATION OF CLIMATIC SETPOINTS:AN ARTIFICIAL INTELLIGENCE APPROACH

Roger Martin-Clouaire*
Thierry Boulard**
Marie-Josée CROS*
Benoit JEANNEQUIN***

*INRA, Laboratoire d'Intelligence Artificielle, BP27, 31326 Castanet, France
** INRA, Station de Bioclimatologie, 84140 Montfavet, France
*** INRA, Domaine Horticole d'Alenya, Le Mas blanc, 66200 Elne, France

I. INTRODUCTION

Farmers of the next decade must become expert managers of all aspects of their farming operations. Good management will be more intensive and more demanding in terms of time and expertise, and more critical to farm survival. Automating or at least supporting the decision making process underlying agricultural production .management is becoming an increasingly important issue in order to master the inherent complexity of the task and make a more rational use of resources. This applies in particular to greenhouse production systems because of the intricacy of the various processes involved and the manifold interactions between the phenomena pertaining to the biological characteristics of the crop and to the physical characteristics of the greenhouse. Moreover, the fast paces at which greenhouse technology, practical know-how and the range of plant varieties evolve do not let the grower acquire a robust management know-how by simple accumulation of experience.

The better management decisions that a knowledgeable and efficient computer system is capable to provide can be translated into higher yields, earlier produce and lower production costs. Rough estimations concerning the French Mediterranean context in which we are interested showed that for the same agronomic objectives (in terms of quantity, quality and timing) up to 25% of fuel energy could be saved by a more careful and wiser management of the greenhouse climate.

The work reported here concerns a computerized approach to support the tactical management of a greenhouse production system yielding tomatoes which are grown during winter and early spring time in the southern part of France. A computer program, called SERRISTE[1], has been developed for choosing automatically appropriate daily climatic setpoints to be maintained by specific devices equipping the greenhouse (heating and aeration systems).

Quantitative modeling of the whole crop and greenhouse systems is hitherto out of reach. For this reason the approach underlying the SERRISTE system rests on a rich body of empirical knowledge that would be used by a highly qualified grower in order to solve the setpoint determination problem. Simply stated, the setpoint determination problem consists in finding values that can ensure a satisfactory balance over conflicting goals and preferences such as: avoiding important damages to plants (e.g. diseases), maintaining good growth-development combination, minimizing heating-energy spending. The appropriate way to achieve these goals is predominantly dependent on the prevailing crop and weather conditions of the current day and on past and predicted conditions. Conventional daily management in greenhouses is based on predetermined sequences of setpoints (i.e. blueprint) that depend on static characteristics (e.g. latitude, equipments) of the greenhouse and the crop (e.g. variety of tomato). These setpoints vary with time, within the 24 hours of a day and along the cultivation period. Actually, they are just guidelines that must be adjusted by the growers depending on the crop conditions observed or predicted on a short term basis. Through such a tuning the growers try to perform a kind of empirical optimization of the dynamic behavior of the production system. The difficulty in doing this tuning stems from the strong interactions that exists between the effects of the actions associated with the setpoints. Moreover, the crop systems are poorly understood compared to many physical systems and, as usual with agronomical phenomena, much of our understanding is qualitative. Consequently taking the truly optimal decision in the setpoint determination problem is practically illusive.

Since quantitative approaches (e.g. optimization algorithms, simulation models) are not yet applicable, we have tackled the problem within an artificial intelligence framework that allows an explicit representation of the pertinent pieces of knowledge and a clear separation between the domain-specific know-how and its use in an intelligible reasoning process.

This paper describes the overall approach of the SERRISTE

---

[1] SERRISTE means greenhouse grower in French.

system with particular emphasis on the presentation of the body of knowledge incorporated in the system and the artificial intelligence techniques underlying the resolution process of the setpoint determination problem. The main part of the empirical know-how taken into account in the setpoint determination process is expressed through numerical constraints (relations) on the variables involved in the problem. SERRISTE solves what is called a constraint satisfaction problem (CSP) which involves the assignment of values to the variables subject to the above-mentioned constraints. Each set of assignments is considered as acceptable and is then evaluated and ranked according to preference criteria (e.g. economics). SERRISTE belongs to the class of knowledge-based systems and is essentially implemented in an object-oriented programming spirit. It runs on conventional PC-computers under the MS-DOS operating system.

The greenhouse decision and the computer frameworks in which the present work is inserted are introduced in Section II. Section III describes the fundamental elements of the body of empirical knowledge involved in the setpoint determination problem. The artificial intelligence approach of the SERRISTE system is exposed in Section IV. Preliminary evaluations of the current version and further developments are are briefly discussed in Section V.

II. GREENHOUSE DECISION AND COMPUTING FRAMEWORKS

In agreement with the widely spread decomposition (Udink ten Cate and Challa, 1984; Baille et al., 1990) of the decisions that lie between strategy formulation and its implementation, we consider a decision process structured in a tree level hierarchy going at the bottom (level 1) from on-line climate control through (level 2) the tactical decision level concerning the determination of daily setpoints and to the higher (level 3) seasonal planning. Each level is associated with a particular goal and some decisions must be taken so that these objectives are satisfied. The decision making process at a particular level determines the objectives assigned to the next lower level except at the bottom one where decisions are directly transformed into actions (Cros and Martin-Clouaire, 1990). The first level encompasses a set of regulation algorithms that aims at controlling the important (and controllable) climate factors. The automation of the short term decision process at this level is fairly well mastered (although some improvements are still possible) in modern, well-equipped greenhouses. The second level deals with deciding the appropriate daily settings depending on the actual status of the crop (growth stage, vigor), weather conditions and timing situation with respect to the overall planning specified at this level. The decisions involved concern the environment setpoints of a particular day but the reasoning needed to reach these decisions must span a several day time scale. Only little research (Challa et al., 1989; Tantau, 1989) has been devoted to the issue of automating middle level decision making. Finally, given the general goal of maximizing the profit of the grower, the upper level, that is still to be explored, deals with the appropriate decomposition of the crop production season in growth stages and the determination of the corresponding mean inside temperatures according to constraints related to mean outside climate, crop growth and development, equipment, cost of production, tomatoes pricing coditions, etc...

Again, hitherto, only the first level is fully automated. In practice, this is realized thanks to the now classical distributed processing architecture in which real-time constraints on particular greenhouse compartments are handled by dedicated control computers that are connected to a central machine. The latter is typically of the PC type and is essentially used as an information system for collecting, storing and displaying environmental data. Beside its utilization for monitoring the important parameters of the different compartments it can be used as the media to modify the settings of the control devices that are regulated by the dedicated control computers.

The higher the level in the hierarchy the less often the decisions must be taken and the less structured the decision types. Thus, it is not surprising that the only really automated level is the first one. The lack of structure of a decision type means that either all steps of the decision cannot be specified before the decision is taken, or the decision depends on qualitative (i.e. incomplete) knowledge or imprecise/uncertain data (e.g. predicted weather). Consequently, such decision types do not yield easily to formal analytical treatment by quantitative decision models.

The work reported here is an attempt to address the middle level decision process (daily setpoint determination) outside of the realm of purely quantitative approaches (e.g. optimization algorithms, simulation models) by relying on a rich core of heuristic knowledge. The knowledge-based system SERRISTE developed for this purpose has been designed to run on the central PC machine of the greenhouse computer environment.

III. BASIC KNOWLEDGE INVOLVED IN THE DETERMINATION OF CLIMATIC SETPOINTS

This section presents an analysis of the basic body of practical knowledge to take into account and how to use it wisely in the determination of daily climatic setpoints. The empirical knowledge we want to articulate pertains to modern greenhouses located in southern France and to winter crops (typically, tomatoes are planted in November and are grown through June with harvest beginning as early as mid-January). Such information is hard to find (very disseminated and incomplete) in the specialized literature because much of it is context dependent (i.e. specific to the climate in the area) or is still informal due to uncertainties pervading many aspects of the subject. This paper focuses only the climate issue. Other important components of the production management problem such as carbon dioxide concentration and fertilization are not considered here, although we agree that preferably they should not be treated independently of the others.

Subsection A presents the important aspects that have to be dealt with in the determination of climatic setpoints. Subsection B provides time decompositions of the days and growing season that are suitable for the task at hand. The order in which the decision process considers the different aspects involved is outlined in Subsection C. Finally Subsection D shows that the pieces of knowledge may readily be translated into constraints and rules.

A. Main factors to care of

In the problem of climatic setpoint determination one has to take care of and integrate the following two classes of variables describing :

- the situation outside the greenhouse, expressed through air temperatures, solar radiation, wind speed;
- the inside situation that we want to manage and which may be evaluated through both quantitative measures of air temperature, saturation deficit and soil temperature, and a qualitative appraisal of some physiological aspects of plants including especially the stage indicators and symptoms of diseases, wilting or too strong vigor.

The setpoints are another class of variables whose values specify the domains outside of which the control computers must command the use of devices such as the heating system (start or stop) or the roof windows (close or open). The setpoint class includes the air heating, soil heating and aeration setpoints. Some elements of the above second class are directly associated with setpoints (e.g. air temperature and the air heating setpoint or air temperature and the aeration setpoint) and some are not (e.g. saturation deficit, vigor).

The choice of setpoints must be done by a reasoning process integrating the above-mentioned variables in an advantageous manner ensuring a profitable, though safe, combination of growth and development factors while keeping the energy spending within acceptable bounds and as low as possible.

Managing the production aims first at controlling the basic physiological functions such as photosynthesis, respiration, assimilation and transpiration that underlie the growth and development of plants. How the above functions individually depend on climatic conditions is rather well known though not sufficiently precisely for a accurate global quantitative modeling (Jones et al., 1989). Roughly, one can consider that growth is essentially affected by the intensity and duration of solar radiation that provides the energy needed in the photosynthesis process, whereas development is directly linked to the amount of heat (temperature) received over a period. Temperature influences also the rate of photosynthesis and, thus, the rate of growth.

The main elements characterizing the inside climate (light, temperature, saturation deficit) interact with each other; few interventions (heating, ventilating, shading) act preferentially on one particular variable but modify also several others. For examples, heating affects temperature but also the saturation deficit; ventilating affects both temperature and the saturation deficit and modify the carbon dioxide concentration (though we shall not consider explicitly this variable in this paper). Each decision has complex repercussions that are sometime opposed in their more or less delayed effects on the crop.

In order to reach the most advantageous decision in the setpoint determination problem one must rely on the currently available understanding of the phenomena, their interactions and above all the adequate decision attitudes in front of the possible classes of problems. Actually the main rules that must be fulfilled by the values of the variables in order to be part of an acceptable (i.e. rational according to the decision maker) solution are known, though imprecisely, by expert growers. For instance, the results of experimental works give us guidance on the most beneficial and forbidden light/temperature combinations in a particular period of the year for suitable growth and development of plants. Solving the determination problem is not easy, however, because several decision alternatives may have to be explored before a solution is reached. The generation and evaluation of these alternatives may be the source of a significant combinatorial complexity that growers may have a hard time to cope with, contrary to computers.

Besides growth and development factors care must be taken beforehand to prevent undesirable situations. This concerns in particular the incidence and development of diseases (mainly grey mold caused by Botrytis cinerea) or infestations by parasites. Essentially the preventive management decisions must ensure that conditions of high humidity (low saturation deficit to be more exact) are avoided. Again the rules governing such decisions are empirically available.

Another kind of undesirable situation is the lost of balance between reproductive and generative functions, characterized by particular aspects (size, shape, color) of different organs (leafs, stems, flowers, inflorescence) developing in the upper part of plants. Note that the interpretations of these aspects may be different depending on the development stage, climatic data of the close past and the time of their observation. The common symptom that the growers are able to perceive is a too strong vigor where an acceleration of vegetative growth is occurring while at the same time the flower production is decreasing. This vigor symptom may have been caused by fertilization or climate-related problems but can be controlled by proper climate settings. Simple rules, induced from the observed practices of experienced growers, tell that the appropriate reaction in case of a too strong vigor is to increase the mean (over the day) air temperature, decrease the soil temperature and lower the humidity (i.e. increase the saturation deficit). A too weak vigor is also undesirable and can be corrected by the converse actions.

A specificity of the weather in the French Mediterranean area of France is that periods of beautiful days with intense solar radiation may alternate abruptly with dark and windy days. A sharp and sudden change of weather is harmful to the plants. It stresses them due to the inertia of some physiological functions such as water absorption. The stress phenomenon occurs when a long period of cloudy and damp days (during which plants have gotten used to a low transpiration activity) is followed by a period of dry sunny days. A wilting of the plants ensues and is noticeably visible at the moment when the solar radiation and saturation deficit are important. It may appear during several consecutive days while the plants are not used to the new climatic conditions. From a physiological point of view this wilting phenomenon corresponds to a demand of water from the environment (strong solar radiation and high saturation deficit) that exceeds the amount absorbed through the roots. The stomatal regulation causes then the stomata-closure which in turn impedes the photosynthesis process. A natural solution to this problem is to prepare the plants to the change of weather that can be foreseen using the commonly available local area forecasts. The preparation consists in increasing the soil temperature (to increase the water absorption through the roots) and increasing the saturation deficit (to increase the water demand of the environment) the last day of the cloudy weather period. One could also reduce the detrimental effect of such a sudden change by increasing the soil temperature and decreasing the saturation deficit the first day of the sunny period.

Finally in deciding the control regime, especially the temperature regime, to be maintained in the greenhouse the decision maker must take into account not only the requirements of the plant but also the cost of providing heat. This economical factor depends on timing, yield of the production, expected level of market prices of tomatoes and price of energy. So far we have not considered the problem in its full generality where the above considerations fluctuate from one year to the other. Implicitly a standard situation is assumed and the solution corresponding to minimal spending in energy is preferred.

As shown in this subsection, the local area weather predictions are worth considering in the quest for good management performance.

B. Stage of growth and periods in a day

It is clear that the above factors are more or less relevant or important depending on the stage of growth of the plants. To each stage correspond specific rules used in the decision process. We are considering four main stages that practically correspond to uniform morphological objectives and management recommendations. Each stage is characterized by a sum of degree-days that the plants must receive within the corresponding interval of time. Table 1 describes the four stages in the case of the Capello tomato variety.

Tomato plants are subject to life-cycles that demand special considerations. Therefore, we have decomposed the 24 hours of a day in four periods :

- from sunrise to sundown during which the objectives concerning photosynthesis and transpiration functions must be satisfied ;
- the first part of the night where respiration and assimilation partitionship is taking place as a continuation of the photosynthesis process of the previous diurnal period ;
- the main part of dark night that requires heating, although the plant activity is lessen (this is where energy saving can potentially be made by adequate management);
- the dawn where the plants have to be prepared to the next daylight period and where humidity has to be taken care of.

The starting point and duration of the above periods are changing through out the season.

A table of suitable ranges of the inside variables has to be given for each stage of growth in the interval of time that is of interest (i.e. from planting time until late in the fourth stage when control is no longer possible due to high temperature and intense solar radiation) and for each of the four periods considered in a day.

C. Decision process

The management of the greenhouse climate by the growers comes down to finding the best compromise between the aforementioned factors underlying the different stages of growth. The decision process must absolutely ensure a correct handling of aspects related to important or irreversible damages that may be caused to the crop and to the formation and development of fruits: avoiding the forbidden zones of the climatic variables, take care of risks of diseases and crop infestations. Then, outside the possibility of occurrence of dramatic situations, the reasoning process must consider choices of setpoints that can maintain a good growth/development combination, an adequate photosynthesis/ respiration balance, a normal vigor of plants and a fair balance of the auxiliary fauna. The choices must also take into account a possible sharp change of weather. Finally among the possible alternatives the decision process must select the best or preferred one with respect to the criteria considered. Usually this amounts to find out the cheapest one in terms of cost of energy to be provided by heating.

The decision process starts from an overall objective expressed as a desired mean temperature over the 24 hours of the day that chiefly depends on the solar radiation forecasted. This objective can be slightly adjusted to take into account the recent history of the crop growth. For instance, the grower may want to make it up for missing degree-days if he is lagging behind in terms of the desired sum of temperature over a growth stage. Conversely, he may wish to slow down the crop development if he is ahead on his degree-days program. Furthermore, the adjustment of the global objective may be made for anticipatory reasons.

The decision process goes through successive refinement of the overall objective by generating consistent sub-objectives associated with smaller intervals of time. At the end of this time-based iteration the setpoints which are nothing but sub-objectives concerning each of the four elementary periods composing a day are reached.

D. From empirical knowledge to constraints and rules

As explained in the sequel, SERRISTE operates on pieces of knowledge that are essentially represented under the form of constraints on the variables involved. For instance, the suitable growth/development combinations are expressed by mathematical functions that associates a solar radiation intensity (which is an imposed datum) with an interval of acceptable air temperatures. This association varies along the season and depends on static factors like the latitude of the greenhouse site or the physical characteristics of the greenhouse. The way of setting heating and aeration in order to obtain a desired mean air temperature and humidity during day time is another example of an empirical and approximate know-how that ultimately is expressed by a numerical constraint on the difference between aeration and heating setpoints.

The prevailing crop and weather conditions of the current day are inducing numerical constraints on the possible values of some of the variables to consider (including setpoints). The context dependency of these constraints is best conveyed by rules. For example, if the plants are too vigorous, then increase the mean daylight air temperature of about one or two degrees, decrease the soil temperature of about two degrees and increase the saturation deficit. Another example is given by the following rule : if damp weather is forecasted enforce a smaller difference between aeration and heating setpoints (that is the way of ensuring that aeration will occur frequently thus preventing high humidity that may contribute to disease incidence).

The precise form of the constraints and the way they are processed in the SERRISTE system for finding climatic setpoints is the subject of the next section.

IV. THE SERRISTE SYSTEM

From an artificial intelligence point of view, solving a problem requires knowledge representation capabilities and an inference machinery appropriate to the reasoning task at hand. We have translated the problem of the determination of setpoints into what is known in artificial intelligence as a constraint satisfaction problem (CSP). In this framework, knowledge is expressed through algebraic constraints restricting the possible values of numerical variables. The core of the reasoning process consists in a resolution mechanism that searches all the values (of the variables) that are compatible with all the constraints.

This section describes the conceptual scheme of SERRISTE and the main features of its implementation. Subsection A outlines the artificial intelligence approach underlying SERRISTE. Subsection B shows briefly how the knowledge is represented. A technical presentation of the main features of the resolution mechanism is given in Subsection C. The global functioning of SERRISTE is described in Subsection D. The implementation issue is briefly addressed in Subsection E.

A. Outline of the artificial intelligence basis of SERRISTE

The problem of determining climatic setpoints is seen as an allocation task in which a set of variables must be assigned values satisfying simultaneously a set of constraints that represent pieces of knowledge to be taken into account. The constraints specify directly or indirectly the allowed (or forbidden) values of the concerned variables.

The field of constraint satisfaction problem (CSP) has been intensively studied and used in artificial intelligence (AI) (Mackworth, 1987; Kanal and Kumar, 1988). A wide range of practical tasks including planning scheduling, design can be modeled in this framework. See (Stone et al., 1992) for an example of application in crop rotation planning.

The core resolution process consists in choosing a not yet instantiated variable and assigning it a value taken in its domain such that all constraints bearing on this variable are satisfied. In doing so, one may reach a dead end in which no value can be given to the variable under consideration. A backtracking is then necessary; one must withdraw the instantiation of a previously considered variable and assign a new value to it. When all variables have been instantiated: a solution is then found. However, there may be zero or several solutions. In the latter case, resolution may be carried on until all solutions are found. This basic principle of search in a set of alternatives is conceptually at the very heart of artificial intelligence. Many algorithms have been developed in AI for this heavily combinatorial kind of problem.

The essential goal of a CSP resolution mechanism is to make the best possible choices in the following two situations:

- selection of the next variable to be instantiated,
- determination of the value to be assigned to the selected variable.

In the SERRISTE system, these choices are made according to two guideline criteria: respecting the logic underlying the setpoint determination process so as to facilitate a step by step understanding of the resolution, and ensuring efficiency of the resolution.

The selection of variables relies on the use of a specific tree structure of variable clusters that is constructed beforehand and that tells anytime which variables to consider next given the last instantiated. This so-called evaluation tree is produced by a careful analysis of constraint-induced dependencies between variables. Each node of this tree is a cluster of strongly interdependent variables.

Usually, before achieving an instantiation, nothing tells how to identify the values that will lead to a global solution. At most one can use a filtering technique to reduce, for each variable, the number of alternatives to be explored. Such a technique aims at eliminating impossible alternatives (i.e. those leading to inconsistencies). The immediate effect of applying the filtering technique on the variables of a cluster is to reduce the domains (i.e. the set of candidate values) of these, or some of these variables. Reducing the domains of a particular set of variables may result in that some of the previously possible values (i.e. that are assumed to be so) of other variables become impossible. The filtering technique propagates automatically this effect through all the concerned variables in the constraint network.

Finally, when several sets of possible values are found by the resolution mechanism (usually around ten, but up to one hundred in some cases), SERRISTE has to perform a sort of tradeoff reasoning by integrating preferences (e.g. attitude with respect to disease occurrence, energy saving strategy) that are not expressed by constraints of the above-mentioned type. This ultimate phase of problem solving that aims at returning the "best" solution among the acceptable ones is not detailed in this article. We shall focus mainly on the issue of determining all the acceptable solutions.

B. Knowledge representation

The first step in applying the constraint satisfaction approach to setpoint determination consists in identifying the variables involved, their associated domains and the relations (i.e. constraints) between these variables.

So far, all the variables involved in the SERRISTE system are numerical and most of them concern temperatures. For illustrative purposes, only temperature variables are used. The set of variables includes those associated with setpoints plus some additional ones, called intermediary variables, that are needed for the reasoning process and for constraint expression. All these variables have continuous domains that are nevertheless treated as if they were discrete for two practical reasons: the CSP approach makes sense only for discrete domains and a too high precision would be meaningless. The discretization reflects actually the level of granularity found in typical human reasoning that deals with the task at hand; two temperatures that are less than 0.5°C away of each other are not considered as significantly different.

It is assumed that the domain of each variable is known and given in the form of an interval which is interpreted by the resolution mechanism as a discrete and finite set of values. These domains depend on the variety of tomato and environmental context. For instance, the domain of ITP3 (average night temperature) for a Rondello variety in the phase preceding the flowering of the first bunch is the discrete interval [11, 18] such that two consecutive values differ by 0.5°C.

Formally, constraints are relations between variables. In SERRISTE they are defined intentionally by linear equations of the general form

$$\sum_{i=1,n} c_i X_i = F \qquad (1)$$

where the coefficients c_i's are scalar values, the X_i's are variables (each having a domain) and F is an interval. In this equation the meaning of the '=' sign should be understood as: given a set of n values v_i assigned to the variables X_i, the constraint is satisfied if $\sum_{i=1,n} c_i v_i$ falls within the interval F.

Any constraint may be parametrized in order to take into account its variation through the cultivation period and its dependency on the specific conditions of the current day. Both the set F and the coefficients may be linked to context dependent parameters.

Under this general form the following constraints can be represented:

- Unary predicates (e.g. in the early stage of development, the average diurnal temperature ITP1 must belong to [14, 28]),
- Linear equations (e.g. the average temperatures at dawn, night and dusk, denoted by ITP2, ITP3 and ITP4 respectively, must satisfy the relation 3.ITP2 + 8.ITP3 + 2.ITP4 - 13.ITN = 0 where ITN is the average temperature of nocturnal time),
- Inequalities (e.g. if the foreseen solar radiation is high, one must have $2 \leq \text{ITD} - \text{ITN} \leq 6$ where ITD is the average temperature of diurnal time).

A convenient way to visualize the dependencies between the variables is to build a constraint network incorporating all the variables and constraints involved in the problem. Figure 1 is a simplified example of such a network.

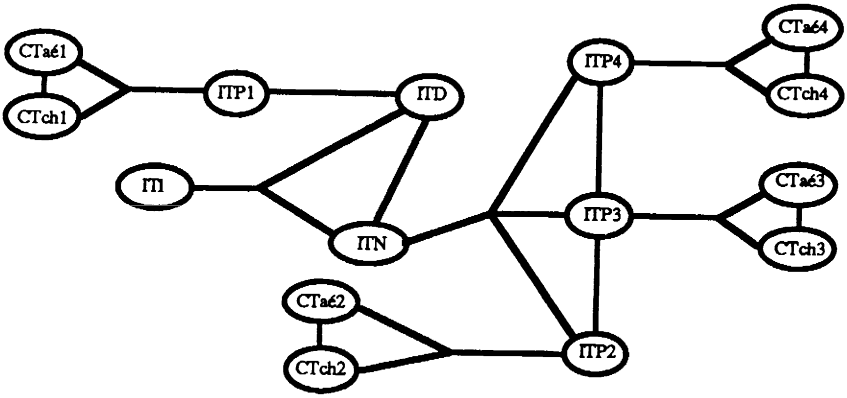


Figure 1. A constraint network

An edge between two variables means that a binary constraint connects these variables. Several edges converging in one point represent an n-ary constraint (i.e. any constraint involving n variables) on the variables located at the extremities of the edges (i.e. circled items). Usually, unary constraints are omitted for preserving a good readability of the network.

C. Resolution process

As outlined in Subsection A, the problem of setpoint determination is an allocation task in which numerical values must be assigned to a set of variables that are connected to each other through constraints. For the efficiency and the intelligibility (i.e. understandability) of the resolution process, a preliminary processing consists in building a structure that will provide an appropriate clustering of the variables and an order for exploring the constructed clusters. The clustering techniques are presented in the subsection below. The overall resolution process combines a filtering technique (that aims at eliminating impossible values) with a search algorithm that keeps trying instantiations and withdrawing them (backtracking) when a failure is detected. The filtering and search algorithms are described in the last two subsections. For more details, see (Cros and Martin-Clouaire, 1991).

Clustering of variables

The idea underlying this preprocessing phase is to exploit the structural properties of the constraint network in order to get together the variables that are strongly interdependent. Intuitively, it is wiser to try a simultaneous instantiation of such a group of variables because it is an effective strategy that avoids many backtrackings, and it brings a focussing effect that eases the task of a user who wants to follow the resolution process. The clusters of variables are organized in a tree that can be explored in a systematic way, providing an appropriate order for the instantiation of the variables.

Let us first define some basic notions used in the following. Two constraints are connected if a particular variable is involved in both (e.g. C1(V1, V2) and C2(V2, V3, V4) are connected because they share the variable V2). Two variables are directly connected if they are involved in the same constraint. More generally, two variables X and Y are indirectly connected if there exits a chain of connected constraints such that the first one concerns X and the last one concerns Y (e.g. with C1(V1, V2), C2(V2, V3, V4) and C3(V3, V5), V1 is indirectly connected to V5).

Clusters of variables are achieved and structured into a tree by running an iterative procedure that examines exhaustively the constraint network starting from a given set of variables defined as the root cluster. This procedure, when applied to a cluster of variables, produces the clusters that immediately descend from it (i.e. its sons). Initially, a set of variables playing a particular role is defined deliberately as the root cluster. The above-mentioned procedure is applied to this root cluster and is then used repeatedly for each newly created cluster. Without going into details (see (Cros and Martin-Clouaire, 1991) for the full algorithm), let us describe the basic principles behind the construction of sibling clusters descending from a given cluster G.

Let Cp1,...,Cpr be the constraints in which at least one variable of G is involved and that have not already been considered by the clustering procedure. Let us suppose that Cj is one of these constraints. $\partial j(G, Cj)$ denotes the set of variables concerned by Cj but not contained in G. This is called the development of G according to Cj. The construction of new clusters is obtained by merging the developments associated with Cp1,...,Cpr if one of the three criteria given below is satisfied, until merging between developments or newly merged developments is no longer possible. Two developments or merged developments $\partial 1$ and $\partial 2$ are merged if they satisfy one of the three criteria:

- $\partial 1$ and $\partial 2$ have a non-empty intersection;
- a variable of $\partial 1$ is directly connected to a variable of $\partial 2$ through a constraint that has not been considered yet (i.e. that has never been involved in the development of a group);
- a variable of $\partial 1$ is indirectly connected to a variable of $\partial 2$ through constraints that have not been considered yet.

In our simple example illustrated in Figure 1, the root cluster which is the starting point of the instantiation process includes only the variable IT1 that stands for the average temperature desired over a 24 hour period and has only one constraint involving IT1. The development of the root cluster according to this constraint is the set {ITD ITN}. Since no other development has to be taken into account, this set is the son cluster of the root. Two constraints involving ITD and ITN have to be considered. The developments of {ITD ITN} according to these constraints are {ITP1} and {ITP2 ITP3 ITP4}. As these two development do not satisfy any of the merging criteria, they give birth to two clusters that in turn are considered in a similar fashion. In this simple example however, no merging actually takes place. Ultimately, this construction process provides a tree of clusters.

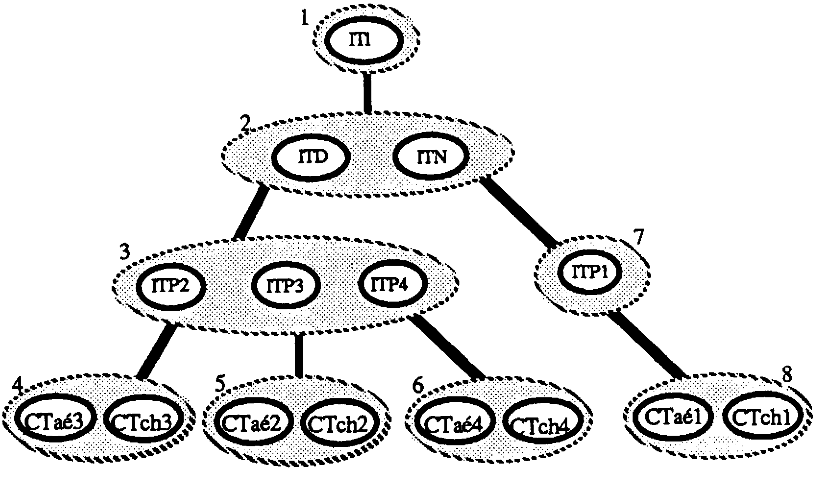


Figure 2 . Tree structure of arranged clusters

Let us assume that a particular systematic procedure for exhaustive exploration of a tree structure is given. Then, the tree of clusters may be used as it stands for the purpose of guiding the instantiation process. However, efficiency and user friendliness of the resolution process may be improved by arranging the order of the sons of each cluster. As far as efficiency is concerned, it is preferable to consider first the clusters containing variables that are difficult to instantiate. In addition, a particular arrangement may turn out to be closer to the grower's pattern of reasoning. These late arrangements are not automated because they are user-dependent).

The network in Figure 1 enables the construction of the tree structure shown in Figure 2. The numbers near the groups of variables indicate the order in which the clusters are explored. Technically, this order corresponds to a left to right depth-first search. In other words, given an instantiation of some intermediary variables, this strategy is intended to find out as soon as possible if there exists a possible instantiation for the setpoint variables which are contained in the terminal or leaf clusters.

Filtering

This process aims at reducing the set of candidate values. In SERRISTE, a so-called Waltz (1975) filtering is used. A general study of this general technique which has been employed in several artificial intelligence applications was made by Davis (1987). Given a set of variables subjected to some constraints, the filtering detects and eliminates the values that cannot take part in a solution. In some cases it may discover that the problem does not have any solution. It relies on an intelligent propagation process that is iterated through the constraints and the variables involved in these constraints.

Let V1, ...,Vn be the variables involved in constraint C and Ai the possible values for variable Vi. The Waltz filtering is based on an essential operation, embedded in the procedure Reduce_domain which when applied to the constraint C(V1, ...,Vi, ...,Vn) and the variable Vi , returns the subset of Ai obtained by removing all values incompatible with all conceivable sets of values that might be assigned to the other variables. A formal definition of this operation is:

$$\text{Reduce\_domain}(C,Vi) = \{ai \in Ai \;/\; \exists\; aj \in Aj,\; j=1,n,\, j \neq i\;\; C(a1, ..., ai, ..., an)\}$$

where C(a1,...,ai,...,an) means that the constraint C is satisfied by the value aj for the variable Vj, j = 1,n.

For continuous variables, the programming of this procedure depends on the nature of the constraint concerned. If C allows to express each variable as a function of the other variables (this is the case for the type of constraints manipulated by SERRISTE) and if each Ai is an interval, then the Reduce_domain operation can be performed by a calculus on intervals which requires only simple computation on the bounds. For example, given the constraint $V1 + V2 = 10$, the initial domains $A1 = [0 , 2]$ and $A2 = [5 , 9]$. The new filtered domains $A'1$, $A'2$ can be obtained as follows: $A'1 = A1 \cap [10 - 9, 10 - 5] = [1, 2]$ and $A'2 = A2 \cap [10 - 2, 10 - 0] = [8 , 9]$. Note that the result would not be improved by reusing Reduce_domain with A'1 and A'2 'instead of A1 and A2 (i.e. the returned sets would not be smaller).

Such a filtering takes advantage of the linearity of the constraints and amounts to treat the variables as continuous. We shall see that the general search algorithm presented in the next subsection works with discrete variables. The domain discretization that is necessary to carry on the resolution is done after any filtering operation. It is a particularity of our approach to use the algebraic nature of constraints and variables to switch conveniently from continuous to discrete domains and vice versa. So, rather than an expensive filtering on discrete universes, SERRISTE performs a much simpler filtering using interval calculus on continuous universes.

In the Waltz algorithm, the Reduce_domain operation is used in the procedure Revise which, for a given constraint C, detects among its associated variables those with a domain that can be reduced and eliminates the corresponding values. The Waltz algorithm applies to a set of constraints that one is willing to take into account. Each of these constraints are considered in turn by the Revise procedure. Each time some values can logically be removed from the domain of a particular variable involved in the constraint under consideration, the algorithm removes these values. The other constraints involving variables with a modified domain are then pushed in the queue of constraints to be considered by the Waltz filtering. In this way, the consequences of reducing a domain are propagated and induce modification in the domains (i.e. of other variables) through the appropriate constraints. The propagation of domain reduction is carried on until there is no more constraint to be considered (natural quiescence) or until a termination condition applies.

Termination criteria (such as the number of times any constraint may be considered) are needed because, in some cases, the Waltz filtering algorithm may go into infinite loops or reach quiescence only after a long time. Even in these cases however, filtering is worth doing since it reduces the number of failures in the instantiation process. The Waltz algorithm is sound, that is, it cannot eliminate good values (it cannot cause a solution to be missed), but incomplete (except with simple constraints) since the reduced domains may still contain elements whose combination is incompatible with some constraints.

Search algorithm

The complete resolution mechanism first performs a filtering over the entire set of constraints in order to attain a certain level of consistency by pruning the values that for sure cannot take part in a global solution. At this stage a global inconsistency may be discovered. It means[2] that in such a case there is no solution and there is no need to push the resolution further. If successful in the initial filtering the resolution mechanism enters in a typical CSP search algorithm that considers one cluster after the other (in the order specified by the tree structure built for this purpose). Each time, it makes tentative instantiations of the variables of the current cluster, checks that they satisfy the directly concerned constraints, propagates by Waltz filtering their effects on other variable domains and backtracks to another tentative or another group when a complete solution is found (so as to seek another one) or when a failure is encountered. This essential algorithm corresponds to the following procedure, called FindSolutions, which (hopefully) is self explanatory.

```
Procedure: FindSolutions
set  CurrentCluster to  root-cluster
1  if  none of the domains of the variables of CurrentCluster is
             empty
      then  chose a combination of values from each of these
             domains append it to the partial solution under
             development suppress these values in their
             corresponding domains
      else  goto  4  (i.e. backtrack)
   if any constraint concerning the current and/or already
             considered clusters  is not satisfied
      then  goto 1 (i.e. try another combination of values)
2  if filtering applied to the other constraints un uncovers
                    inconsistency
          they undo this local filtering
                    maintein (clean) the data structure of the partial
                    solution
                 goto 1  (i.e. backtrack)
3  if the CurrentCluster is the last one
      then  save the partial solution under development
              (since it is indeed an acceptable solution)
            goto 1 (i.e. try another combination of values)
      else  set  CurrentCluster to  the next one
              (according to the order induced from the tree)
            goto 1 (i.e. try another combination of values)
4  if the CurrentCluster is the root-cluster
      then  return  (i.e. all the acceptable solutions, if any, have
```

---

[2] Incidentally , this indicates that there is something wrong in the knowledge base since the system should be able to provide setpoints in any situation.

```
              been found then)
    else  undo the local filtering
              (put back the domains as they were after the initial
              filtering)
          set  CurrentCluster to the preceding one
          maintain (clean) the data structure of  the partial
                solution
          goto  1 (i.e. try another combination of values)
```

From an artificial intelligence point of view, the salient features of our approach are the processing of variables by clusters rather than individually and the incorporation of a filtering operation before any instantiation of the variables contained in a cluster. We also use advantageously the fact that the domains of the variables may be considered as continuous for filtering purposes and discrete in the final stage of search.

D. The global functioning of SERRISTE

The global functioning of SERRISTE can be summarized in a seven-step process. The first two steps (Phase A) are performed only once in a cultivation period. They consist in an initialization of some computational structures and collection of data that remain valid over the entire cultivation period. After these two steps have been completed the resulting structures and collected data are saved in an appropriate manner so as to be reused every day of the cultivation period. The sequence going from Step 3 to Step 7 (Phase B) corresponds to a typical routine (daily) use of SERRISTE for the determination of setpoints.

Step 1: "Compilation" of the knowledge base

This step aims at analyzing the knowledge base in order to set up some data structures mainly for efficiency of the resolution. In particular, this is where the variables are clustered into groups which are themselves organized in the so-called evaluation tree.

Step 2: Input of data pertaining to the greenhouse and the crop

The user has to answer to a sequence of questions concerning in particular:

- the location of the greenhouse (latitude);
- the characteristics of the greenhouse (size, type of heating system, type of cover, availability of thermal screens,...);
- variety of tomato used.

Step 3: Input of data pertaining to the current conditions

The questions concern:

- the current date and current stage of crop growth;
- a qualitative assessment of the state of the crop (vigor, presence of disease);
- the forecasted weather for the current day (upper and lower temperatures, wind speed, cloud cover in qualitative terms like cloudy, blue sky, rainy);

- the forecasted cloud cover for tomorrow;
- effective cloud cover observed yesterday;
- effective climate conditions observed (measured) yesterday inside the greenhouse;
- short term objectives (e.g. increase vigor, compensate a temperature backlog).

Step 4: Computation of needed information based on data obtained in Steps 2,3

From the data collected in the preceding steps the values of the parameters involved in the constraints and other useful parameters such as the time of sunrise, the estimated numerical value of maximal solar radiation are computed by applying appropriate formulae.

The domains of some variables vary depending on the stage of growth and the variety of tomato. The domains corresponding to the considered situation are deduced by firing appropriate rules.

Step 5: Resolution of the constraint satisfaction problem

In this step, which has been explained in Subsection C, SERRISTE makes an initial filtering to check the global consistency of the problem and ease the search of solutions. If the filtering does not discover an inconsistency the system enters the FindSolutions procedure. By the end of this step, the system has produced a (possibly empty) set of assignments to the variables.

Step 6: Choice of the "optimal" assignment

The system computes among all assignments (which represent acceptable solutions) the "best" one according to the specified criteria (typically minimize consumption of energy of the heating system). A very simple model of energy balance is used.

Step 7: Save results and prepare reports according to the demand of the user

The user may wish to keep a daily report file containing the chosen solution together with the conditions (input and computed) characterizing the current day. For analysis purpose, the user may request to save also additional information such as the set of all acceptable solutions or a detailed trace of the run which are very useful to pinpoint various kind of problems (e.g. why there is no solution or which constraint does not work properly and is responsible for aberrant solutions).

E. Basic principles of the implementation

The SERRISTE system has been developed with KAPPA (IntelliCorp, 1991) which is essentially a knowledge representation tool kit. It is a strongly object-oriented hybrid environment that allows to mix different features objects, rules, functions, demons and conventional programming techniques when developing an application. KAPPA's objects can be structured in hierarchies composed of classes, subclasses and instances. The properties of an object are expressed in slots or attributes that may themselves be characterized by restrictions concerning for instance the type of information allowed as value (text, numerical, Boolean, object) or the number of values allowed in case of multivalued slot. The behavior of the objects is defined by procedural code in the so-called methods which can be inherited as any other property through hierarchical links. In addition, to any slot one can attach demons that automatically respond to alterations on the slot. The rules can be used in forward or backward modes to perform reasoning tasks involving the objects. The programming language KAL used within KAPPA (for writing methods for instance) is closely integrated with the C programming language, in that external C functions can be called directly. KAPPA (version 1.2) runs under MS-WINDOWS and the MS-DOS operating system.

The CSP-related capabilities of SERRISTE are embedded in four classes, Quantites, Contraintes , Solutions and Groupes. Quantites has two subclasses: one for the variables involved in the constraints and one for the various kinds of parameters. The hierarchy below Contraintes provides the prototypes of the different kinds of constraints. They are distinguished by the number of variables that they concern. The object Solutions has a subclass Solutions-Acceptables whose purpose (as its name indicates) is to gather the different acceptable solutions found by the system. The slots of this object are numerical and have names corresponding to the variables. When a solutions is found during the search process, an instance of SolutionsAcceptables, named SolutionAcceptable_n (n being the number of the solution) is created, and each slot is assigned the value of the corresponding variable. The subclass Solutions-Preferees is designed to receive the instance of the preferred solution (so far we have assumed there is only one optimal solution). Groupes is the prototype of a cluster of variables, i.e. its instances are the clusters involved in a particular application and are created in Step 1. There is also an important object, named Resolution, which involves the top level methods of the resolution process. The main properties and methods attached to these objects and the role they play in the resolution process are described in (Martin-Clouaire et al., 1992).

The aspects of the real world which are relevant and important for the climate management task are represented through objects such as Serre (meaning greenhouse), Culture (meaning crop) and Temps (meaning time). In addition, instances of the two classes Meteos (meaning weather) and ClimatsMesures (meaning measured climate data) express weather forecasts and greenhouse climate measurements for a given day. The instances are named Meteo_j or Climat_j, where j refers to the date of the considered day.

The essential part of knowledge specific to a particular application is expressed by instances of the classes Variables and Contraintes. This section gives concrete examples of two instances used in the current version of the knowledge base. For clarity, they are presented in a slightly edited manner to avoid the burden of syntax definition.

```
-----------------------------------------------------
CTsol  Instance of subclass Variables
-----------------------------------------------------
Slots with their local values if any :
        Valeur
        TMinMin = 10
        TMaxMax = 30
        ValeurMin
        ValeurMax
        ValeurMinApresFiltrageInitial
        ValeurMaxApresFiltrageInitial
        ValeurMinReduite
        ValeurMaxReduite
Local methods:
-----------------------------------------------------
```

Figure 3. The slots of the instance CTsol

The object CTsol (Figure 3) represents the soil temperature setpoint variable. It is shown in a state where only two slots have received values. There is a slot Valeur that contains the value of the variable when available. The other slots go in pairs and convey lower and upper bounds of the variable value at four different stages of the resolution process. They correspond, in the order of Figure 3, to the default situation before any processing (e.g. a soil temperature setpoint is always within the bounds 10 and 30), the post-initialization state, the situation once the initial filtering has been performed, the current situation while searching the acceptable solutions. Note that the instance CTsol does not have any local method.

```
-----------------------------------------------------
C10  Instance of subclass ContraintesBinaires
-----------------------------------------------------
Slots with their local values if any :
        V1 = ITN
        V2 = ITD
        a1 = 1
        a2 = -1
        a3Min = 2
        a3Max  : Demon Determiner_a3Max
Local methods:
        Determiner_a3Max
                { GetValue(DTD_N:Valeur); }
-----------------------------------------------------
```

Figure 4. The slots and method of the instance C10

The constraint C10 represents the following piece of information: the difference between the average temperature of diurnal time (denoted ITD) and the average temperature of nocturnal time (denoted ITN) must be within the interval $[2, \Delta TD\_N]$ (i.e. $2 \leq ITD - ITN \leq \Delta TD\_N$). The parameter Δ TD_N depends on the prevailing weather conditions of the day. The instance C10 (see Figure 4) is a member of the class ContraintesBinaires which has slots for the terms composing a typical binary constraint of the form $a1.V1 + a2.V2 = [a3Min, a3Max]$. At the time of the definition of C10 the value of a3Max is not known. So the demon method Determiner_a3Max has been associated to the corresponding slot and gets the value as soon as it is needed. The code of the local method Determiner_a3Max simply tells to get and return the value of the slot Valeur of the object DTD_N which represents a particular parameter.

```
Rule DCTsolAdapt6
If
          Culture:Botrytis = non And
          Culture:Vigueur = bonne And
          Culture:ObjVigueur = diminuer And
          Meteo_j-1:qMRgj = faible And
          Meteo_j:qPRgj = eleve
Then
          DCTsolAdapt:Valeur := 1
```

Figure 5. The rule DCTsolAdapt6

In the current knowledge base of SERRISTE, forward chaining rules are used to deduce the values of some parameters that depend on the data entered in the initialization phases. The rule shown in Figure 5 tells literally that if the crop is not affected by Botrytis and its vigor is good and the objective with respect to vigor is to make it lower and the solar radiation was poor yesterday and the solar radiation should be high today then, the slot Valeur of the parameter DCTsolAdapt must be set to 1.

As far as the size of the problem is concerned, the current knowledge base contains 24 variables, 43 parameters, 27 constraints and 62 rules.

V. DISCUSSION

The knowledge-based system SERRISTE aims at providing the versatile decision support capabilities that are needed for an efficient management of daily greenhouse climate. The main motivation behind the work reported here was to gather an adequate body of heuristic knowledge and develop the artificial intelligence software SERRISTE capable of representing this knowledge in way appropriate to the execution of the task on conventional greenhouse computers. Before the advent of accurate mathematical models of crops and greenhouses becomes a reality it is our belief that the best practical source of help in the management task is provided by a mixture of a general scientific background (on plant physiology and greenhouse engineering) with empirical agronomic expertise about growing particular crops in particular greenhouses under specific conditions. See, however, the papers by (Jones et al., 1991) and (Seginer, 1991) for recent contributions on the side of numerical approaches.

SERRISTE rests on a body of knowledge that is typical of what expert growers are using and that we have expressed essentially under the form of constraints on numerical variables. However, strictly speaking, SERRISTE is not mimicking human reasoning in solving the setpoint determination problem; the constraint satisfaction approach is too combinatorial to be carried out by a human brain. Thus, SERRISTE is more a knowledge-based system than an expert system.

Few works have been reported on computer decision support systems applied to the management of greenhouse climate. Of special interest, however, is the paper by (Gauthier and Guay, 1990) that presents an experimental object-oriented design of an ambitious system addressing both management and control of greenhouse climate.

A. Interesting features and current status

An expert-based decision making allows a rigorous management of interactive parameters such as humidity and temperature and, thus, greatly contributes to better prevent diseases (which otherwise cause expensive interventions). Moreover, the decision support capabilities of SERRISTE, if used regularly, bring along coherence in the crop management; harmful effects of discontinuity that are encountered when several decision makers are involved in the choices of setpoints can be avoided.

Besides relying on expert knowledge and AI constraint satisfaction techniques, the management approach undertaken through the SERRISTE system is novel in that it exploits weather predictions for better satisfying the physiological requirements of the plants, and for reducing energy costs. It also takes into account qualitatively assessed data such as symptoms of diseases and vigor of the plants.

The structuring of expertise, that has been discussed in Section III, is a required step before incorporating the knowledge in a decision making computer system. The formalization of growers' know-how has brought valuable benefits. Experts make progress in the field because they have to think about their practices, clarify what is implicit and uncover gaps in knowledge. Another significant by-product of this basic work is that it contributes to facilitate the transfer of knowledge.

Basically, SERRISTE solves an assignment problem through a general artificial intelligence approach. Although so far the constraints have chiefly been linear, we did not use mathematical programming techniques for the following reasons. Firstly, a clear objective function is lacking (reaching the truly optimal solution is illusory given the qualitative nature of the knowledge to take into account). Secondly, it is highly desirable to produce an understandable resolution for users so as to facilitate the maintenance and development of the knowledge base. Finally, we wished to keep open the possibility of incorporating easily new pieces of knowledge and using heuristic choices if necessary.

The computer program centered on the constraint satisfaction techniques has been tested extensively to uncover and fix programming errors. We have started the validation process of the whole SERRISTE system, that is, the application-specific knowledge possessed by the system and the overall computational approach. We are now in the process of verifying the consistency and the quality of the results provided by SERRISTE. Indeed a currently undergoing experiment consists in applying the setpoints of SERRISTE in one compartment of a greenhouse and compare the results with those obtained in a neighbor compartment managed by a human grower that applies conventional rules. The experiment is not finished yet and it is not the subject of this paper to discuss its results but the available empirical evidence has shown so far positive benefit from using SERRISTE.

Although SERRISTE is presently only a prototype system the overall approach seems sound. Deeper testing and evaluation are required: the system must be confronted to a significant number of cases and it must be confronted with a larger problem including the management of carbon dioxide and nutrition. On the problem solving side SERRISTE suffers from deficiencies related to the following aspects. First and most importantly, the system is too slow. Second the expressive power of the knowledge representation capabilities is too limited. It is impossible to code faithfully the expert knowledge because of the inherent imprecision pervading it. The crisp constraint framework is not satisfactory in this respect. Third, sometimes there are too many solutions among which some are not significantly different.

B. Future works

The overall project of computer-aided greenhouse management under-taken through the SERRISTE system is, of course, far from complete.

Currently, we focus the effort on an extension (Martin-Clouaire, 1992) of the representation capabilities in order to deal adequately with soft constraints. Basically, we use approximate reasoning techniques that have their roots in the theory of possibility (or fuzzy set theory). In addition, this extension allows to make distinctions among the constraints; some being more important than others.

Other directions that are in need of further exploration and development are discussed in the remaining part of the section.

The current knowledge base needs to be improved on several aspects. In particular, what remains to be addresses is the problem of managing the production in the second half of the spring season characterized by higher outside temperature. As already mentioned, the present version of the system deals only with climate issues. The next step to consider is the management of carbon dioxide enrichment and nutrition conjointly with the climate factors.

Before considering a routine use of an extended system of the SERRISTE type one must carry out an in-depth validation analysis by using it in real contexts (real greenhouses) under various and numerous conditions. The robustness of the approach could be evaluated by applying it to other geographic contexts. Moreover, other crops may be considered.

In order to be really useful, a greenhouse management system must provide solutions (if any) or advice in critical or exceptional situations that usually leave growers in helpless positions. Such situations occur, for instance, when a particular device used to modify the greenhouse climate has not worked properly (due to a failure for example). The kind of knowledge required to cope with such problematic cases has still to be identified. Whether or not the current reasoning capabilities of SERRISTE are sufficient remains to be shown.

The exploitation by a decision making system of the property that tomato plants act as efficient temperature integrators over a several day period enables anticipatory decisions which can potentially contribute in a significant manner to energy saving. The rationale behind these "reservoir" capabilities of plants and the basic principles behind its profitable exploitation are briefly addressed in (Boulard et al., 1991). So far SERRISTE leaves the burden of reservoir management to the grower who is given the possibility to adjust the overall objective of the day (see Subsection D in Section III). Providing a decision support that could use the reservoir capabilities is difficult for several reasons including: the incompleteness of knowledge on the topic, the necessity to reason over a several-day time scale and, of course, to decide under uncertain predictions.

Acknowledgements: This work has been partially supported by the AFME agency (Agence Française pour la Maîtrise de l'Energie). It has been achieved in the framework of an INRA project involving colleagues from Bioclimatology (M. Mermier, P. Reich, S. Simon) and Agrarian Systems and Development (J. Lagier) Departments. J. Jihad, K. Kováts and V. Terrazzoni have contributed to the implementation of SERRISTE.

REFERENCES

Baille A., Boulard T., Gary C. (1990) Les critères d'optimisation dans la gestion du climat et de la production sous serre. Comptes Rendus de l'Académie d'Agriculture, 30 Mai, 25–36.

Boulard T., Jeannequin B., Martin-Clouaire R. (1991) Analysis of knowledge involved in greenhouse climate management : Application to the determination of daily setpoints for a tomato crop. Proc. IFAC/ISHS Workshop on Mathematical and Control Applications in Agriculture and Horticulture, Matsuyama, Japan (Sept. 30 - Oct. 3), 265–270.

Challa H., Nederhoff E., Bot G., van de Brack N. (1988) Greenhouse climate control in the nineties. Acta Horticulturae: 230, 459–470.

Cros M.-J., Martin-Clouaire R. (1991) Détermination de consignes par satisfaction de contraintes. Proc. International Conference on Expert Systems and their applications, Avignon, France (27–31 Mai), Vol. 3, 401–413.

Davis E. (1987) Constraint propagation with interval labels. Artificial Intelligence: 34(1), 1–38.

Gauthier L., Guay R. (1990) An object-oriented design for a greenhouse climate control system. Transactions of the ASAE: 33(3), 999–1004.

IntelliCorp (1991), KAPPA Reference Manual (version 1.2). IntelliCorp Inc., Mountain View, CA, May 1991.

Jones J.W., Dayan E., van Keulen H., Challa H. (1989) Modeling tomato growth for optimizing greenhouse temperatures and CO2 concentration. Acta Horticulturae: 248, 285–293.

Jones P., Jones J., Hwang Y. (1991) Simulation for determining greenhouse setpoints. Transactions of the ASAE: 33(5), 1722–1728.

Kanal, L., Kumar V. (Eds.) (1988) Search in Artificial Intelligence. Springer-Verlag, New-York.

Mackworth A. (1987) Constraint satisfaction. In the Encyclopedia of Artificial Intelligence, Wiley-Interscience Publication (Shapiro S. Ed.), 205–211.

Martin-Clouaire R. (1992) Dealing with soft constraints in a constraint satisfaction problem. To be presented at

"International Conference on Information Processing of Uncertainty in Knowledge-based Systems", Palma de Mallorca, Spain (July 6–10).

Martin-Clouaire R., Kovats K., Cros M.-J. (1992) Determination of greenhouse climate setpoints by SERRISTE: the approach and its object-oriented implementation. Internal Report. To be published.

Seginer I. (1991) Optimal greenhouse temperature trajectories for a multi-state-variable tomato model. Proc. IFAC/ISHS Workshop on Mathematical and Control Applications in Agriculture and Horticulture, Matsuyama, Japan (Sept. 30 – Oct. 3), 73–79.

Stone N., Buick R., Roach J., Scheckler R., Rupani R. (1992) The planning problem in agriculture: farm-level crop rotation planning as an example. Artificial Intelligence Applications – Natural Resources, Agriculture, Environmental Studies: 6(1), 59–75.

Tantau H.-J. (1989) On line climate control. Acta Horticulturae: 248, 217–222.

Udink ten Cate A., Challa H. (1984) An optimal computer control of the crop growth system. Acta Horticulturae: 148, 267–276.

Waltz D. (1975) Understanding line drawing of scenes with shadows. In the Psychology of Computer Vision (P.H. Winston Ed.), McGraw-Hill, New-York, 19–91.

Chapter 10

COMPUTER OPPORTUNITIES IN AGRICULTURE AND HORTICULTURE

W. Day

Process Engineering Division
Silsoe Research Institute
Wrest Park, Silsoe, Bedford, UK

I. INTRODUCTION

In every industry, the ongoing revolution in computing technology is opening up new methods and approaches to process optimisation and control. In general, the agricultural industries will utilise many developments that have been motivated by the demands of other industries. Thus the major opportunity for control systems on-board vehicles is being taken up by the automobile industry for the car market. The scope for utilising the technology on tractors and other farm vehicles will follow from these developments, given the considerably greater size of the car market, and therefore ability to support research and development. However the agricultural sector does present many interesting and novel problems that can be tackled as new technology appears. This chapter seeks to illustrate key themes that will be of major importance to the advance of computer applications in agriculture and horticulture.

Computer technology is making rapid advances in processing power and speed, in compactness and in decreasing unit cost, and with these advances new applications in agriculture and horticulture will become feasible. These applications will reflect the wide variety of agricultural processes, and will bring with them specific challenges to the technology. The biological variability in agricultural systems, the complex interactions between the physiological processes that underlie crop growth and the uncertainty in the prediction of process performance, because of the influence of future weather, are important issues in many agricultural and horticultural systems.

Developments in hardware and software will allow practical implementation of computer techniques in produce grading, crop yield prediction, and crop and animal management. A number of example areas have been chosen to illustrate how computer applications can be expected to develop in the coming years. The relationship between scientific research, computing systems and the application area are discussed and linked to future opportunities.

II. IMAGE ANALYSIS

In horticulture, many processes require human operators to make decisions based on visual information. This is particularly apparent for the operations of produce harvesting and grading. It also applies more widely in agriculture, where stockmen assess the health and well-being of their animals in part using visual information, for example on activity and gait, and farmers identify plant diseases through their visible symptoms. Applying the technique of computer-based image analysis to these problems raises a number of interesting research challenges. It is particularly important to recognise that the objects of interest are generally biological, and therefore there is considerable variability from object to object within a batch. The objects are also often viewed in an unstructured environment, again complicating the decision-making task.

The advance of practical applications of image analysis in agriculture and horticulture has been crucially dependent on two key attributes of the revolution in computing hardware - decreases in system cost and increases in processing speed. Standard TV camera technology can produce images resolved into 512 x 512 pixels, with each pixel giving intensity as an 8-bit number, representing a grey level from 0 (black) to 255 (white). Such information is generally adequate to represent many of the scenes of interest, and standard colour cameras can expand this to more complex discrimination. Because this capability has been developed via TV based technology, which has a mass market, unit costs are low and the agricultural application becomes feasible. However, it is not sufficient just to create the information. Decisions in agricultural processes must be made rapidly if process automation is to be successful, given the relatively low value of the agricultural output. To deal with a quarter of a million items of information in an image rapidly has required the availability of fast processors. Processing speeds are estimated to be doubling every three years, and speeds have now reached around ten million instructions per second with memory access times down to tens of nanoseconds. If these changes continue, image processing tasks of even greater complexity may be possible.

Practical examples of the application of computer-based image analysis in agriculture and horticulture already exist, in the execution of straightforward repetitive tasks like those associated with produce grading (Marchant, 1990). Research can be expected to make progress with more complex tasks, and specific applications will continue to be vitally important to establish the scale and form of the real-world problems (Tillett, 1991). There is also major scope for new advances through the establishment of generic approaches to issues like 3D structural interpretation, tracking and analysis of moving images in a scene, and training of computer programs. The development of more rapid processing systems, and particularly transputers which can provide parallel processing and sophisticated intercommunication between processors, will be particularly important.

Advanced image interpretation

The use of artificial intelligence will be a key tool to enable the future advance of image interpretation. The biological variability inherent in agricultural objects, accompanied by geometrical complexity and considerable subtlety in the differences in form or colour that are important in judging quality, present a complex web of information. Intelligent knowledge-based systems can provide a way of utilising past knowledge whilst making decisions about the current task. The knowledge is entered by the operator, whilst the capability to reason is present in the computer in the form of an 'inference engine'.

An alternative approach to achieving intelligence is to allow the computer to learn from examples. The 'neural network' concept is based in principle on the learning and memory abilities of the human brain. The training period is used to teach the program from examples of 'correct' and 'incorrect' objects. The neural network can then apply this knowledge to classify other objects. The approach is particularly useful where the quality attributes are fairly easy for human decision but are difficult to quantify as rules or numbers.

Ben Hanan et al. (1991) have presented an example of such a use of neural networks in classification of apples. A more complex example of the approach has been to determine the position of nodes in an image of a chrysanthemum plant for micropropagation (Davis, 1991). The nodes are characterised by rapid changes in direction of the image outline. Classifying the whole outline in this way will specify many positions which could be nodes, whereas only a few are. A training routine can then be used in conjunction with a number of images for which nodes have been separately identified, and after this training the program is capable of identifying nodes with a high degree of success.

Intelligent systems can be expected to be of increasing importance, and will probably have particular significance in agriculture where the variability in size and geometry will make definition of precise rules and formulae difficult.

Image analysis as a sensor for robot control

The opportunity for use of image information in real-time control of automation will be a logical consequence of advances in image processing methods. The use of robots to undertake mechanical tasks that entail a high degree of 'mental' and physical agility is very dependent on sensors, and automatic guidance of robots using image analysis will be an important feature in the unstructured environments that are typical of agricultural operations.

Robotic handling will set major demands for image analysis and computation if accurate control is to be achieved. As an example, the image analysis task is likely to be crucial to the success of robotic mushroom harvesting (Tillett and Reed, 1990). Identification of mushrooms from the growing medium is relatively straightforward, given the high contrast between the growing medium and the mushrooms. Locating each mushroom and defining its size is made easier by the regular shape of the mushroom. When viewed and lit from above the mushrooms appear brightest at the centre, and the brightness decreases towards the edges as the mushroom surface becomes more vertical. The next mushroom to be picked is determined by analysing sizes against a target grade and assessing the optimal order in which to harvest closely packed mushrooms if damage to adjacent ones is to be minimised.

Image analysis can have a more extensive role in the control of the robotic process. Thus, in the automated system for processing geranium cuttings for propagation described by Simonton (1991), the image data is classified into position, orientation and interconnection of the primary plant parts so that various processing decisions can be made. The grade of cutting must be determined, the grasp location for the robot end-effector must be chosen and a decision on the need for leaf removal must be made. Coordinating robotic systems that are robust and operate with the rapidity necessary for practical processes will continue to demand advances in computing systems and algorithms.

III. CROP MODELS AND OPTIMISATION

Agricultural research has made considerable progress over many years in the use of computers to analyse and predict crop growth and crop response to the environment. Crop simulation models have been constructed from mathematical descriptions of component processes, like radiation interception by the leaf canopy, photosynthesis and dry matter growth. It is now becoming apparent that these models can have direct application to the farming industry.

Computer models of field crop performance

Models of field crop growth and development have been extensively researched, but few have been used yet for real-time prediction of crop performance. This reflects the complex interaction between biological processes and variables defining soil and atmospheric environment of the crop. Some component processes, like crop development, have been described in models developed for prediction (eg. Travis et al., 1988), and such models are now being used as part of crop consultants' advice to farmers, for example on the best time to apply fertilisers.

More complete models of crop performance have proved difficult to construct with sufficient stability in relation to soil or weather factors to provide convincing predictions to farmers. The wide range of factors that can affect crop growth, and the problems associated with defining and measuring the most appropriate state variables for the process are certainly contributing factors. The GOSSYM model of cotton growth has been taken further than most, and has been interfaced with an expert system so that farmers' concerns for example on fertiliser application strategy, can be evaluated through simulation under the control of the expert system. This allows intelligible answers to be given to the farmer (McKinion et al., 1989), who would not wish to work directly with the simulation model.

It is likely that this sort of approach will eventually be of direct importance in farm management. At present, knowledge of the detailed interactions between growth processes is incomplete and methods to allow for the uncertainty in the definition of conditions and in the models themselves are also lacking.

A major computer application in relation to crop performance will be in predictions of the effect of changes in environmental conditions or government policies on crops and cropping systems. There is widespread concern about the effects of climate change on agricultural systems. Experimental studies of field systems are both difficult and costly, and in some circumstances may well require the guidance of model predictions before a sensible experimental design can be finalised. Other environmental issues, like the desire to reduce nitrate emissions from farming systems and to decrease use of pesticides, can also be reviewed through the use of crop models.

These evaluations can be expected to increase, though there is still only very limited validation of the capability of such models to represent accurately the responses to variations in field conditions. The issue of model validation must continue to be stressed as an essential parallel activity to model–development, as ever more extensive policy issues, like the review of national agricultural capability (Maracchi et al., 1988), are tackled with crop models. If confidence in the use of computer models can be developed in a sensible way, then quite complex issues can be tackled. Thus both from an agricultural and an environmental point of view, it is important that the use of animal wastes and other organic materials as a source of crop nutrients is based on a sound understanding of all the competing issues. These include amounts and availability of nutrients to crops, gaseous emissions of spread waste and loss of nutrients to water courses. Models of crops and cropping systems are likely to be important in developing a real scientific basis for these practices.

Models for optimal control of the crop environment

In greenhouse crop production, many of the environmental variables that make prediction of responses of field crops so difficult are held constant. The models that are required to define key environmental responses may therefore be simplified or possibly omitted entirely. It is therefore not surprising that greenhouse crop models are likely to be the first to be used in real–time control of crop management.

The physical environment within greenhouses has been measured, interpreted and controlled for many years (Bailey, 1991). Information from the plant that is indicative of its physiological state has been defined as an additional source of knowledge, opening up the 'speaking plant' approach to control (Hashimoto et al., 1981). Thus changes in leaf temperature can be monitored and interpreted in terms of the changes in stomatal aperture caused by the onset of water stress (Hashimoto, 1982). Leaf temperature could then be used to control irrigation, so that crop performance is maintained at a high level.

More complex processes can be considered as part of the control strategy, now that computing power is increasing to the point where models of crop response to the environment can be evaluated in real time. This will permit more complicated issues to be incorporated in the control strategy. Dynamic CO_2 optimisation serves as an important example, where the benefit to be gained from enhanced CO_2 concentrations can be determined by evaluating models of photosynthetic response to light and CO_2, conversion of photosynthate to growth and yield, and ultimately to additional crop value (Day et al., 1991; van Henten and Bontsema, 1991). The cost equation must relate the value of additional yield to the price of CO_2 gas and the rate of supply necessary to maintain an enhanced concentration, which will depend directly on the greenhouse ventilation rate. Some major simplifying assumptions about the processes may be involving to allow realistic real time optimisation of these models, but the finite response time of the greenhouse system to changes in CO_2 enrichment is of the order of 20 - 60 minutes, (Chalabi and Fernandez, 1992), providing a significant time window for processing. Currently available personal computers should be capable of providing optimal CO_2 set points to greenhouse climate control systems to allow these optimisation concepts to be tested soon at least in the context of experimental systems.

Practical realisation of these opportunities for optimal control in the greenhouse will require novel approaches to defining the process and the constraints on optimisation. Challa and Straten (1991) have pointed out the complexity of the problem, which requires information from diverse sources, including climate conditions inside and outside the greenhouse, forecasts of weather and market prices, and from the grower, on his policy and personal assessment of control criteria, forecasts and crop condition.

IV. INFORMATION TECHNOLOGY ON FARMS

The development of Information Technology (IT) - the use of computer-based systems to acquire, store, process, transform and disseminate data and information - has had major impacts on all industries. IT has made some inroads into farming, but the major computer application in this area is just beginning.

Moncaster (1988) has described the IT farm as a comprehensive and integrated management information system. Some components already exist as farm business management packages, bureaux services for market price information, weather and disease forecasts, and some automatic data collection, eg. in dairy farming. If we look into the future, these components will be supplemented by others, and integrated into a system that can provide information for both strategic and tactical decision-making on the farm.

Within the IT farm, some components will operate on quite short timescales and have close control over processes. The protected crops industry provides a current example of IT systems in direct control over a farming process. The greenhouse control computer, gathering information from sensors, makes decisions on operational control of heating and ventilation systems to maintain an effective growing environment for the crop. Developments of this implementation of IT, to include real time optimisation of system models defining the optimal environmental setpoints for current weather and market conditions, have been discussed earlier.

The opportunities for direct control in field agriculture are not so well formed as yet. The concept of selective or spatially variable field operations is an early practical example. The approach involves tailoring operations to meet the requirements of soils or crops in parts of the field, rather than applying the same uniform operation over the whole field.

Information technology systems will also be important to the extension worker, who will continue to provide an effective interface between the farmer and many new technologies. They can provide him or her with rapid access to management information, forecasts and the results of research and development work, so that his decisions are up to date and based on the knowledge of many experts. Expert systems will be particularly important in this role.

Selective field operations

Farming operations can be managed to cope with area units over an enormous range, from a fraction of a hectare to many hundreds of hectares. In the large fields that are common in many intensive production systems, many factors can vary across the field. There will often be more than one soil type; there may be considerable variation in slope and in water table height; long term weed problems may exist in some parts of the field but not in others. These variations could be taken into account to ensure that operations are closely tailored to the requirements of each part of the field if information gathering and interpretation can be linked to operational control systems. Benefits would be both economic and environmental. If soil conditions and fertiliser supply can be matched more closely to the crop's needs, more cost effective production will result. This will often be achieved by reducing input costs and reducing the use of agrochemical inputs. This will be beneficial in meeting environmental demands on agriculture, which now increasingly form part of the legislative framework in developed countries.

Selective field operations depend upon making decisions about field management on a local basis, and controlling farm machinery in order to implement these decisions. The prime components in the execution of selective operations in the field will be quantifying the key parameters that vary across the field, interpreting their influence on the optimal form of the field operation and then effecting appropriate control of this operation throughout the field. In order to reach sound decisions about the most appropriate level of an operation, it may well be necessary to integrate a number of pieces of information, which reflect both current conditions and past performance of specific parts of the field. Thus databases that can be called on both in the execution of selective operations and in reviewing their effectiveness by assessing historical changes are likely to feature strongly in this area.

Spatially selective application of herbicides, or "patch" spraying is an example that is currently being researched. Its feasibility depends on whether weeds grow in patches or are uniformly spread throughout fields. Some studies of cereal fields in the UK (Marshall, 1988; Wilson and Brain, 1990) have shown that some grass weeds do grow in persistent patches. Though a field-wide uniform spray may be necessary to cope with broad leaved weeds early in the season, subsequent sprays to tackle persistent weeds, particularly grass weeds, could be applied to limited areas. The environmental benefit of significant reductions in total herbicide use is matched for the farmer by substantial cost savings, as the chemicals to tackle persistent weeds are often expensive. The necessary techniques to control the application of agrochemicals dynamically during spraying operation are now available in injection metering systems (eg. Frost, 1988). The concentration of active ingredient can be adjusted, either in relation to vehicle forward speed (to maintain a uniform dose rate) or in response to other information, for example on the need for agrochemical in that part of the field.

Many aspects of advanced computer technology will need to be developed if this opportunity for dynamic control within fields is to be realised. Image analysis techniques may be able to identify weeds and quantify weed numbers in real-time as part of a sprayer control system. However the demands on resolution and interpretation are enormous, particularly given the low population level of some weeds that can significantly decrease yields (eg. cleavers at 1 to 2 plants m^{-2}) (Thompson et al., 1990). An alternative method would be to build a historical database from information gathered at times when weeds are particularly distinct. For example, the contrast between weeds and crop immediately prior to or after harvest could permit rapid interpretation of aerial photographs. When supplemented with direct observation by farmers, this could provide the essential data for controlling the application for the problem weed in the next season. This would depend upon the spatial stability of the weed patch, but there is growing evidence for a number of weed species that stability is considerable (G. W. Cussans, personal communication, 1992). The database must then be available in the form of a map, and there must be a means of locating the sprayer vehicle in the field. This calls for position fixing technology, based on satellites or other communication systems. The information on the map can then be turned into a decision on spraying (Stafford et al., 1991).

This type of system can provide a mechanism to respond to increasing demands for lower applications of agrochemicals, by targeting them more carefully. It will be demanding of many features in computer technology, and will form a challenging application in coming years. Other angles on more precise targeting of treatments can be expected to increase. 'Spot' treatment, even requiring identification of the treatment site within an individual plant, may be possible, for both chemical and mechanical processes.

It is also feasible that, for some field operations, all the decisions could be made in a purpose-build unmanned machine. Such autonomous guided vehicles (AGVs) can provide a platform for a range of actions, utilising intelligent machines to maximise the accuracy and the effectiveness of the task.

The expert system link from research to practice

Information technology will be essential to the development of rapid and efficient systems of providing agricultural users with all the knowledge and data they need to make the most cost effective decisions for their businesses. Strategic decision making, particularly in areas where research is bringing rapid advances, will be as important a goal for IT systems as the more obvious tactical and operational uses. The extension worker will have a continuing role at the interface, but needs to be able to call rapidly on quantitative information, and tailor it to the user's particular requirements. Expert systems provide the essential elements for this knowledge transfer, and will be a major area for application of computer systems.

The expert system provides a heuristic approach to problem solving, utilising judgemental reasoning in conjunction with formal reasoning. The expert system can also be transparent, explaining and justifying the line of reasoning. By separating the knowledge base from the inference procedures, updating the expert system to encompass new knowledge can be particularly straightforward. The expert system can also respond directly to the available information, using it to define the most appropriate path or to pinpoint missing information that will have a major influence on the final decision. The qualitative routes of the expert system can also be coupled to quantitative calculations, of considerable importance to defining the cost effectiveness of the solutions available. All these issues have particular relevance to agricultural problems, and will find solutions for the agricultural situation particularly demanding. The importance of uncertain factors such as weather on process performance, the annual cycle of many of the processes, and the need to interface business decisions with ever-increasing governmental and environmental constraints will all pose technical and other challenges to the development of appropriate IT for agricultural systems.

Decisions on the utilisation and disposal of animal wastes, particularly to meet problems of pollution and nuisance form a key example. There is already a wide range of constraints on the farmer yet he may still be required to adapt as potential pollution or odour nuisance problems arise, or legislation changes. Research has provided quantitative information on the capabilities of different engineering approaches for the handling and treatment of wastes, and current strategies of advice can elicit the options available to the farmer to react to particular problems. The expert system can provide the effective means of harnessing computer technology to link these information sources together, and provide front line expertise to the planning process for the farmer (Beaulah et al, 1990; Cumby, 1992).

V. CONCLUSION

The importance of information and its interpretation to the success of farming systems has long been recognised, and the crucial role of the extension worker bears witness to the commitment of the industry to obtaining and utilising necessary information for many of the longer term decisions. But computer technology can satisfy enormous numbers of other opportunities for informed decision making, particularly where specific and rapid decisions are required utilising local information, or where the extent of the relevant knowledge is too great for any individual. There will be many challenges to overcome if computers are to achieve their potential in this application area. Cost is obviously a significant factor in a primary industry that is of such an extensive nature. Technical issues relating to the variability of biological systems, the difficulty of sensing some of the key attributes that determine process performance, and the unpredictability of future conditions that will affect both the process and product value will also present major opportunities for research to contribute.

ACKNOWLEDGEMENT

This paper is an expanded and extensively modified version of a paper (Day, 1991) previously published in the proceedings of an IFAC Workshop.

REFERENCES

Bailey, B.J. (1991). Climate modelling and control in greenhouses. In J. Matthews (Ed) "Progress in Agricultural Physics and Engineering", CAB International, Wallingford.

Beaulah, S.A., Brewer, A.J., Hall, C.A., Cumby, T.R. and Phillips, V.R. (1990). The Waste Engineering Expert System (WEES): An expert system applied to pollution from animal wastes. Part I: Initial developments and programming aspects of the odour control and aeration modules. Divisional Note DN 1575, AFRC Institute of Engineering Research, Silsoe.

Ben Hanan, U., Gutman, P.-O. and Peleg, K. (1991). Classification of apples with a neural network based classifier. In "Mathematical and Control Applications in Agriculture and Horticulture" Y. Hashimoto and W. Day (editors) IFAC Workshop Series, 1991 Number 1, Pergamon Press, Oxford.

Chalabi, Z.S. and Fernandez, E. (1992). Spatio temporal responses of a glasshouse to gaseous enrichment. Journal of Agricultural Engineering Research, 51, 139-151.

Challa, H. and van Straten, G. (1991). Reflections about optimal climate control in greenhouse cultivation. In "Mathematical and Control Applications in Agriculture and Horticulture" Y. Hashimoto and W. Day (editors) IFAC Workshop Series, 1991 Number 1, Pergamon Press, Oxford.

Cumby, T.R. (1992). Expert systems to tackle intensive livestock production (abstract) Journal of Agricultural Science, Cambridge, 118, 397-398.

Davis, P.F. (1991). Orientation independent recognition of chrysanthemums nodes by an artificial neural network. "Computers and Electronics in Agriculture", 5, 305-314.

Day, W. (1991). Computer applications in agriculture and horticulture: a view. In "Mathematical and Control

Applications in Agriculture and Horticulture" Y. Hashimoto and W. Day (editors) IFAC Workshop Series, 1991 Number 1, Pergamon Press, Oxford.

Day, W., Chalabi, Z.S., Bailey, B.J. and Aikman, D.P. (1991). The path to dynamic optimization of carbon dioxide in the greenhouse. In "Mathematical and Control Applications in Agriculture and Horticulture" Y Hashimoto and W Day (editors) IFAC Workshop Series, 1991 Number 1, Pergamon Press, Oxford.

Frost, A.R. (1988). Methods of controlling the application rate of chemical from a crop spraying machine. In S.W.R. Cox (Ed.) Engineering Advances for Agriculture and Food Butterworths, London. pp 61–62.

Hashimoto, Y. (1982). Dynamic behaviour of leaf temperatures: a review. Biological Science (Tokyo), 34, 68–75.

Hashimoto, Y., Morimoto, T. and Funada, S. (1981). Computer processing of speaking plant for climate control and computer aided plantation (computer aided cultivation). Acta Horticulturae, 115, 317-325.

Maracchi, G., Conese, C., Miglietta, F., Bacci, L. and Parikh, J.K. (1988). An information system for agricultural productivity. In J.K. Parikh (Ed.). Sustainable Development in Agriculture. Kluwer Academic, Dordrecht. pp 59–97.

Marchant, J.A. (1990). Computer vision for produce inspection. Post Harvest News and Information, 1, 19–22.

Marshall, E.J. (1988). Field scale estimates of grass weed populations in arable land. Weed Research, 28, 191–198.

McKinion, J.M., Baker, D.N., Whisler, F.D. and Lambert, J.R. (1989). Application of the GOSSYM/COMAX system to cotton crop management. Agricultural Systems, 31, 55–65.

Moncaster, M.E. (1988). Introduction to the application of information technology to agriculture. In S.W.R. Cox (Ed.) Engineering Advances for Agriculture and Food. Butterworths, London. pp 321–326.

Simonton, W. (1991). Automated plant handling and processing in a robotic work cell. In "Mathematical and Control Applications in Agriculture and Horticulture" Y. Hashimoto and W. Day (editors) IFAC Workshop Series, 1991 Number 1, Pergamon Press, Oxford.

Stafford, J.V., Miller, P.C.H. and Ambler, B. (1991). Mapping weedy selective herbicide application. Proceedings of conference "Locating systems for agricultural machines", Godollo, Hungary.

Thompson, J.F., Stafford, J.V. and Miller, P.C.H. (1990). Selective application of herbicides to UK cereal crops. American Society of Agricultural Engineering paper 90-1629.

Tillett, R.D. (1991). Image analysis for agricultural processes: a review of potential opportunities, Journal of Agricultural Engineering Research, 50, 247-258.

Tillett, R.D. and Reed, J.N. (1990). Initial development of a mechatronic mushroom harvester. In Proceedings of International Conference, "Mechantronics: Designing Intelligent Machines", Institution of Mechanical Engineers, London pp 109-114.

Travis, K.Z, Day, W. and Porter, J.R. (1988). Modelling the timing of early development of winter wheat. Agricultural and Forest Meteorology, 44, 67-79.

Van Henten, E.J. and Bontsema, J. (1991). Optimal control of greenhouse climate. In "Mathematical and Control Applications in Agriculture and Horticulture" Y. Hashimoto and W. Day (editors) IFAC Workshop Series, 1991 Number 1, Pergamon Press, Oxford.

Wilson, B.J. and Brain P. (1990). Weed monitoring on a whole farm - patchiness and the stability of distribution of Alopecurus mysosuroides over a ten year period. Proceedings of the European Weed Research Society Symposium, Helsinki.

SECTION 4
Relating Plant Sciences

Chapter 11

ENVIRONMENTAL SIGNALS, RECEPTORS, PHYSIOLOGICAL AND GENETIC REGULATION IN PLANTS

Aubrey W. Naylor

Department of Botany
Duke University
Durham, North Carolina

I. INTRODUCTION

The development of organized form and behavior in living things presents many inspiring intellectual, philosophical and experimental challenges. How this is achieved in the simplest to the most complex life form is a many faceted question that until recently has been out of reach of experimental explanation. Advances in methodologies in physics, chemistry and biology have at last reached the point where theories and hypotheses can be rationally formulated, tested, recast and refined. There are still many, many undiscovered pieces to be identified in this jig-saw puzzle, but, at last, progress toward understanding is on an ever steepening incline.

All organisms are the vectorial resultant of the interplay of their genetics and their internal and external environment. In the achievement of growth and form there is a bewildering array of control points. These are ever present at the environmental, genetic and biochemical levels. Without sensory molecules and systems, basic to directed responses, we could not have distinctly different forms and functions among plants and animals - only amorphous protoplasm. There could be no coordinated massing of functions we casually associate with cell walls, membranes, cytoskeleton, nucleus, endoplasmic reticulum, chloroplasts, peroxysomes, mitochondria, chloroplasts, golgi apparatus and vacuoles. Signals for the production of all of these cell components must be highly coordinated if the cell is to survive in nature. Failure or even momentary disruption of the orderly sequencing of internal and external signals to the genetic blueprint can lead to the cessation of growth, lack of necessary repair of components of the metabolic systems, disarray of chromosome behavior in nuclear division, leakage of crucial metabolites by membranes and even death. Yet slight differences in timing in coordination of chemical and physical signals to the genetic dictionary and metabolic pathways is necessary if cells are to become different from one another and distinct tissues such as food conducting and water conducting tissues are to be produced. A still higher level of coordination of signal response and selectivity in use of genetic information is required in the generating roots, stems, leaves, flowers, fruits and seeds. Obviously, an understanding of the many factors underlying initiation of the differentiation process is necessary if we are to exercise control over it.

II. SIGNALING AND ITS RECEPTION

All living cells have a common requirement to sense their environment and respond to it appropriately. This point is emphasized spectacularly when the Venus fly-trap senses the presence of a fly on its leaf. A chain of reactions is set sequentially in operation. The two halves of the leaf fold closed swiftly enough to capture the fly. This is followed by specialized cells secreting hydrolytic enzymes that digest the fly's soft body parts. The resulting compounds are sensed and absorbed through the epidermis. Thus, through a sequence of signals, the Venus fly trap is supplied with preformed amino acids, carbohydrates and lipids. A less dramatic example is seen in a stem's apex. If the local internal and external environmental signals are sequenced in one way meristematic cells will divide and grow in size producing stems and leaves; with a slight change in the signals the planes of division may change from random to highly directional resulting in sepals, petals, stamen and pistils or the cells may cease to divide and go into dormancy.

Physical and chemical signals are ever present activating, inhibiting, neutralizing and stopping metabolic activities and often altering molecular populations in the cell at the level of the gene causing cells to respond in hundreds of different ways. Keeping all of these communications precisely timed, and coordinated is a formidable task; and finding out how it is done is one of the most fascinating challenges in science. If the communications go wrong a teratoma, failure to reproduce or death can swiftly follow.

Early in evolution, single celled organisms developed ways to identify crucial signals in their external environment and get that information across their cell membranes into their cytoplasm and subsequently to DNA - the genetic code. These communication routes appear to have been largely retained by multicellular organisms through evolutionary history. The question then becomes one of how are the physical and chemical signals perceived, and then how does transduction occur and what are the consequences? The principal environmental factors impinging on plants are light, temperature, water, minerals, pH and gravity. How does the plant perceive these signals?

LIGHT

Light is perceived by at least three different pigment systems. One of these, chlorophyll, is an energy converter the other two, phytochrome and cryptochrome are especially important in photomorphogenesis. Chlorophyll and its associated proteins makes it possible for the energy of blue and red light to be converted to chemical energy - the high energy compounds being NADPH and ATP. It is the energy stored in these compounds that drives most all of the life processes of unicellular and multicellular plants and animals.

Phytochrome is activated by red light and transformed to a molecular form that has a multitude of actions controlling a wide range of developmental processes such as initiation of germination, promotion of formation of leaf primordia, expansion of leaves, chlorophyll accumulation, stem elongation, flowering and senescence depending upon the tissue and cell type in which it finds itself. Upon activation phytochrome is thought to involve calmodulin, a transducer of small Ca^{2+} concentrations, in major physiological changes. Activated calmodulin can bind and control a wide range of regulatory enzymes in cells. Six of these are already known. More on this point later. Thus far, genes for several phytochromes have been isolated and their products characterized. Additionally, progress has been made in learning how these genes receive their signals and are regulated (Furuya, 1987; Kay, et al. 1989).

Far less is known about how cryptochrome, a blue light activated pigment, transduces the energy it receives than is known about phytochrome. It is clear, however, that a train of reactions is set in motion by blue light resulting in many apparently disparate responses, such as growth of stems toward a light source, orientation of leaves and opening of stomatal pores in leaves (Singer, 1987). The latter phenomenon comes about through activation of a kinase driven proton pump in the plasma membranes of the guard cell pairs - ultimately resulting in an increase in turgor of these cells. As with phytochrome, there is evidence more than one cryptochrome exists (Konjevic, et al. 1992). Evidence is also accumulating indicating that this pigment is a 120K Da plasma membrane bound flavoprotein that upon exposure to blue light is activated by binding with gnanosine triphosphate (Warpeha et al. 1992; Short, et al. 1992).

TEMPERATURE

Temperature is sensed by every chemical and chemical reaction in the cell. Ions, atoms, molecules, and supra-molecules are all affected. Life exists generally speaking, in unstable equilibrium with the environment within the narrow temperature range of 0° to 85°C. Two basic concepts: "rate effects" and "weak bond" effects are keys to controlling metabolic pathways and the complex biological molecules on which those pathways and their chains of reactions are dependent.

From the standpoint of rate effects, changes in the average kinetic energy of the atoms and molecules of the organism, that is, the temperature of the organism, will be translated into changes in the rates at which the chemical reactions comprising metabolism occur. This would include respiration, photosynthesis, nucleic acid synthesis, protein synthesis, lipid synthesis and the staggering array of other reactions required in making and maintaining the multitude of constituents of the new and old protoplasm. Since all of these diverse reactions with their differing rate maxima must be coordinated in a highly structured time frame, it is readily understandable why the temperature window for life is so narrow. One illustration will suffice. It is well known that rye is a cool weather crop growing best at about 22°C. An important reason for this is that at elevated temperature (32°), plastids become deficient in chloroplast ribosomes, chlorophyll synthesis fails and ribulose-bisphosphate carboxylase-oxygenase fails to be regenerated as the leaves become chlorotic (Feierabend and Schrader-Reichardt, 1976; Rademacher, and Feierabend, 1976). In wheat, it has been found that at the time the number of chloroplasts per cell reaches a maximum the amount of chDNA declines to about a fourth of amount present at max (Boffey and Leach, 1982). Subsequently, there is no chloroplast replication. At the other end of the temperature scale, a maize mutant has been found that lacks the ability to produce chloroplasts ribosomes at 20°C and does not become green. But at 30°C the plant grows normally (Hopkins and Elfman, 1984). Earlier Pringsheim and Pringsheim (1952) reported the creation of an albino strain of *Euglena gracilis* by culturing at elevated temperatures for a prolonged period. Electron micrographic investigations (Kivic and Vesk, 1974) revealed that a genetic lesion prevents plastid differentiation beyond the proplastid stage. Presumably heat accelerated RNA-ase and DNAase to a greater degree in the chloroplast than in the nucleus thereby destroying the chloroplast without killing the cell.

The cause of the narrowness of the life permissible temperature window becomes even clearer when the importance of weak bonds, meaning van der Waal forces, hydrogen bonds, ionic bonds and hydrophobic interactions in living organisms are taken into consideration. It is at the molecular and macromolecular levels that the crucial relation between biological structure and function can be most appreciated and visualized.

Virtually all of the higher orders of bio-molecular structure as seen in the 3° and 4° structure of proteins, glyco-proteins, membrane structure, nucleic acid structure and most of the biochemical interactions demanding a high degree of stereochemical specificity (e.g. the binding of substrates of enzymes) are highly if not entirely dependent on weak bonds. Many enzymes are made up of two or more subunits. An example of this is ribulose bisphosphate carboxylase-oxygenase, the enzyme of paramount importance in CO_2 fixation. This enzyme consists of eight large and eight small subunits. These subunits must be precisely arranged if maximum photosynthesis is to be achieved. They are assembled by means of mutually attractive forces and through the aid of a protein called chaperonin (Gatenby and Ellis, 1990). These 16 molecules are held together as a single macromolecule by means of multiple weak bonds without the aid of a single covalent bond. Understandably, the molecule is sensitive to temperature changes, pH shifts, dissolved gases, and ionic composition of its surroundings. Thermal injury to the enzymes and cytoskeleton microtubule proteins can also be traced to loss of their tertiary or quarternary structure as weak bonds give way (Carter and Wick, 1984; and Guy 1990).

It is a common observation that as the temperature is raised or lowered beyond the optimum for growth the spectrum of newly formed proteins in cells change. This automatically alters the favored metabolic pathways. Ultimately, one or more metabolic component(s) will become limiting and life threatening. Such points establish the upper and lower temperature limits for each plant species or variety. Often these limits can be extended through a process of acclimation called cold or heat hardening. Many temperate zone but few tropical plants have the ability to acclimate. Sometimes this failure is caused by the presence of cold-labile enzymes that literally dissociate their sub-units at non-freezing temperatures because of excessive hydrophobic bonding (Guy, 1990 see Table 2). Those plants that are successful in acclimating to low temperatures show profound changes in the lipid and protein components of their membranes and cytosol. Intermolecular disulfide bonding becomes prominent. Important alterations also occur in protein census and configuration (Huner, et al. 1982; Mohapatra, et al., 1987). Conformational changes in enzymes may distort the active site sufficiently to cause partial to total loss of activity.

pH AS A REGULATORY FACTOR

Both external and internal pH can have profound effects on plant growth. Root growth is generally most favorable at slightly acidic values - in the range of 5.5 to 6.5. Roots have a tendency to acidify the solution that bathes them because of proton exchange which accompanies the uptake of cations such as $(NH_4)^{2+}$, K+, Ca^{2+}, and Mg^{2+}. Not unexpectedly, the microflora adjacent to roots varies change as the pH changes. Fungi generally are dominant in the acid pH range while at higher pH values bacteria become dominant. Parenthetically, one of the serious deleterious effects of acid rain is a decrease in soil pH with a concomitant rise in the population of root pathogenic fungi. Fluxes in pH of the root zone considerably affect the solubility and mobility of minerals and hence their uptake. Phosphorus, for example, is available only over the narrow pH range 4.0 – 7.5 and this ion is available only after the roots have acidified their surrounding water layer (Olsen et al. 1981). pH regulatory effects begin within the plant at the interface of the plant with its environment. Because of the ionic and electrical properties of cell walls, which ordinarily contain a high concentration of polygalacturonic acid, roots selectivity accumulate cations. This makes more difficult entry of anions into the cell wall matrix. The plasma membrane provides a still more formidable regulatory barrier with its array of anion/H^+ symports, H^+–ATPase (proton pumps), cation transport channels and sugar H^+/symports. These membrane components make it possible to have cell compartments that are maintained at different pH levels. Compartmentation of metabolic processes like photosynthesis, respiration, protein synthesis etc. is essential because each process has it own optimum pH. John Raven (1985) has provided us with the pH values of the major compartments of green plant cell in the light and in the dark. In the light, the chloroplast has a pH as low as 4.0 but rises as high as pH 7.8 in the dark, meanwhile the mitochondria operates at a pH of 7.6 and the vacuole flucuates around pH 5.0. At the same time the roots are bathed in a soil solution at pH 6.8.

Enzymatic action is very sensitive to pH. This sensitivity is inevitable because the catalytic groups (imidazole, carboxyl, amino) are catalytically functional in only one of their ionization states. Typically pH plots of the rates of enzyme–catalyzed reactions are bell shaped, corresponding to two sigmoid curves, one of the ionizable groups acting as an acid and the other as a base. While pH effects on the efficiency of enzyme catalysis usually are a manifestation of the ionization charactertistics of the catalytic groups, there is another aspect to the matter – pH affects conformation of the enzyme. These changes can be sufficient to completely distort the active site rendering it inoperable.

Numerous enzymes occupy more than one compartment. These are usually isozymes. Isozymes are enzymes with similar catalytic functions but are encoded by different genes. Their pH optima often differ; they may exhibit temporal and tissue specificity. They can also play a role in developmental regulation and play a role in adaptation to environmental change.

WATER

Water typically constitutes 80–95% of the mass of growing tissue. Even the stems of old trees may around 30 percent water by weight. Among the driest plant structures are mature seeds and even they die if their water content drops below 5 to 15 percent. All life as we know it is dependent upon water and there is no substitute for it.

Water serves life in many ways. It is the most abundant and best of solvents. In this role it is involved in solubilizing soil minerals and serves as a vehicle of transport into and throughout the plant. In this role it greatly influences the molecular structures and properties of proteins, membranes, nucleic acids, and other cell constituents. It is also the medium in which most of the biochemical reactions occur. Furthermore, it participates in a number of chemical reactions such as hydration, dehydration and hydrolysis. It is also a truism that for every carbon atom processed in photosynthesis a molecule of water is used as is made evident in the formula for glucose ($C_6H_{12}0_6$). Glucose is a building block that can be used as the sole carbon source for many organisms.

Water availability is often the most important controlling factor in the establishment and maintenance of a given natural ecosystem or crop type. Its multitudinous roles makes it obvious that limitation will result in numerous alterations in mineral nutrition and metabolic activities. The ability of water to form hydrogen bonds gives rise to its extraordinary thermal, cohesive and adhesive properties. Not unexpectedly, the numerous physiological processes show differential sensitivity to water status of the plant. T.C. Hsiao and Acevedo (1974) have expressed graphically the effect of varying water potential on hormone balance, solute uptakes photosynthesis, stomatal opening, cell expansion, and protein synthesis. In order of sensitivity to water loss from the intact plant was 1) cell expansion, 2) wall synthesis, 3) protein synthesis, 4) stomatal conductance and 5) photosynthesis. All of these functions were completely shut down by mild stress conditions, i.e., –2 MPa. Under identical conditions, however, there was a steady accumulation of abscisic acid and solutes.

Knowledge of the contribution of water to the energetics of assembly and disassembly of carbohydrates, proteins and lipids has a long history. As a result of the elegant quantitative work of Colombo, Rau and Parsegian (1992) at the molecular level we now know that hydration/dehydration reactions contribute far more to the conformational changes of enzymes than previously thought, and, indeed, water can behave as a ligand providing allosteric effects as a result of hydrogen bond binding. This has been illustrated with three widely different proteins (Colombo et al. 1992; Kornblatt and Hoa, 1990; Rand and Fuller, 1992) - hemoglobin, cytochrome oxidase and hexokinase. Hemoglobin hydrates, while cytochrome oxidase shows a hydration-dehydration cycle, and hexokinase dehydrates. In each instance a number of water molecules are involved in the conformational changes (Rand, 1992). V.A. Parsegian and members of his laboratory have shown that forces of surface-perterbed water between phospholipid layers and between macromolecules as little different from bulk water as a fraction of a calorie/mole, can dominate interaction energies when many molecules of water are involved. For example, in the process of substrate binding, glucose hexokinase releases about 100 molecules of water, a total mass 10 times that of glucose. The net result is a tight fit between enzyme and substrate (Rand and Fuller, 1992) thereby increasing the efficiency of the process. Observations such as these make it possible to understand how mild water deficiencies can cause the drastic effects mentioned above.

GRAVITROPISM

Late in the 19th century, Charles Darwin together with his son Francis concluded from their experiments with roots that perception of gravity occurs in their tips (Darwin, 1880). It was then reasoned that some influence was transmitted up the root to the zone where actual bending occurs. Since Darwin's time, fundamental progress on the phenonmenon of gravitropism has been slow. One conclusion generally agreed upon has been that the point of perception is limited to a small area in the root cap. The obvious next question is how is the unilateral gravity signal perceived.

Mutations of genes relevant to growth response to gravity make it clear the responses of roots and hypocotyls can be cleanly separated (Mirza et al., 1984). Likely, there are at least two means of detecting gravitational forces. The one that is most obvious and most studied is achieved through the gentle pressure of starch grains settling under the influence of gravity on endoplasmic membranes as root or shoot reorientation occurs following a change in position. Gentle as these pressures are they are now known to be sufficient to cause Ca^{2+} to be expelled into the cytosol setting up an electrochemical gradient (Braam and Davis 1990; Knight et al., 1991; Trewavas and Gilroy, 1991). Strong evidence is accumulating that the cytoskeleton is involved (Sievers et al. 1991). This is probably done by the statoliths changing position and triggering the transduction mechanism via actin filaments. Quickly, this results in an assymetry in the signalling system in the root cap that is transmitted to the elongating cells of the root several cells away where asymmetric cell elongation reorients the root until uniform signaling is restored around the circumference of the root. Meanwhile, the root becomes oriented downward once again.

III. CHEMICAL AND INTERNAL ENVIRONMENTAL SIGNALS

Early in evolution, single celled organisms developed ways to identify crucial chemical signals in their environment and get that information across their cell membranes into their cytoplasm. These communication routes appear to have been largely retained by multicellular organisms. Several classes of information systems are used by cells: all involve a receptor. The receptor is the primary signal discriminator. It sensitively and often specifically picks up signals from the environment. Each cell must, of necessity possess many receptors. These, in turn, vary in kind and number according to the locale. Speed of response can also vary enormously.

One class of receptors, perhaps the simplest, is located in the cell's cytosol. The chemical signals activating these receptors are lipophilic easily penetrating the plasma membrane. After binding to its receptor in the cytosol the activated receptor enters the nucleus and targets a specific DNA sequence coding for a specific protein.

A second class of receptors are large molecules that span the plasma membrane. The signal molecules in this case are often water soluble and cannot pass through the cell membrane because of its large lipid component. When one of these binds to the surface exposed segment of the receptor on the outside of the cell a change in the shape of the receptor molecule occurs sufficient to initiate further reactions involving the portion of the receptor that is inside the cell. In this case the signal molecule remains outside the cell.

The third class of receptors is made up of two important variants of the membrane spanning signal receptors. One of these functions as an enzyme family. Members are tyrosine kinase enzymes that calalyze phosphorylation of tyrosine in certain key proteins. These proteins are the receptors triggered by growth factors and hormones. A second variant serves as a regulatable pore in the plasma membrane. This membrane spanning receptor contains within it a channel that can allow passage of small-cations-such as sodium or potassium-into or out of the cell. This movement is fundamental to the generation of electrical currents. Most of the time this channel is closed, but when a proper signal molecule binds to it a configuration change occurs resulting in opening of the channel.

A fourth receptor system is the most complex. It consists of receptors that lack intrinsic activity as either enzymes or ion pores. These are trans-membrane receptors coupled to enzymes in the membrane by a class of proteins known as guanine-nucleotide binding proteins or G-proteins.

G-coupled-protein receptors include those for red and blue light and certain hormones. Upon activation of the G-protein by a signal outside the cell, the guanosine diphosphate (GDP) moiety becomes phosphorylated yielding GTP. In the GTP form, the G-protein activates a cascade of reactions that eventually results in a cellular response. The functions vary depending on the character of the G-protein's protein moiety. Essentially what the G-protein does is either regulate the presence or activity of an enzyme or so alter membrane channels that passage of cations is permitted or stopped.

Products of the activated enzyme's action or ion fluxes (like Ca^{2+}) can serve as "second messengers" in amplifying and propagating the original signal within the cytoplasm. Sometimes the second messenger activates a protein kinase which proceeds to phosphorylate a specific protein, or the messenger may alter the activity of an enzyme or cause a previously quiescent protein to bind to a specific operator protion of a DNA strand in the nucleus thereby altering the kinds and amounts of proteins synthesized. Eventually the GTP-protein is inactivated by hydrolysis to GDP-protein. Thus it becomes available for reactivation when another signal reaches it from outside the cell. G-proteins seem to function in a regulatory way in cell functions as diverse as blue light perception, cell multiplication, secretory processes and responses to hormones.

Thus far, seven distinct G-proteins have been characterized and can be distinghished by their responses to antibodies. Many more will surely be discovered in the future. Those characterized have a lot in common. All have molecular weights of approximately

100 kD and consist of three polypeptide subunits: alpha, beta and gamma. Specifity among the G-proteins resides principally in the alpha subunit and to some degree in the gamma component.

Plants have a very important Ca^{2+} based signal transducing system involving membrane bound G-proteins and an inositol phosphate cycle. This multistep system begins with a protein signal receptor that spans the outer cell membrane. Adjacent to it on the membrane's inner face is a guanine nucleotide binding protein. The receptor, upon receiving a signal it recognizes, activates its associated G-protein which in turn activates its neighbor the enzyme phospholipase. Phospholipase then hydrolyzes a nearby phospholipidmoleculesphosphatidal-4,5-bisphosphateformingtwo second messengers - 1,2-diacylglycerol, which remains in the membrane, and inositol-1,4,5 triphosphate, a water soluble compound free to move in the cytosol. IP_3 then induces release of Ca^{2+} stores from the endoplasmic reticulum or vacuole. Factors such as light and plant hormones are known to alter levels of IP_3 and, in turn, cytosol Ca^{2+} levels in plant cells (Gilroy et al. 1990; Einsphr and Thompson 1990).

The principal receptors of Ca^{2+} signals in the cell are members of a calcium binding protein family appropriately called calcium-modulated proteins - many of these are well characterized. Their diversity helps to account for the several known Ca^{2+} responses. The calcium binding domain of these proteins bind reversibly with dissociation constants in the micro to submicromolar range thereby providing a sensitive regulatory mechanism. The importance of calcium as a factor in regulating cell growth is difficult to overestimate. It is essential in sustaining cell viability and replication. Also, progression through the cell cycle is particularly sensitive to calcium concentration.

The best characterized calcium modulated proteins in plants are calmodulin and the calcium dependent protein kinase (CDPK). Calmodulin has no enzymatic activity of its own but in combination with Ca^{2+} it has a very wide range of functions. Furthermore, it is essentially ubiquitous in distribution and loss of the gene for it is lethal (Davis et al. 1986). Calcium activated calmodulin regulates many enzymes including protein kinases and phoshoprotein phosphatases.

Diacylglycerol, which remained in the membrane following release of inositol-1,4,5-triphosphate into the cytosol, together with Ca^{2+} activates its neighboring protein kinase C whereupon it phosphorylates a specific population of intracellular proteins some of which are involved in cellular replication. Another potential role is in the regulation of the plasma membrane's proton pump, which generates a gradient of pH and electrical potential essential for cell growth. The cell membrane has been referred to as a complex switching device that accepts many kinds of signals, gauges their relative strengths then routes the signals to second messengers to the specific metabolic pathways that assure cell reactions appropriate to a changing environment.

IV. REGURATION

METABOLIC REGULATION

In his book *Cellular Energy Metabolism and its Regulation* Daniel E. Atkinson (1977) said "Regulation is not a late (evolutionary) development superimposed on metabolism after catalysis had become well established . . . Regulation is the most fundamental difference between living and non-living systems, and it must have coevolved with other properties of life from the beginning." Thus, ordering of the external and external signals that are sensed by membranes, the cytoplasm, mitochondria, plastids and nucleus is a fundamental attribute of all organisms. Flexibility in response to numerous simultaneously imposed signals is essential in maintaining an inherently unstable system – which is protoplasm.

Even the "simplest" of living organisms such as the archaebacteria are metabolically complex–containing carbohydrates lipids, nucleic acids and well over 1000 different proteins – most of which are enzymes. Some of the proteins, however, serve as food reserviors while others have a structural function making up the cytoskeleton. Indeed, it is the catalytic proteins that are ultimately responsible for the synthesis of the carbohydrates, lipids and nucleic acids and proteins including the cytoskeleton. Thus, metabolic complexity is the resultant of numerous biochemical pathways operating simultaneously. Chemical chaos would be inevitable if these metabolic processes, often competing for the same substrates, were not separated in some way. Even in what previously has been considered undifferentiated cytoplasm it is becoming clear there are distinct metabolic regions and compartments. The cell is not simply a bag of enzymes with reactions taking place purely by chance encounter. There are several levels of organization.

Living organisms are unique in that they build themselves from relatively few raw materials. This is done with energy ultimately derived from the sun and a multitude of enzymes acting on relatively few, but essential, raw materials. These catalytic proteins enormously accelerate the interconversion of small molecules, polymerize giving macromolecules, and facilitate transport of metabolities into and out of cells. They are also involved in the establishment of gradients, production of electron transport chains, replication of chromosomes, dynamics of the cytoskeleton, replacement of unstable molecules and in cellular defense against pathogens. Thus, enzymes must be expected to be ubiquitous in the organism. Indeed, they are present in cell walls, membranes the cytosol and organelles.

Organization of enzymes occurs in a hiearchy of levels. Many enzymes are highly specific in what they do, a few are multifunctional as is the nitrogen fixing enzyme, nitrogenase which has multiple catalytic domains on a single protein. This enzyme carries out the multistep process of transforming N_2 to $2NH_4^+$ – a form of nitrogen that can be utilized by the plant. At the next level of organization, one finds multienzyme aggregates. Sere (1987) uses the word "metabolon" to describe an aggregation of enzymes that carry out an orderly sequence of reactions. Metabolons may or may not be bound to a membrane. For example the entire anaroebic respiration pathway employing numerous enzymes is represented by isozyme pairs that are differentially compartmented in the cytoplasm and chloroplasts (Copeland and Turner, 1980; Gottlieb, 1982).

REGULATION BY COMPARTMENTATION

As we saw in the discussion of pH differences, from 4.0 to 7.8, in different parts of a single cell, compartmentation allows reactions with vastly different environmental requirements to take place in very close proximity. Also, division of the cell into small volumes permits the concentration of metabolic intermediates in a favorable environment for interaction. Furthermore, lengthy pathways that interact through supplying one another with substrates but requiring differing environments within close proximity to one another make possible efficient interchange. Most of the known compartments with their specialized metabolic activities are sufficiently different in appearance to be easily identifiable as organelles. The major organelles in plant cells are the nucleus, the command center of the cell; mitochondria, the center of O_2 use in respiration and the source of immediately available energy and reducing power in the cell and provider of a large number of amino acids; the Golgi apparatus, the source of cell wall materials microrvacuoles - transporters of various compounds throughout cells and through membrances to the surface; the vacuole, this may grow to occupy most of the cell. It contains inorganic ions, organic acids, sugar, many hydrolytic enzymes and a variety of secondary metabolites some of which may play a role in plant defense; plastids, depending upon their environment serve in photosynthis, carbohydrate storage, or pigment bodies; glyoxysomes, the site of lipid conversion to forms utilizable cooperatively by associated mitochondria. In addition there are rough and smooth endoplasmic reticula consisting of a network of membranes in which glycoproteins and lipids are synthesized. Many mechanical functions in the cell are reliant on another component of the cytoplasm, the cytoskeleton. This apparatus is made up of microfilaments, microtubules and intermediate filaments. Cytoskeletal filaments, aside from giving the cell strength and rigidity, control the highly choreographed chromosomal movements seen in mitosis and meiosis. All of these several compartments are enclosed by the plasma membrane which serves as the primary environmental sensing device, switching device and communicator of the cell. Surrounding the cell as a whole is the cell wall, once considered little more than a jacket but now known to contain important enzymes such as invertase, glucosidases and peroxidases plus compounds that serve as inhibitors or poisons to potential predators.

Metabolic pathways once thought to operate in single compartments are now known to involve two or three. In photorespiration, for example, the chloroplast, peroxysome and mitochondria are involved; and in lipid synthesis the endoplasmic reticulum supplies triacylglycerole while plastids supply fatty acid components. Unfortunately, as yet, relatively few facts are known about how the several specialized cellular compartments communicate with one another and coordinate their activities. Of course, when they are in physical contact communication is simplified. Equally vague, is how proteins coded for in the nucleus and assembled in the cytoplasm are targeted to the proper compartment. This holds true even for very closely related proteins such as isozymes. Still more fascinating is that after arriving in the appropriate organelle each component of protein complexes such as multienzymes become located in ordered sequence. Apparently, some ordering is achieved through physical processes of self-assembly and some with the aid of enzymes called chaperonens. Adding further to the puzzle is that many of these protein associations can be quite different in flower, leaf, stem and root tissues (Beevers, 1991). Thus, coordinating factors of organ, tissue and cellular types have significant roles to play.

Characteristically, there is continuous turnover of macromolecular cellular components. During growth, rates of synthesis obviously exceed those of degradation. Compartmentation allows these opposing processes to be independently regulated. They can be separately initiated, modulated by effectors or hormones that are targeted to specific organelles through receptors or transport systems or halted. Coordinated regulation between compartments can be achieved through substrate and or co-enzyme movements between organelles. Ion gradients driving transport or generating ATP have been observed in plastids mitochondria, plasma membranes, tonoplasts and Golgi vesicles.

V. GENETIC REGULATION

In the last analysis, all of an organism's features are encoded in the gene strings that make up the core of the chromosomes residing in the nuclear compartment. This collection of chromosomes makes up the genome. Just when and how much of the information encoded there is expressed will determine the differences between cells, tissues and organs of an organism. Since all cells, except the gametes, of an individual have exactly the same information, it is obvious only a portion of the total amount present is used in any cell. Thus, regulation and selective use of the genes is the fundamental basis underlying differentiation of cells, tissues and organs.

Gene expression results from responsive external and internal environmental signals. While many of the signals are responded to at the metabolic level, even in the absence of the nucleus, many others are responded to at the genetic level. Gene expression results from molecular activation of specific regions of the genome. These molecular interactions selectively turn on some genes and repress others. As cells assemble new combinations of molecules, using products of earlier expressed genes, subtle changes occur in the physiology and local environment. Continuous feedback of information regulates the next steps in expression of the genome and at any given time the cell or tissue's behavior reflects events that occurred in earlier stages.

Regulation of gene expression involves both qualitative and quantitative components, that is, whether or not a gene will be expressed and the length of time it is expressed. Since transfer of information from the genome to the functional protein in the cell is a multistep process, involving transcription of a DNA segment into messenger RNA, followed by translation of the message into an amino acid string and then a processing stage before the functioning protein is finally produced, there are several points where modulation can and does occur. For instance, some m-RNA molecules are ephemeral while others are long lived and may persist for months and in some seeds for years. Furthermore, many metabolic products can alter the entire gene expression process. Rapid responses to environmental perturbations involving protein synthesis usually take place without the synthesis of new m-RNA.

Adjacent cells, such as those of xylem and phloem, although they start out looking alike, become different from one another because, early on, they synthesize and accumulate different sets of proteins. Ultimately, these differences are made manifest in terms of form and function of these cells. According to the environmental history of the cell and the immediate external and internal signals impenging on the genome, specific genes can be activated or suppressed. Following selective targeting and activation of a gene, there are at least six control points for a given protein's formation before that protein becomes functional. The timing of each step is important.

Any or all of the six steps leading from DNA to functioning protein can be involved in establishing differences in population densities for a given protein in adjacent cells. Qualitative and quantitative protein differences can arise by 1) regulation of when a given gene is transcribed, i.e. transcriptional control; 2) regulation of the processing of the primary RNA transcript, i.e. processing control; 3) regulation of which of the many completed m-RNAs in the cell's nucleus are exported to the cytoplasm, i.e. transport control; 4) determining which m-RNAs in the cytoplasm will be translated into "raw" polypeptides by ribosomes, i.e. translational control; 5) selectively stabilizing specific m-RNAs against degradation by RNAses, i.e. m-RNA degradation control; and 6) final processing of the "raw" polypeptide into a fully functional form, i.e., functional control. It is at this point that "fine tuning" of enzymes takes place through interaction with metals, nucleotides, amino acids, carbohydrates and sometimes end products of enzyme action.

Gene regulatory proteins can either activate or repress gene transcription. This is the role of a large set of sequence-specific DNA binding proteins. Each of them recognizes specific DNA sequences 8–15 nucleotides long. Binding can either facilitate (positively regulate) or inhibit (negatively regulate) transcription of an adjacent gene. Good evidence that transcription control occurs during cellular and tissue differentiation comes from an analysis of m-RNA populations in different tissues of tobacco. Each tissue measured had about 25,000 to 30,000 different m-RNAs. Of these, some 8,000 were common to all tissues, while the remainder were specific to given tissues.

Gene regulatory proteins interact with one another in different combinations in regulating certain genes. This is a very effective way in which enzymatic population diversity arises foreshadowing cellular specialization. Of enormous significance is the fact not all regulatory proteins are equal. Some must be considered master gene regulatory proteins. These have a decisive role in coordinating the activity of sets of genes. Such genes are referred to as homeotic genes. An excellent example of a master gene of this type is displayed in the flowering process.

As a plant grows, the dome-like meristem at the apex of the stem rides upward as new cells are produced beneath it. On the flanks of the dome, bulges appear that grow into leaves. While the bulges are very small, cells in their axils divide and produce a new meristem that has the potential of becoming a branch. This continuous production of stem, leaves and buds is finally halted by the terminal meristem producing the parts of a flower-sepals, petals, stamens and carpels. This sudden change in the meristem's program is traceable to the action of a special factor that activates a master gene that controls whether or not a plant will flower and subsequently bear fruit. This master gene has been named *floricaula* (Coen, et al. 1990). Sequence analysis shows that *flo* encodes a protein, not yet isolated, but that is clearly unique. Based on the speed of visibly detectable action caused by this suddenly activated gene, its effects are profound. Two days after a photoperiodically sensitive plant like the snapdragon (*Antirrhinum majus*) receives the environmental stimulus to flower an abrupt change is seen to occur in the stem's developmental pattern. Evidently the flo product starts a cascade of genetic events that in rapid ordered sequence result first in the production of a whorl of sepels, then petals, then stamens, and lastly a pistle or pistles, i.e. the completed flower.

Apparently, three combinatorial gene sets are prodded into action by *flo*'s product. Timing of activation and duration of activity of each of these gene sets is crucial in the production of each of the four parts. First to go into action is gene set 1 which is responsible for the sepals; then gene sets 1 and 2 control development of the petals. Meanwhile, gene set 1 ceases to function and gene sets 2 and 3 are active in the production of the stamens; finally, in its turn, gene set 2 ceases to function while gene set 3 continues to function in the production of the pistil. Failure of any one of these sets to function or if there is mistiming of a set's expression abnormal flowers result. The master gene for each of these three sets has been identified and each possess a stretch of code typical of regulatory genes. Apparently, the proteins they encode act by binding to specific targets in the genome. Their role is to activate, in order, the several genes making up each of the three sets (Cooen, 1991). If master gene 3, for example, fails to "fire", the flower will have no pistil, or if master gene 2 fails to "fire" the flower will have no petals.

Perhaps when the protein coded for by *flo* is isolated and fully characterized it will become possible to design molecules that cause the gene to be activated at the convenience of the farmer or orchardist.

VI. CONCLUSIONS

The performance of a plant is the resultant of interactions of its genome with environmental factors any one of which may become limiting whether it be light, temperature, pH, mineral nutrients or water. All have an impact on anabolic and catabolic processes, the two major aspects of metabolism. The metabolic possess are regulated by enzymes, catalytic proteins which owe their origin to the genome. The enzymes in turn are often regulated by end products of their activity and by the presence of small organic or inorganic ions and molecules. Fine tuning of metabolic pathways keeping them from proceeding too rapidly or too slowly is accomplished to a major degree by feedback control which can work almost instantaneously and is reversible. In addition, end products may activate genes that cause production of enzymes leading along other pathways.

In the differentiation process the positioning on the plasma membrane of signal receptors of physical and chemical signals is of key importance in the establishment of differences in cells and tissues. This is a key process in the establishment of polarity. This leads to use being made of different gene sets in the genome by nearby cells. The unused gene sets may become inactivated to the point of becoming very difficult to turn back on. This may even be seen in the undifferentated cells of root and shoot meristems. A root tip will not generate leaves nor will a stem tip generate a root cap. However, once a plant has passed into the "adult" phase of its life the same meristem which formerly gave rise to leaves and axillary buds can be induced to synthesize a chemical agent that activates a genetic cascade causing the flower parts to be produced in sequence thereby using up the entire meristem.

Much has been learned about genetic and non-genetic control mechanisms operating in normally growing plants. Much of Modern agriculture is based on a conscious and subconscious use of this information in crop production. To a major degree biological engineering is based on knowledge of ways to manipulate metabolic control mechanisms. With increased information about how to regulate the operation of the genome and metabolic pathways we can expect enhanced crop production, better meeting the needs of our geometrically expanding world population.

REFERENCES

ap Rees, T. (1987). Compartmentation of plant metabolism. In D.D. Davies, (ed) The Biochemistry of Plants 12:87–115.

Atkinson, D.E. (1977). Cellular Energy Metabolism and its Regulation. Academic Press, New York.

Beevers, H. (1991). Metabolic compartmentation in plant cells. pp. 1–21 In Emes, M.S. (ed) Compartmentation of Plant Metabolism in Non–Photosynthetic Tissues. Soc. Exptl. Biology Seminar Series 42. Cambridge Press.

Boffey, S.A. and Leech, R.M. (1982). Chloroplast DNA levels and the control of chloroplast division. Plant Physiology 69:1387–1391.

Braam, J. and Davis, R.W. (1990). Rain,-wind,- and touch induced expression of calmodulin and calmodulin-related genes in Arabibopsis. Cell 60:357-364.

Carter, J.V. and Wick, S.M. (1984). Irreversible microtubule depolymerization associated with freezing injury in Allium cepa root tip cells. Cryo-Lett. 5:373-382.

Coen, E.S., Romero, J.M., Doyle, S, J.M., Elliott, R., Murphy, G. and Carpenter, R. (1990). Floricaula: A homeotic gene required for flower development in Antirrhimum majus. Cell 63:1311-1322.

Coen, E.S. (1991). The role of homeotic genes in flower development and evolution. Ann. Rev. Plant Physiol. and Molec. Biology 42:241-279.

Colombo, M.F., Rall, D.C. and parsegian, V.A. (1992). Protein solvation in allosteric regulation: A water effect on hemoglobin Science 256:655-659.

Copeland, L. and Turner, J.F. (1980). The regulation of glycolysis and the pentose phoshate pathway. In P.K. Stumpf and Conn, E. (eds). The Biochemistry of Plants. 11:102-128.

Darwin, C. (Assisted by Francis Darwin) (1980). The Power of Movement in Plants John Murray London.

Davis, T.N., Urdea, M.S., Masiarz, F.R. and Thorner, J. (1986). Isolation of the yeast colmodulin gene: calmodulin is an essential protein. Cell 47:423-431.

Einsphr, K.J. and Thompson, Jr., G.A. (1990). Transmembrane signaling via phosphatidyl inositol 4,5-bisphosphate hydrolysis in plants Plant Physiology 93:361-366.

Feierabend, J. and Schrader-Reichhardt, U. (1976). Biochemical differentiation of plastids and other organelles in rye leaves with a high-temperature-induced deficiency of plastid ribosomes. Planta 129:133-145.

Furuya, M. (ed.) (1987). Phytochrome and Photoregulation in Plants. Academic Press, Tokyo.

Gatenby, A.A. and Ellis, R.J. (1990). Chaperone function in the assembly of ribulosebisphosphate carboxylase-oxygenase. Ann. Rev. Cell Biology 6:125-149.

Gilroy, S., Read, N.D., and Trewaras, A.J. (1990). Elevation of cytoplasmic calcium by caged calcium or caged inositol triphosphate initiates stomatal closure. Nature 346:769-771.

Gottlieb, L.D. (1982). Conservation and duplication of isozymes in plants. Science 216:373-380.

Guy, C.L. (1990). Cold acclimation and freezing stress tolerance: Role of protein metabolism. Ann. Rev. Plant Physiol. and Molec. Biol. 41:187–223.

Hopkins, W.G. and Elfman, B. (1984). Temperature-induced chloroplast ribosome deficiency in virescent maize. J. Heredity 75:207–211.

Hsiao, T.C. and Acevedo, E. (1974). Plant responses to water deficits, water use efficiency and drought response. Agric-Meterology 14:59–84.

Huner, N.P.A., Hopkins, W.G., Elfman, B. and Hayden, D.B. (1982). Influence of growth at cold-hardening temperature on protein structure and function. pp. 129–144 In P.H. Li and Sakai, S. (eds.) Plant Cold Hardiness and Freezing Stress: Mechanisms and Crop Implications. Vol. 2. Academic Press, N.Y.

Kay, S.A., Keith, B., Shinozaki, K., L-Chye, M. and Chua, N.-H. (1989). The rice phytochrome gene: structure, autoregulated expression, and binding of GT-1 to a conserved site in the 5' upstream region. Plant Cell 1:351–360.

Kivic, P.A. and Vesk, M. (1974). An electron microscope search for plastids in bleached Euglena gracilis and Astasia longa. Canad. Jour. Bot. 52:695–699.

Knight, M.R., Smith, S.M. and Trewavas, A.J. (1992). Wind induced plant motion immediately increases cytosolic calcium. Proc. Natl. Acad. Sci. 89:4967–4971.

Konjevic, R., Khurana, J.P. and Poff, K.L. (1992). Analysis of multiple photoreceptor pigments for phototropism in a mutant of Arabidopsis thaliana. Photochem. and Photobiol. 55: 789–792.

Kornblatt, J. and Hoa, G.H.B. (1990). A nontraditional role for water in the cytochrome c oxidase reaction. Biochemistry 29:9370–9376.

Mirza, J.I., Olsen, G.M., Iverson, T.-H. and Maher, E.P. (1984). The growth and gravitripic responses of wild type and auxin-resistant mutants of Arabidopsis thaliana. Plant Physiol. 60:516–522.

Mohapatra, S.S., Pook, R.J. and Dhindsa, R.S. (1987). Changes in protein patterns and translatable messenger RNA populations during cold acclimation of alfalfa. Plant Physiol. 84:1172–1176.

Olsen, R.A., clark, R.B., and Bennett, J.H. (1981). The enhancement of soil fertility by plant roots. Amer. Sci. 69:378-384.

Pringsheim, E.G. and Pringsheim, O. (1952). Experimental elimination of chromophores and eyespot in Euglena gracilis. New Phytol 51:65-76.

Rademacher, E. and Feierabend, J. (1976). Formation of chloroplast pigments and sterols in rye leaves deficient in plastid ribosomes. Planta 129:147-153.

Rand, R.P. (1992). Raising water to new heights. Science 256:618.

Rand, R.P. and Fuller, N.L. (1992). Water as an inhibiting ligand in yeast hexokinase activity. Biophysical J. 61(2,pt2);Abstr. #1986, p. A345.

Raven, J. (1985). pH regulation in plants. Sci. Prog. 69:495-509.

Sievers, A., Buchen, B., Volkman, D. and Hejnowicz Z. (1991). Role of the cytoskeleton in gravity perception. pp 169-182 In C.W. Lloyd (ed) The Cytoskeletal Basis of Plant Growth and Form. Academic Press, London.

Singer, H. (ed.) (1987). Blue Light Responses. CRC Press, Boca Raton, Florida.

Short, T.W., Porst, M. and Briggs, W.R. (1992). A photoreceptor system regulating in vivo and in vivo phosphorylation of a pea plasma membrane protein. Photochem. and Photobiol. 55:773-781.

Srere, P.A. (1987). Complexes of sequential metabolic enzymes. Ann. Rev. Biochem. 56:21-56.

Takahashi, H., Scott, T.K., and Suge, H. (1992). Stimulation of root elongation and curvature by calcium. Plant Physiol.98(1): 246-252.

Trewavas, A. and Gilroy, S. (1991). Signal transduction in plant cells. Trends in Genetics 7(11/12):356-361.

Warpeha, K.M.F., Kaufman, L.S. and Briggs, W.R. (1992). A flavoprotein may mediate the blue-light-activated binding of guanosine 5'-triphosphate to isolated plasma membranes in Pisum sativum L. Photochem. and Photobiol. 55: 595-603.

Chapter 12

PHYSIOLOGICAL REGULATION OF WATER TRANSPORT IN CONTROLLED ENVIRONMENTS

J. S. Boyer* and H. Nonami**

*College of Marine Studies, University of Delaware,
Lewes, DE19958, U.S.A.
** Department of Biomechanical Systems,
College of Agriculture, Ehime University,
Tarumi, Matsuyama 790, Japan

I. INTRODUCTION

Most plants possess the ability to carry out photosynthesis. Using the products of the process, all the organic constituents of the cells can be derived as long as CO_2 is available from the air, and H_2O and a few salts are available from the soil. The CO_2 and H_2O combine to form sugars that together with the elements from the soil are the building blocks for the other cell constituents.

The CO_2 enters the cells by dissolving in water at the cell surfaces, and thus the cells must expose wet surfaces to the air inside the leaf. This causes a large water loss (transpiration) from the plant because water tends to evaporate from the wet cell surfaces. The water vapor diffuses out of the leaf and into the surrounding air.

The H_2O entering the air causes special problems for controlled environments intended for plants. In a confined space, the evaporation causes humidity to rise. The more plants there are, the higher the humidity. High humidity affects the quantity and quality of the plant; leaves and fruits tend to become larger without a corresponding increase in structural material. Soils often become water-logged, and salt uptake may diminish. Root development may become inhibited as well. Because the humidity level depends on how many plants there are in the chamber, the plants vary in these characteristics according to plant density. Reproducible plant growth is thus difficult to obtain without humidity control.

The amount of water lost to the air depends mostly on the amount of radiation striking the leaves. The radiation causes leaf stomata to open. The greater the radiation input, the more the stomata open. The diffusive resistance to water loss is thus decreased by high radiation, and transpiration increases. In addition to these diffusive effects, the incoming radiation is partitioned between transpiration and sensible heat according to the humidity, wind speed, and temperature of the leaf and air. The lower the humidity, the greater is the fraction of the incoming radiation devoted to transpiration. Leaf temperatures decrease as transpiration becomes more rapid and may fall below air temperature.

The effect of these interacting factors is to cause humidity control to become more difficult as humidity is brought down to those approaching the external environment. To deal effectively with this problem, it is not enough to achieve humidity control in an empty chamber. Moreover, one cannot fill a chamber with plants and determine their water use at unspecified humidity. Design must be aimed at chambers fully laden with plants exposed to the required humidity.

II. PHYSIOLOGICAL CONTROL OF WATER MOVEMENT THROUGH THE PLANT

Because humidity can have significant effects on plant growth, it is important to consider the principles that are involved. The amount of water lost by transpiration is much larger than the amount required as a substrate for photosynthesis and cell growth, and this imposes larger requirements on water extraction from the soil than would otherwise occur. The situation can best be understood by considering the water lost from the plant to be part of what we shall term a "mass water budget" that includes all the water lost and gained by the plant (Boyer, 1985). The budget would have the form:

$$A + T + G = H \qquad (1)$$

where the water taken in is the absorption flux A, water loss is the transpiration flux T, the water used in all growth processes is the growth flux G, and the water stored in the leaf is the storage or hydration flux H. Each flux has units of $mol \cdot m^{-2} \cdot s^{-1}$ and is based on the same unit area, usually the projected leaf area, but the area is arbitrary and could be chosen anywhere in the plant as long as it is the same for each flux.

The flux A is toward the plant and accordingly is a positive flux whereas the flux T is a loss and thus a negative flux. Because the flux G involves water used in cell enlargement and metabolism, it also is negative. The hydration flux affects the water content of the tissue and depends on the balance of A, T, and G. Thus, H is positive when A exceeds T plus G, or negative when A is less than T plus G. In many cases H is zero because the other fluxes balance each other.

This budget includes G because cell enlargement and metabolism are essential plant functions even though the flux may be only 1/100 of T. The physical requirements are different for G and T, and conditions favoring T may inhibit G and vice versa. In particular, rapid T can dehydrate the shoot tissues and decrease G.

Water absorption for A depends to a large extent on the development of tensions (negative pressures) in the plant that extract water from the soil (Boyer, 1985). When T plus G exceed A, H becomes negative and the cells dehydrate. This dehydration is common early in the day and causes tensions to develop that are applied to water in the plant and ultimately the soil. As sufficient water is being absorbed to equal the requirements for T plus G, further dehydration is prevented and H becomes zero. In this condition, the shoot is dehydrated but stable. The process reverses as transpiration diminishes late in the day. Water enters faster than T plus G require. The H is positive, the shoot rehydrates, and tensions become less than T plus G are again balanced by lower A. This sequence of events indicates that absorption always lags transpiration, and it illustrates what has been called the "absorption lag" (Kramer, 1937; 1938).

The tensions that form as a result of dehydration are determined by the tissue water potential (i.e., the chemical potential of water expressed in units of pressure). The water potential (Ψw) varies with tissue water content and is determined mostly by solute (Ψs) and pressure (Ψp) components according to:

$$\Psi w = \Psi s + \Psi p \qquad (2)$$

Water potentials develop in cells because photosynthesis generates solute at concentrations usually much higher than in the solutions surrounding the cells. The high concentrations lower the water potential of the cell solution by the amount of osmotic potential which causes water to move into the cells (Fig. 1). As water enters, pressures develop because the increased volume is opposed by the rigidity of the cell walls. The increased pressure raises the water potential by the amount of turgor (Fig. 1).

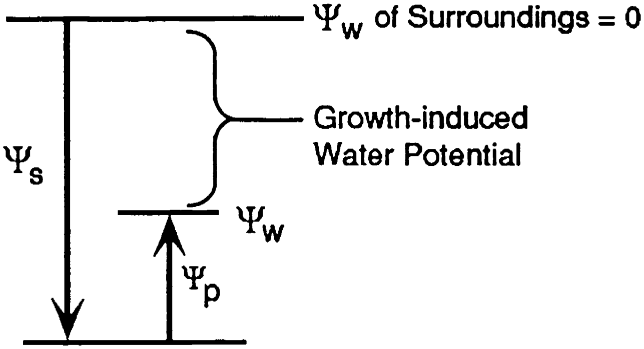


Figure 1. Potential diagram illustrating the osmotic potential (Ψs), pressure potential (Ψp), and water potential (Ψw) of a cell in a growing tissue. The tissue is in contact with water having a water potential of zero. In this case, the growth-induced water potential forms because pressure potential is not as high as it would be in the absence of cell enlargement. The pressure potential is suppressed by the yielding of the cell walls.

Because solute concentrations are generally low outside the cell (Scholander et al., 1965; Boyer 1967; Klepper and Kaufmann, 1966; Nonami and Boyer, 1987), the water potential is transmitted to the solution surrounding the cells mostly as a tension, which moves water into the cells and ultimately from the soil. When water is optimally available to plants, tensions tend to run in the 0.2 to 2.0 MPa range (Scholander et al., 1965; Boyer, 1967).

As tissues dehydrate (H is negative), their water potential becomes lower and greater force is exerted on the surrounding water. Tensions as high as 5 to 6 MPa (50 to 60 bars) have been measured in dehydrated tissues (Oechel et al., 1972). The increased tension increases A and counter balances T plus G. In effect, the balancing occurs because water potential changes in a corrective direction whenever an unbalance occurs and H is not zero.

The concept that the mass budget is balanced by the effects of water potential can be applied to the water lost by transpiration except that the loss is not uniform for all the cells, and the gradients are thus different from those required for growth (Boyer, 1974). The external leaf cells (epidermis) are coated with waxes and there are waxes on the surfaces of internal cells close to the stomatal pores (Boyer, 1985; Nonami et al., 1991). The waxy layers retard evaporation and it appears that transpiration occurs mostly from cells close to the veins. Thus, the cells close to the leaf surface are bypassed by the water vapor diffusing in the intercellular air spaces (Boyer, 1974; Boyer, 1985; Nonami and Schulze, 1989; Nonami et al., 1991). It also has been suggested that water moves readily to the epidermis and evaporates from the underside close to the stomatal pores (Meidner, 1976; Sheriff and Meidner, 1974). Whatever the fact, water appears to bypass many of the leaf cells (Rayan and Matsuda, 1988) in contrast to the situation for growth. Little is known about water potential gradients associated with transpiration except that they are substantial over large distances (e.g., from root to shoot) and they are detectable in individual leaves (Nonami and Schulze, 1989; Nonami et al., 1991).

III. WATER MOVEMENT THROUGH STOMATA

Because the pathways of water movement through stomata during transpiration are little understood, the water pathway through stomata has been studied by measuring cell water status of leaf cells (Nonami and Schulze, 1989; Nonami et al., 1991).

When water flows from the xylem towards a guard cell during transpiration, the hydraulic resistances within the tissue surrounding the stomatal pore will influence the water status of cells and may, thus, regulate stomatal responses to air humidity and transpiration. When air humidity becomes lower, transpiration usually increases, because the evaporative demand in air increases. However, if evaporative demand exceeds the supply of water in the plant, stomata of plants will be closed in order to prevent the excessive dehydration of the plants. In such an extreme case, it may be postulated that the proper adjustment of the water supply in response to environmental demands may cause all plants to exhibit a decrease in transpiration rate as the vapor pressure difference (VPD) between the inside of the leaf and the ambient air increases.

In order to demonstrate such an example, we have chosen *Tradescantia virginiana* as an experimental material. It has never been observed prior to this study that this plant exhibits a decrease of transpiration with an increasing VPD. *Tradescantia virginiana* has large cells, allowing cell water status to be determined easily by using the cell pressure probe for cell turgor determinations and the nanoliter osmometer for cell osmotic potential determinations (Nonami and Schulze 1989). Changes in the xylem water potential and light intensity could be set up, and it could be demonstrated that transpiration did indeed decrease as the VPD increased. The electronmicroscopy showed that the inner cuticle layer covers the walls of guard cells and subsidiary cells facing toward the stomatal cavity in *Tradescantia virginiana* (Nonami et at., 1991). Furthermore, by examining cell water status and water flow through cells surrounding the stomatal pore, mechanisms of stomatal closure occurring at an increasing VPD will be explained.

The experimental setups are as follows: A well-expanded leaf (25 - 35 cm long and 3rd or 4th from the youngest leaf) was washed with distilled water and wiped dry to remove dust and salt from the leaf surface, and all leaves of the plant except the washed leaf were excised, and cut surfaces were coated with petroleum jelly. The entire pot and basal part of the plant were covered with a plastic film and further covered with aluminium foil so that evaporation could only take place from the one leaf. A single, well-expanded, mature leaf was used in order to simplify water flow pathways and relations between transpiration and water status of cells, because the xylem is already differentiated completely and developed well in the mature leaf, and thus, hydraulically well-conductive. Further, because expanding young leaves consume water for growth (G) in addition to transpiration (T), and also because expanding cells tend to osmoregulate at low water potentials (Boyer 1985), water relations including young leaves may be more complex. Removing sources of extra water loss by both excising all leaves except the one leaf and covering the plant and pot thoroughly was important in order to minimize changes in water status in roots during the experiment. Also, the removal of extra tissues reduces the capacity of storage or hydration (H). By using such a setup, Equation 1 can be reduced to

$$A = -T \qquad (3)$$

Thus, the water status measured in plants is related entirely to functions of water absorption (A) and transpiration (T).

The plant was placed in a split chamber with a controlled environment which was modified from the system described by Shackel and Brinckmann (1985). The leaf gas exchange chamber was separated from the plant chamber and contained a 1 cm-thick brass bar along the bottom of the leaf chamber. The adaxial side of the leaf was pasted with petroleum jelly on the brass heat sink, and the leaf was tied with 50 μm-diameter constantan wire on the heat sink at three positions, ensuring close contact of the leaf with the heat sink. The temperatures of both the heat sink and the air of the chamber were controlled with the same temperature-control unit, and the temperature, which was recorded with a thermocouple located between the heat sink and leaf, did not deviate significantly from the air temperature (i.e., less than 0.1 °C) in all transpiring conditions. Because the whole plant was kept at a constant temperature, the enthalpic change (ΔH) in the plant cells during transpiration in this system could be considered negligible, and thus, water movement through the plant cells can be analyzed by simply measuring water potentials of cells.

The gas exchange method used was the open-air system and measured by the difference in H_2O and CO_2 concentrations between the inlet and outlet of the leaf chamber with a Binos gas analyzer. All gas flows were measured with calibrated rotameters. Air was taken from the outside of the building, filtered through charcoal, bubbled through heated-water and introduced into a temperature-humidity control unit. Air temperature in the leaf chamber was regulated to 28.0 °C ± 0.1 °C. The average CO_2 concentration introduced into the leaf chamber was 340 ppm. The rate of air flow through the leaf chamber was sufficient to give an air speed across the test leaf of approximately 0.5 $m \cdot s^{-1}$. The light source was a metal halide lamp set in a projector and filtered with 2 layers of infrared reflecting mirror. The photosynthetically active radiation was set to 2 different levels, i.e., 300 and 400 $\mu mol \cdot m^{-2} \cdot s^{-1}$. Air humidity of the inlet air of the leaf chamber was measured with a dewpoint mirror and controlled to 4 different levels of VPD between the air and the leaf, i.e., 0.4 kPa, 1.1 kPa, 2.3 kPa and 3.0 kPa VPD at 28 °C, by a feedback loop. Because the leaf temperature was regulated to closely match the air temperature, 0.4 kPa, 1.1 kPa, 2.3 kPa and 3.0 kPa of VPD corresponded to 90%, 70%, 40% and 20% RH of air humidity at 28°C. The VPD was changed stepwise at 4 levels from small to large (i.e., 0.4 kPa → 1.1 kPa → 2.3 kPa → 4.0

kPa), and then decreased stepwise in the reverse direction (i.e., 4.0 kPa → 2.3 kPa → 1.1 kPa → 0.4 kPa). At each step, at least 2 h of steady rates of transpiration and assimilation were maintained prior to measurements of the water status of the cells with the cell pressure probe and the nanoliter osmometer (Nonami et al., 1987; Nonami and Schulze, 1989). Simultaneously, stomatal width and length were measured directly with a microscope and the average opening of the stomatal pore area was calculated from 15 - 20 measurements by assuming that the stomatal aperture could be represented by an ellipse. After the gas exchange and water status measurements were taken, the plant was removed from the gas exchange system and transferred to a box with water-saturated air.

Xylem water potentials were estimated from water potential measurements obtained from the mature region of the roots. The root water potential was measured with the isopiestic psychrometer (Boyer and Knipling 1965) at the mature region located at the middle of the potted soil so that the potential was well equilibrated throughout the tissue and the surrounding soil. Thus, the root water potential can be said to represent the xylem water potential in the root. Because the *Tradescantia* plants have well-developed xylem vessels, resistance to water flow inside the xylem should have been very small. In fact, the xylem water potential changes between the basal part and the tip of the *Tradescantia* leaves were very small under various transpiring conditions (Shackel and Brinkmann 1985; Frensch and Schulze 1988). Therefore, the xylem water potential estimated from the water potential of the mature region of the roots may be safely assumed to apply throughout the entire plant, especially under conditions of this experiment, in which all leaves except one were removed and water loss, except from the leaf in the leaf chamber, was entirely prevented. A thermocouple chamber (Boyer 1966; Boyer and Knipling 1965) was coated with melted and resolidified petrolatum (Boyer 1967) and loaded with 6 root segments about 1.5 cm in length in a water-saturated humid chamber immediately after the segments were excised from the mature region of roots located at the center of the planting pot.

When the xylem water potential (Ψx) was -0.15 MPa, transpiration of plants at light intensity 400 $\mu mol \cdot m^{-2} \cdot s^{-1}$ increased almost linearly as the VPD between the leaf and air increased from 0.4 to 3.0 kPa (see open symbols of -0.15 in Fig. 2A for Ψx = -0.15 MPa). Reversing the process by increasing air humidity resulted in the same values of transpiration as when the VPD increased (see closed symbols of -0.15 in Fig. 2A for Ψx = -0.15

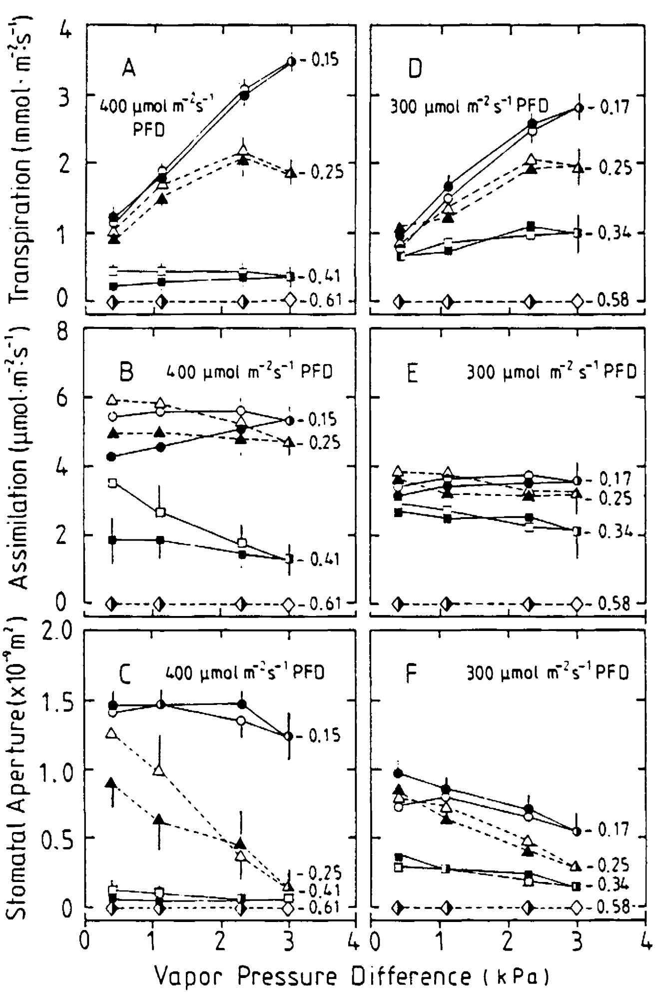


Figure 2. Transpiration (A,D), assimilation (B,E) and stomatal aperture (C,F) of *Tradescantia* plants having various xylem water potentials in relation to vapor pressure difference between the inside of the leaf and the ambient air under light intensity of 400 µmol·m^{-2}·s^{-1} (A–C) and 300 µmol·m^{-2}·s^{-1} (D–F). The xylem water potential was estimated from the water potential of the mature region of roots located in the middle of the pot, and written along the corresponding curve as a negative number expressed in units of MPa. Open symbols indicate stepwise increase of the VPD, and closed symbols indicate stepwise decrease of the VPD in the reverse direction. Vertical bars indicate 95% confidence intervals.

MPa). When the xylem water potential decreased to −0.25 MPa, transpiration increased as the VPD increased from 0.4 to 2.3 kPa, but a further increase of VPD caused a corresponding decrease of transpiration (see open symbols of −0.25 in Fig. 2A for $\Psi x = -0.25$ MPa). This decrease was related to a decrease in the size of the stomatal aperture (see open symbols of −0.25 in Fig. 2C for $\Psi x = -0.25$ MPa). In a reversed step-change of air humidity, a decrease of VPD from 3.0 to 2.3 kPa caused an increase in the size of the stomatal aperture (see closed symbols of −0.25 in Fig. 2C for $\Psi x = -0.25$ MPa), and an increase of the transpiration rate (see closed symbols of −0.25 in Fig. 2A for $\Psi x = -0.25$ MPa). At xylem water potentials between −0.15 and −0.25 MPa, assimilation rates were constant and not affected by changes of VPD (see $\Psi x =$ −0.15 and −0.25 MPa in Fig. 2B). When xylem water potential decreased further, the transpiration rate ceased to increase, and the assimilation rate decreased steadily as the VPD increased (see $\Psi x = -0.41$ MPa in Figs. 2A and 2B). Under light intensity of 300 $\mu mol \cdot m^{-2} \cdot s^{-1}$, both transpiration and assimilation exhibited tendencies similar to those under 400 $\mu mol \cdot m^{-2} \cdot s^{-1}$, but a concurrent decrease of transpiration with an increase of the VPD was not clearly observed (Figs. 2D and 2E).

Both the transpiration and assimilation rates increased rapidly with the increasing size of the stomatal aperture up to a value of about $1 \times 10^{-10} m^2$ (Figs. 3A and 3B). For apertures larger than $1 \times 10^{-10} m^2$ the increase was small (Figs. 3A and 3B). An increase of the VPD caused a proportional increase of transpiration rates at the same stomatal aperture (Fig. 3A), suggesting that a simple diffusion relation was working in the present experiment despite non-linear relations between stomatal aperture and water vapor flux through stomatal pores. Different light intensities gave rise to different assimilation rates at the same stomatal aperture (Fig. 3B), suggesting that diffusion of CO_2 into the leaf was limited by photosynthetic activity in chloroplasts. Because in the present study the ambient CO_2 concentrations were the same at both light intensities, the internal CO_2 concentration at 300 $\mu mol \cdot m^{-2} \cdot s^{-1}$ must have been higher than at 400 $\mu mol \cdot m^{-2} \cdot s^{-1}$ at the same stomatal aperture. Assimilation did not affect transpiration, because transpiration rates plotted together using two light intensity conditions did not deviate from each other, as shown in Fig. 3A in which data points contained data points from both 300 and 400 $\mu mol \cdot m^{-2} \cdot s^{-1}$ in Fig. 3B. Thus, transpiration in the present study was regulated by stomatal aperture size and VPD, but not by the internal CO_2 concentration.

In order to clarify the relations between the speeds of gas diffusion through stomatal pores and the size of stomatal openings, average gas flow rates through stomates were calculated by dividing the total transpiration rate by the total stomatal aperture area on the leaf surface (the average stomatal pore area x stomatal density x leaf area) at a given condition. Because both transpiration rate and

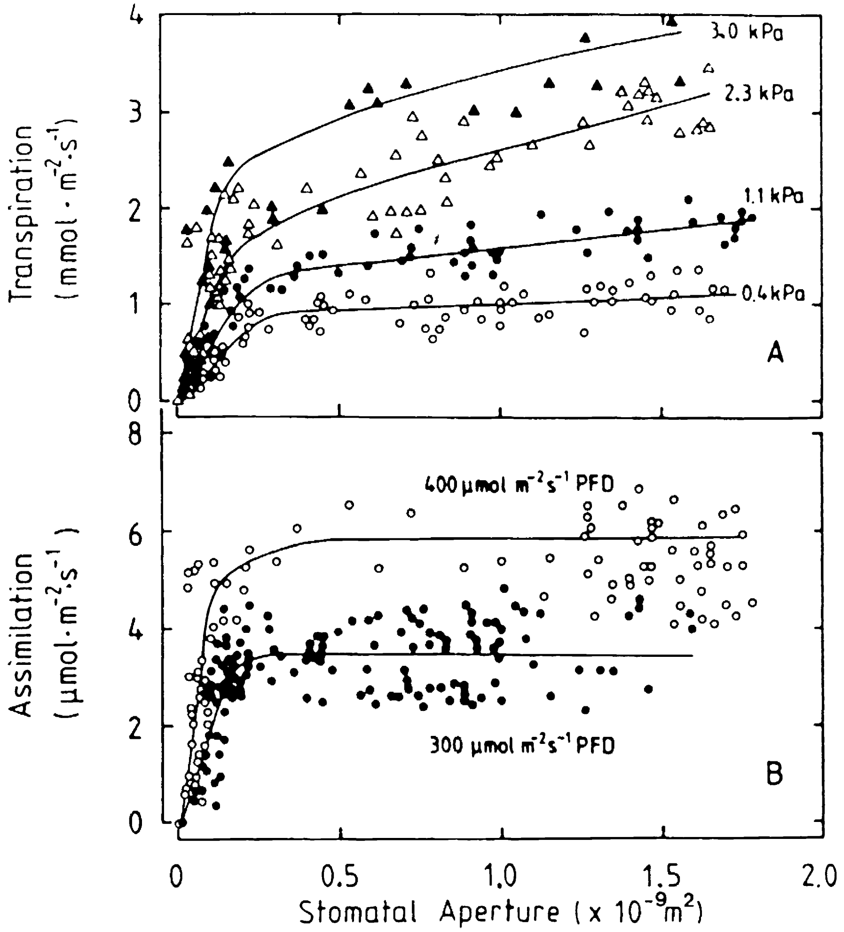


Figure 3. Transpiration rates (A) and assimilation rates (B) for unit leaf area in relation to stomatal aperture. Rates in A are plotted for 4 different VPD's (i.e., 3.0 kPa, 2.3 kPa, 1.1 kPa and 0.4 kPa), and rates in B are plotted for 2 different light intensities (i.e., 400 $\mu mol \cdot m^{-2} \cdot s^{-1}$ and 300 $\mu mol \cdot m^{-2} \cdot s^{-1}$). Each point represents one determination.

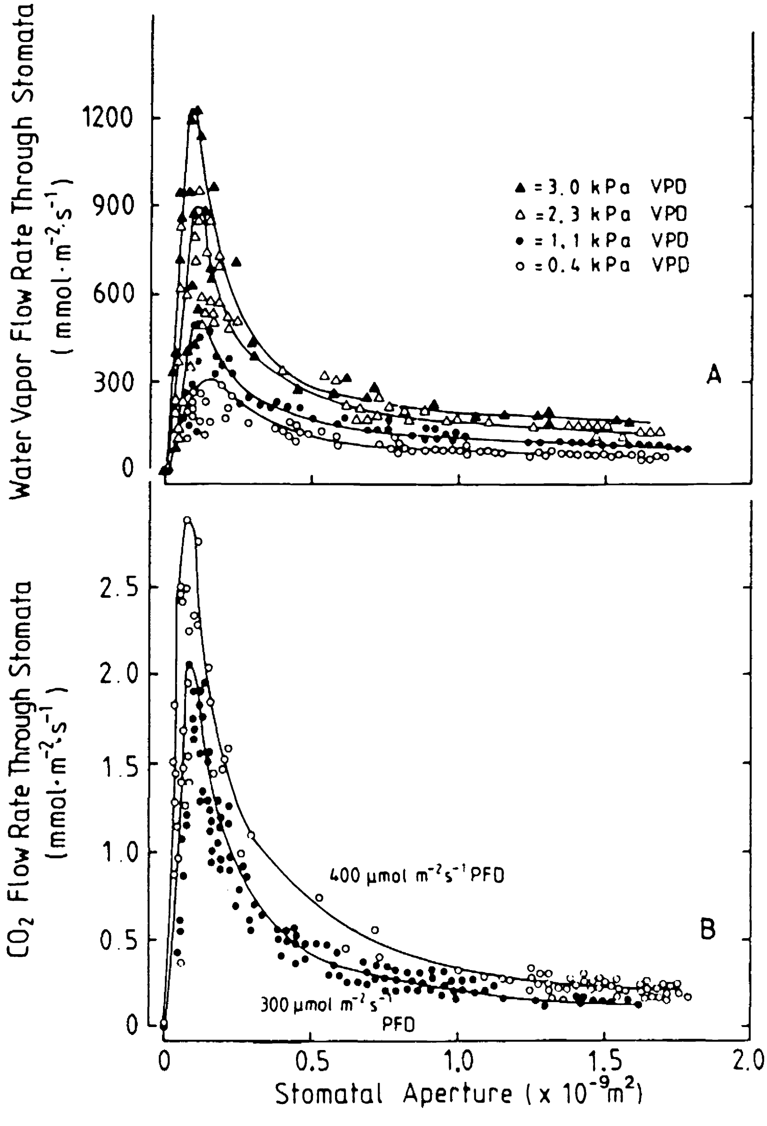


Figure 4. Water vapor flow rates (A) and CO_2 flow rates (B) through stomatal opening in relation to stomatal aperture. Rates in A are plotted for 4 different VPD's (i.e., 3.0 kPa, 2.3 kPa, 1.1 kPa and 0.4 kPa), and rates in B are plotted for 2 different light intensities (i.e., 400 µmol·m^{-2}·s^{-1} and 300 µmol·m^{-2}· s^{-1}). Each point represents one determination.

stomatal aperture size could be measured simultaneously with independent methods, and because the air flow in the gas exchange chamber was maintained at a constant velocity, calculation of the gas flow rates through stomates should have been an appropriate means to evaluate the gas diffusion rates through stomatal openings in the present study. Water vapor flow rates through stomates increased proportionally as VPD increased at the same stomatal aperture (Fig. 4A). As the stomatal aperture increased from zero to 1 x $10^{-10}m^2$, the water vapor flow rates increased linearly, and thereafter decreased almost hyperbolically over the wider stomatal apertures (Fig. 4A). Flow rates of CO_2 through stomates had speed profiles similar to those of water vapor in relation to stomatal aperture size (Fig. 4B), suggesting that stomatal regulation of gas exchange predominated when the size of the stomatal aperture was small, and that gas exchange was not predominantly regulated by movement of guard cells when the size of the stomatal aperture was larger.

In order to understand how stomatal movement and cell water status are interrelated with the inner cuticle covering the guard cells and the subsidiary cells (Boyer, 1985; Nonami et al., 1991) under transpiring conditions, we measured cell turgor, osmotic potential and water potential of epidermal, subsidiary and mesophyll cells when the plants had −0.25 MPa of Ψx under 400 $\mu mol \cdot m^{-2} \cdot s^{-1}$. Conditions for measurements were chosen so that changes in stomatal aperture size could occur over the widest possible ranges (Ψx = −0.25 MPa of Fig. 2C). These conditions were also chosen to see how the cell water status may change when transpiration rates are decreased with increasing VPD (see Ψx = −0.25 MPa in Fig. 2A).

Cell turgor was highest in mesophyll cells, and lowest in subsidiary cells, but subsidiary cell turgor was similar to epidermal cell turgor (Figs. 5A and 5D). Turgor of all cells decreased when air humidity became lower (Fig. 5A), and increased when air humidity became higher (Fig. 5D). Osmotic potentials of subsidiary cells decreased as VPD increased (Fig. 5B), and increased as VPD decreased (Fig. 5E). However, osmotic potentials of both epidermal and mesophyll cells were unaffected by either an increase or a decrease of the VPD (Figs. 5B and 5E).

The water potentials of mesophyll cells were the lowest at all VPD conditions (Figs. 5C and 5F). Because the inner cuticle layer covered the guard cells and subsidiary cells (Boyer, 1985; Nonami et al., 1991) and could prevent those cells from losing water, the

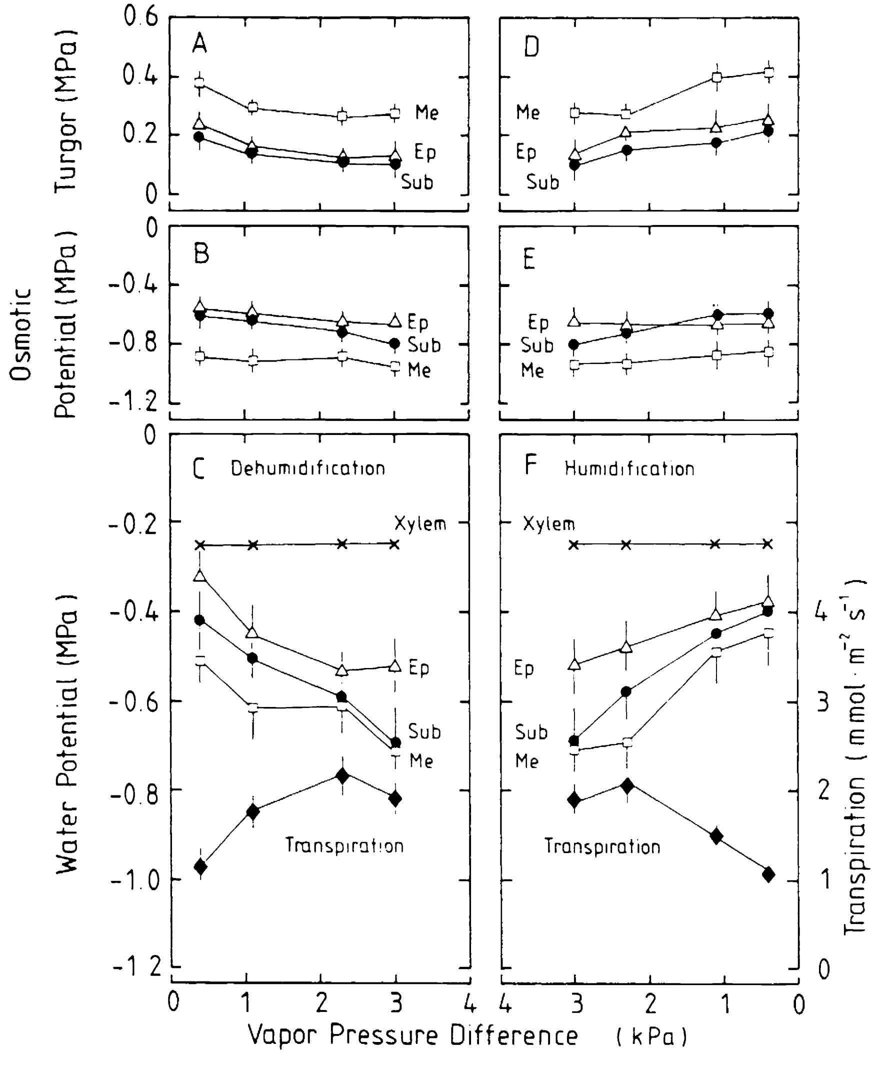


Figure 5. Cell turgor (A,D), cell osmotic potential (B,E) and cell water potential (C,F) of mesophyll cells (Me), epidermal cells (Ep) and subsidiary cells (Sub) in relation to VPD when plants had -0.25 MPa of xylem water potential. Transpiration rates are also plotted in C and F. Air humidity was first decreased (A–C), and afterward, increased in the reverse direction (D–F). Vertical bars represent 95% confidence intervals.

finding that the lowest water potential occurred in the mesophyll suggests that most of the water lost was evaporating from the mesophyll cells. When the VPD increased, subsidiary cell water potentials decreased further and became closer to mesophyll cell water potential, and subsequently, the transpiration rate decreased despite an increase of VPD from 2.4 to 3.0 kPa (Fig. 5C).

When the VPD decreased from 3.0 to 2.4 kPa, transpiration increased and water potential of the subsidiary cells increased more than those of epidermal cells and mesophyll cells (Fig. 5F). A further increase of VPD caused both a decrease of transpiration rates and an increase in water potential in all cells (Fig. 5F).

Proper adjustment of xylem water potential and light intensity caused the transpiration rate of *Tradescantia virginiana* to decrease as the VPD increased. Thus, one may safely assume that this phenomenon, a concurrent decrease of transpiration rate as VPD increases, can be observed in any species of plants which are subjected to specific environmental conditions. In the present study, the plants kept under well-watered, slightly dehydrated or lower light intensity conditions did not exhibit a decrease of transpiration rate with an increase of VPD, suggesting that only a specific combination of conditions involving the stomatal aperture, cell water status and water flow through cells surrounding the stomatal pore should cause the transpiration rate to decrease with an increase of VPD.

Because the occurrence of water loss through the cuticle layer at the leaf surface was negligible in *Tradescantia* plants, it may be safe to assume that water loss from locations covered with cuticle in the stomatal cavity is also negligible compared with locations without cuticle. Therefore, water evaporation is most likely to take place predominantly from surfaces of mesophyll cells during transpiration.

During the transpiration process, water flows from the xylem either through the bundle sheath extensions to the epidermis or directly to mesophyll cells. Since water flows through most permeable regions and the cuticle is not very permeable to water, water flows converge on the mesophyll region and water evaporates into stomatal cavities. Figure 6 illustrates water flow paths from the xylem to the ambient air. In the gas phase, movement of water was represented by streamlines (Batchelor 1967).

Through the stomatal opening, CO_2 gas also diffuses according to the concentration gradient of CO_2. When the internal CO_2 concentration was measured directly and compared with the internal CO_2 concentration estimated from CO_2 and water vapor fluxes, both values were practically identical (Sharkey et al. 1982). The assumption for the calculation of the internal CO_2 concentration was that the diffusion pathway of CO_2 gas can be considered as identical to that of water vapor but in the opposite direction (Moss and Rawlins 1963). Because CO_2 gas is assimilated in chloroplasts in mesophyll cells, CO_2 molecules are believed to move along the streamlines of water gas molecules evaporating from the mesophyll cells. Thus, streamlines illustrated in Fig. 6 can be also considered to represent those of CO_2 flux through the stomatal pore.

Previously, the greatest quantity of water loss occurring during transpiration was believed to originate from the peristomatal region (Maercker 1965; Maier-Maercker 1979, 1983; Meidner 1975, 1976; Cowan 1977; Rand 1977; Tyree and Yianoulis 1980). However, Boyer (1985) has suggested that the evidence for water loss from the peristomatal region should be reconsidered. In the experiments conducted to obtain evidence regarding the source of water loss (Maercker 1965; Maier-Maercker 1979, 1983; Meidner 1975, 1976; Cowan 1977; Rand 1977; Tyree and Yianoulis 1980), the natural flows of water taking place in intact plants had been disturbed, and furthermore, the existence of an inner layer of cuticle extending deeply into the stomatal pore has not been previously considered.

The actual sites of evaporation of water in the stomatal cavity should have a significant effect on the calculation of internal concentrations of CO_2 in the leaf because it is assumed that the diffusion pathway for CO_2 is identical to the diffusion pathway for water vapor (Moss and Rawlins 1963). The estimated internal CO_2 concentration should be the average concentration of CO_2 at the sites of evaporation. If most water evaporates from the subsidiary and guard cells as proposed previously (Maercker 1965; Maier-Maercker 1979, 1983; Meidner 1975, 1976; Cowan 1977; Rand 1977; Tyree and Yianoulis 1980), then direct measurement of the internal CO_2 concentration will deviate from the internal CO_2 concentration calculated by using rates of water vapor flow because there should be a drop in the CO_2 concentration between the sites of evaporation and the site of CO_2 assimilation. Only if most of the water lost through transpiration comes from mesophyll cells, the estimated internal CO_2 concentration should be equivalent to the directly measured internal CO_2 concentration (Sharkey et al. 1982). Therefore, by considering the existence of the inner cuticle layer extension (Boyer, 1985; Nonami et al., 1991), the cell water status of cells surrounding the stomatal cavity (Fig. 5) and confirmation of identical diffusion pathways of both CO_2 gas and water vapor (Sharkey et al. 1982), it is safe to conclude that water evaporates predominantly from mesophyll cells during transpiration.

Because the ambient CO_2 was kept constant in the present study, although air humidity was varied, relations between CO_2 assimilation rates and the stomatal aperture can be analyzed more easily than those between transpiration rates and the stomatal aperture. If the diffusion pathway of CO_2 gas is identical to that of water vapor through the stomatal opening, movement of CO_2 gas will simulate that of water vapor in the cavity, though diffusion takes place in the opposite direction. When the xylem water potential was kept higher than -0.25 MPa and VPD's were varied, assimilation rates were almost constant under 400 $\mu mol \cdot m^{-2} \cdot s^{-1}$ of light intensity (Fig. 2B), even when the stomatal aperture became small (Fig. 2C). This means that the same amount of CO_2 was still absorbed through the stomatal opening even though the aperture became smaller. Thus, because the number of stomates on the leaf was unchanged, when the opening of stomates became smaller, streamlines at the stomatal opening must have been compressed in order for the total amount of CO_2 absorbed to have remained constant, resulting in faster flow rates of CO_2 gas through the opening (Fig. 6). This was observed in Fig. 4B. Likewise, streamlines of water vapor flux must have been compressed, resulting in faster flow rates through the stomatal pore when the stomatal opening became smaller (Fig. 4A).

When a decrease of transpiration rate concurrent with an increase of VPD was observed, both CO_2 flow rate and water vapor flow rate through the stomata reached the maximum observed levels (Figs. 2A, 2B, 2C, 4A and 4B). At this critical stomatal aperture, the predominant limiting process for the gas diffusion through the stomatal opening shifted from a process caused by the concentration gradients of gases between the stomatal cavity and the ambient air to a process regulated by changes in the size of the stomatal aperture. During such a transition, the water status of cells surrounding the stomatal cavity was likely to have been altered. Water potential of subsidiary cells became lower than that of the surrounding epidermal cells and became equal to that of mesophyll cells (Fig. 5C). Because stomatal closure was occurring when the xylem water potential was unchanged, it is only possible to conclude that a shortage of water supply to the guard cells through the subsidiary cells occurred, causing turgor in the guard cells to be

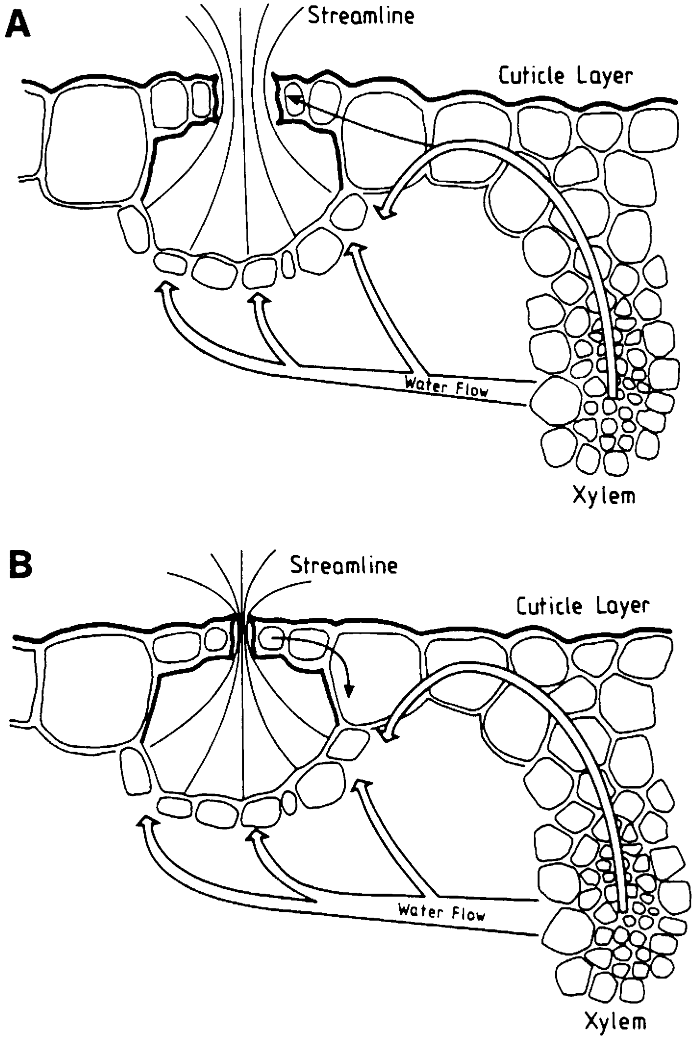


Figure 6. Schematic diagrams of water flow from the xylem to a stoma when the stoma opens fully (A) and partially (B). Arrows indicate water flux pathways. Thin lines from the bottom of the stomatal cavity to the outside of the leaf through the stomatal opening indicate streamlines for water vapor diffusion flux and CO_2 diffusion flux between the ambient air and the inside of the leaf. Note that the leaf internal cuticle extends to the edge where mesophyll cells are attached to epidermal cells (see Nonami et al. (1991) for details).

lower than in the subsidiary cells. Possibly, some water may have leaked out from the guard cells toward mesophyll cells via the subsidiary cells as illustrated in Fig. 6B.

When VPD was decreased by increasing air humidity, transpiration and water potential of the subsidiary cells increased (Fig. 5F). An increase in the stomatal aperture indicates that turgor of the guard cells became relatively larger than that of the subsidiary cells. Water must have flowed into the guard cells through the subsidiary cells, as illustrated in Fig. 6A, when higher air humidity diminished the excessive evaporative demand.

Thus, a concurrent decrease of transpiration rate with an increase of VPD may be interpreted to be a phenomenon related to stomatal closure caused by a shortage of water supply from the xylem to the guard cells when the evaporative demand in the air becomes too large despite relatively high values of xylem water potential. In such a situation, the transpiration rate is primarily regulated by the size of the stomatal opening rather than the VPD between the inside of the leaf and the ambient air. As a result, plants seem to conserve excessive water loss.

Bulk leaf water potentials have usually been used to describe the water status of leaves. Because bulk leaf water potentials are equivalent to the volume-averaged value of cell water potentials, local variations of water potentials are ignored. While epidermal and mesophyll cell water potentials changed from -0.5 MPa to -0.8 MPa, the xylem water potential ranged from -0.15 MPa to -0.61 MPa (Nonami and Schulze, 1989; Nonami et al., 1991). Because epidermal and mesophyll cells comprise most cells in the leaf, if one assumes that the volume-averaged water potentials of the two tissues represent the bulk leaf water potential, it may be concluded that stomatal movement may be independent of the bulk leaf water potential under most ranges of stomatal aperture. Such a conclusion may be erroneous, however, because changes of the xylem water potential can be ignored in the process of determination of the bulk leaf water potential. Because the xylem is the only water source for all leaf cells, changes in the xylem water potential should affect water flow through cells surrounding the stomatal cavity.

IV. GROWTH-INDUCED WATER POTENTIALS

The concept that the mass water budget is balanced by the effects of water potential has major implications for plant growth in controlled environments. Leaves and indeed all plant organs enlarge in a coordinated fashion by absorbing water and thus, water for the growth process must enter all the cells (Molz and Boyer, 1978). As a result, cells closest to the water supply must transmit water at rates many times their own needs. Despite the fact that they may have a relatively high water conductivity, the additional flow necessary for the other cells causes substantial water potential gradients to form. For growth, water potentials in the outermost cells can be 0.2 to 0.4 MPa lower than in the inner cells (Molz and Boyer, 1978).

The water potential associated with growth can be studied if plants are grown in the dark in saturating humidity where T is zero. Under steady conditions, H also is zero and Equation 1 becomes:

$$A = -G \qquad (4)$$

The water potential of the tissue is then devoted entirely to absorbing water for the growth process.

Figure 7 shows water potential along the stem of a soybean seedling grown in this way (Cavalieri and Boyer, 1982; Boyer 1988). The water potentials are averages and are about 0.15 MPa lower in the rapidly growing (rapidly elongating) tissue at the top of the stem compared with the slowly growing (nonelongating) tissue at the base. In contrast, the nonelongating tissue is in near-equilibrium with the water potential of the root medium because there is no water uptake by the cells (A, T, G, and H are zero in this tissue). Thus, the water potential in the vascular supply also equilibrates with the water potentials of these cells.

On the other hand, because the water potential of the elongating tissue is lower than in the nonelongating tissue, the water potential must be lower than in the vascular supply. The amount by which it is lower represents the force applied to the water in the vascular system by the enlarging cells. The result is that water moves out of the vascular system at a rate sufficient to feed the enlargement process.

Although these potentials are averages, they indicate that a three dimensional gradient exists around the vascular system (Fig. 7, middle graphs). This gradient has a higher potential close to the vascular tissue and decreases radially as water moves outward from the vascular system (Molz and Boyer, 1978). It also decreases as water moves inward from the vascular tissue. Because of its three-dimensionality, it is most properly considered a potential field

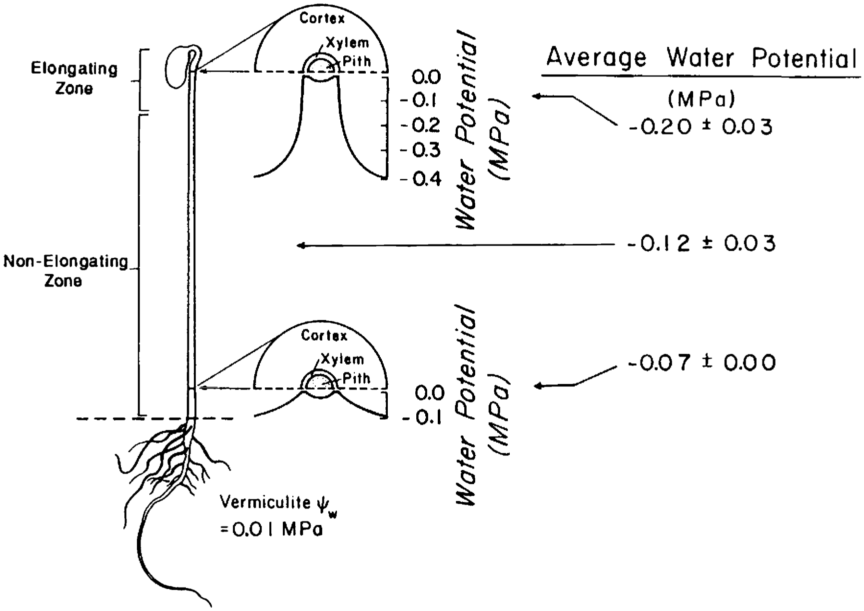


Figure 7. Water potential (Ψw) at various positions along the stem (hypocotyl) of soybean seedlings growing rapidly in saturating humidity in the dark. The average water potential is shown (right-hand table) together with the profile of water potential along the diameter of the stem (center graphs). Note that the profile is steep next to the vascular supply (xylem) and flat at the outer surface of the stem. The profile, when rotated around the longitudinal axis of the stem, gives the 3-dimensional potential field.

arrayed around the vascular system (Nonami and Boyer, 1987; Nonami and Boyer, 1989).

The potential field is responsible for moving water into each cell in a coordinated fashion so that the entire stem enlarges. Because the cells closest to the vascular supply must transmit water for the outlying cells, the potential field is steepest there. The cells farthest from the system require only water for their own needs and the field becomes fairly flat at this position (Fig. 7).

V. ORIGIN OF GROWTH-INDUCED WATER POTENTIALS

The potential field associated with growth is a fundamental property of all growing plant tissues (Boyer, 1985; Nonami and Boyer, 1987). It was suggested that the field forms from the growth activity of the cells because the enlarging cell walls would yield to internal pressures, preventing them from becoming as high as they otherwise would if the walls were rigid (Boyer, 1968). The effect would be to create a water potential lower than in the vascular system (Fig. 1). A different view is that the cell protoplasts could be surrounded by a concentrated solution in the cell walls sufficient to cause water potentials to be low in the cells (Cosgrove and Cleland, 1983). Nonami and Boyer (1987) found that high concentrations were not present, however. They observed instead a significant tension on the water surrounding the cells. Thus, the original concept based on wall yielding is the most likely, and the tension in the wall solution probably arises by transmission of the low potential of the cell protoplasts to the surrounding solution in the usual way.

Water entering the enlarging cells probably flows simultaneously through the cells and around them in the cell walls (apoplast) according to the relative resistances of each path (Boyer, 1985; Molz and Ferrier, 1982; Steudle, 1989; Steudle and Frensch, 1989; Steudle and Jeschke, 1983). The magnitude of the tensions supporting this flow indicates that enlarging tissues have a significant frictional resistance. This probably occurs because the vascular supply is not well developed in enlarging tissues. Also, there are undifferentiated tissues throughout the enlarging region (Steudle and Boyer, 1985).

The presence of the field has two consequences. First, if its shape is modified it will alter the flow of water to the outer tissues for growth. For example, if the water potential decreases in the vascular supply, the initial effect is to invert the field immediately adjacent to the vascular system (Fig. 8). This inversion prevents water from moving out of the vascular system and blocks water transport to all the outlying tissue. However, the outlying tissues are not initially changed in their water potentials and most of the potential field is still intact (Fig. 8).

Such a block rapidly inhibits the enlargement process (Nonami and Boyer, 1989; Nonami and Boyer, 1990a), probably within seconds, because there is a change in water potential in only a few of the cells. Rapid inhibitions have been observed under

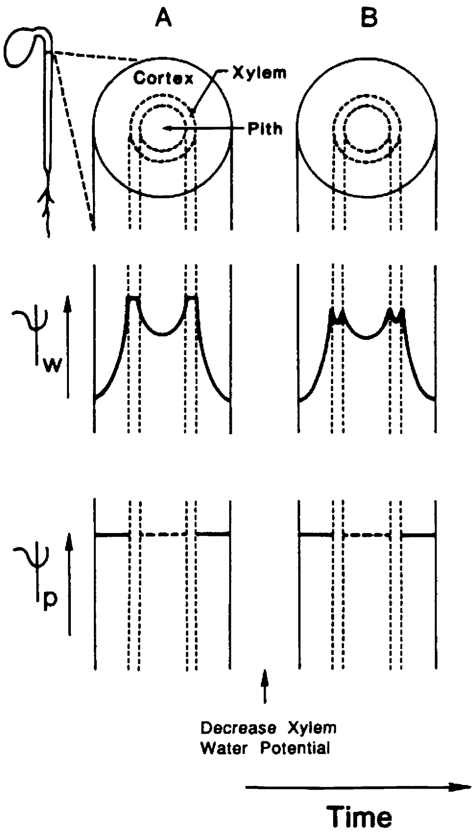


Figure 8. Changes in the growth-induced water potential (Ψw) profile after a decrease in the water potential of the vascular supply. A) Tissue grows rapidly before decrease and a water potential field is present that moves water out of the xylem. The pressure potential (Ψp) is uniform in all the cells. B) Growth is inhibited immediately after xylem water potential begins to decrease. C) Water potential field collapses after 1-2 h, and turgor decreases in cells close to vascular system. Note that water movement and cell enlargement are inhibited at B and C because profile is inverted next to the xylem. For growth to resume, the profile must move downward to re-establish a favorable water potential field.

conditions that would change the field (Acevedo et al., 1971; Nonami and Boyer, 1990a). In time, the potential field of the outlying cells adjusts, a new field is established, and growth resumes gradually (Nonami and Boyer, 1990a). The gradual resumption is in contrast to the rapid inhibition initiated by the vascular system. The difference is probably attributable to the number of cells involved.

Thus, growth competes with transpiration for the water in the vascular system, and growth is sensitive to the rate of transpiration (Westgate and Boyer, 1984; Fiscus et al., 1983). In effect, transpiration lowers the potential in the vascular supply and the growth of the cells needs to compete for water at a correspondingly lower potential.

The growing cells respond to vascular conditions by adjusting their osmotic potential (Westgate and Boyer, 1984) which maintains the turgor and the growth-induced water potential (Fig. 1). The osmotic potential determines the maximum force that can be applied for the growth process. The osmotic potential is lowered by solute entering the cells faster than it is utilized. The solute is derived mostly from photosynthesis (or from reserves previously accumulated by photosynthesis) and consists of sugars and amino acids (Meyer and Boyer, 1981; Morgan, 1984). It is used in biosynthesis for growth and will accumulate if it is unused in the growth process. There is evidence that growth slows as the solution in the vascular system decreases in potential, and solute accumulation results (meyer and Boyer, 1981). It has been shown (Meyer and Boyer, 1981; Michelena and Boyer, 1982; Sharp and Davies, 1979; Westgate and Boyer, 1985) that the accumulation is particularly large in the growing regions. Because the accumulation increases the quantity of sugars and amino acids, the quality of the tissue is changed.

The field provides a way to explain how growth responds to the effects of transpiration and variations in the water supply around the roots, and why the quality of the tissues also changes. It is important to note that, because the potential of the growing cells is not in equilibrium with the surroundings, a change in potential in the vascular supply will not cause the same change in the growing cells. Initially, the water potential of most of the cells is unchanged and later water potential adjusts to a new water potential. Thus, growth experiments require the water potential to be measured. If the growing tissue is excised, the water potential also will change. Excision removes the external water supply but solute transport cannot occur, the osmotic potential cannot adjust, and enlargement continues until turgor decreases to a threshold value at which enlargement is no longer possible (Cosgrove et al., 1984; Boyer et al., 1985). The decreased turgor results from cell wall relaxation that occurs as the walls yield but no water enters. The effects are quite small (water potential decreases about 0.5 to 1.0 MPa) and are delayed by attached mature tissue which can act as an internal water supply (Matyssek et al., 1988). For some growth experiments, however, these changes will need to be taken into account.

VI. GROWTH REGULATION UNDER WATER STRESS

Although water entry into cells is the major cause of cell elongation, the properties of the cell wall and membrane influence the rate of water flow into cells. The process of cell elongation depends on the extension of the cell walls together with the uptake of water and solute. Turgor (Ψp) extends the walls when a threshold turgor, the yield threshold (Y), has been exceeded (Cleland, 1959; Probine and Preston, 1962; Green et al., 1971). This extension depends on the extensibility (m) of the walls that is determined by metabolic activity (Probine and Preston, 1962; Green et al., 1971; Green and Cummins, 1974). The parameters affecting cell enlargement can be expressed as averages for the tissue in which case the rate of wall extension for growth can be related to m ($s^{-1} \cdot MPa^{-1}$) and ($\Psi p - Y$) (MPa) by the equation (Green et al., 1971):

$$G = m(\Psi p - Y) \qquad (5)$$

where G is the relative growth rate (1/Vte)(dVt/dt). Vte is the total volume of the elongating tissue. During growth in a localized growing region, Vte is constant. The growth rate, (dVt/dt), is then the rate at which the volume of new mature tissue is produced. Equation 5 implies that demand for water is determined both by the amount of turgor above the yield threshold and by the extensibility of the wall.

The simultaneous uptake of water is necessary because turgor would otherwise decrease as the walls extend, inhibiting cell enlargement. However, water uptake is not sufficient to completely maintain maximum turgor because there are resistances to water movement into the enlarging cells (Nonami and Boyer, 1987). This creates a water potential (Ψw) that is lower than in the water source (Ψo), usually the xylem. The ($\Psi o - \Psi w$) is the growth-induced water potential. The rate of water uptake necessary to support cell enlargement can be related to hydraulic conductance (L) ($s^{-1} \cdot MPa^{-1}$) and ($\Psi o - \Psi w$) (MPa) by the equation (Boyer and Wu, 1978; Boyer et al., 1985):

$$G = L(\Psi o - \Psi w) \qquad (6)$$

where G is the relative growth rate (1/Vwe)(dVw/dt). Vwe is the volume of water in the elongating tissue. During growth in localized growing regions, Vwe is essentially constant. The growth rate, dVw/dt, is then the rate at which water enters the elongating region and is converted to an equivalent volume of water in the mature tissue. Equation 6 implies that supply for water will be determined both by the potential difference and the conductance.

It should be noted that, because tissue volume is proportional to the water volume contained by the cells, (1/Vwe)(dVw/dt) = (1/Vte)(dVt/dt), and thus, G is the same for Eqs. 5 and 6.

Combining Eqs. 5 and 6 by using relationship $\Psi w - \Psi s = \Psi p$ yields:

$$G = (mL/(m+L))(\Psi o - \Psi s - Y) \qquad (7)$$

which is the rate equation governing tissue enlargement. Equation 7 shows both the effects of wall extensibility and tissue conductance which determine water demand and water supply for the enlarging cells.

All parameters in Eqs. 5, 6 and 7 could be measured simultaneously in intact seedlings by using the guillotine thermocouple psychrometry (Boyer et al., 1985; Nonami and Boyer, 1990a). Nonami and Boyer (1990a) studied the kinetics of physiological parameters regulating plant growth under water stress by using stems of soybean seedlings which were exposed to water stress. Figure 9 shows the sequential changes in most of the growth parameters which are expressed in relative values compared with non-stressed seedlings. The results show that ($\Psi o - \Psi w$) was the first to respond to the water stress. The simultaneous decrease in G and ($\Psi o - \Psi w$) indicates that the change in growth- induced water potential was likely to have caused the decrease in G. The recovery of ($\Psi o - \Psi w$) and G after some time is further evidence that ($\Psi o - \Psi w$) played a regulatory role. However, despite the complete recovery of (Ψo - Ψw), the recovery of G was incomplete. This indicates that another factor was responsible for limiting for G.

After the coincident decrease in G and (Ψo - Ψw), m and L decreased (Fig. 9). The recovery of (Ψo - Ψw) was followed by m, (Ψp - Y), and L. G recovered as the last of the growth parameters (m and L) recovered. Therefore, the factors limiting G must have been m and L.

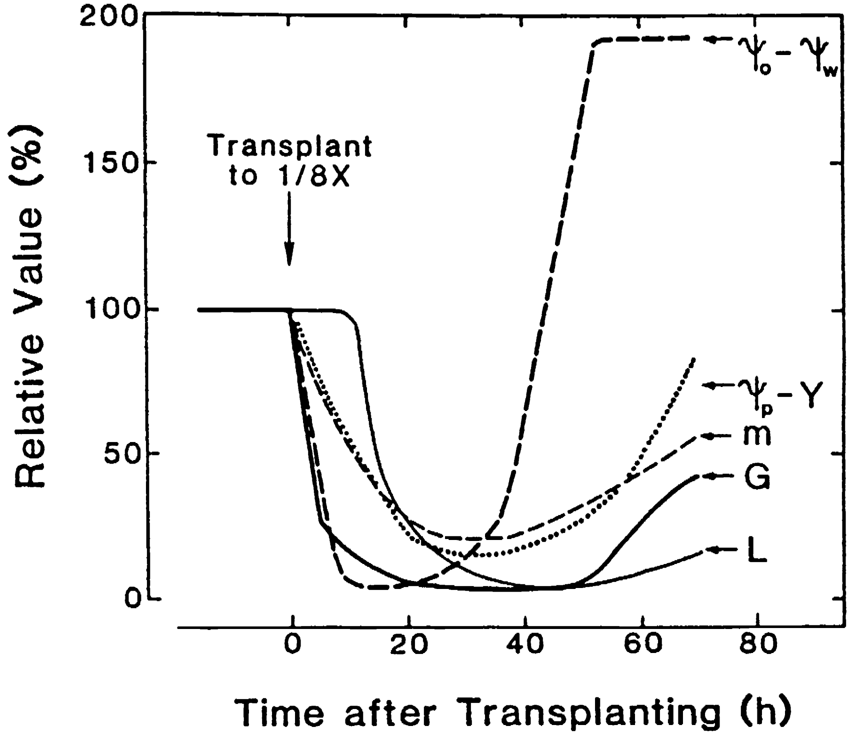


Figure 9. Relative values of the stem elongation rate (G), growth-induced water potential (Ψo - Ψw), growth-active turgor (Ψp - Y), wall extensibility (m), and tissue conductance for water (L) at various times in stems of intact soybean seedlings transplanted to water deficient vermiculite (i.e., vermiculite having 1/8x of water content of the control), compared with the control (100 %, before transplanting).

This order indicates that the inhibition of cell elongation was caused first by a physical factor ($\Psi o - \Psi w$) and secondarily by metabolically controlled factors (m and L). Although the physical factor recovered, G continued to be suppressed until the metabolically controlled factors recovered. Therefore, both during inhibition and recovery, the sequence of changes was consistent with the transmission of a physical signal to the cells that resulted in a metabolic response. The transmission required about 15 h during the inhibition and 10 to 20 h during recovery.

In order to check the observations which indicate that the inhibition of stem growth at low water potentials is accompanied by decreases in cell wall extensibility and tissue hydraulic conductance to water, Nonami and Boyer (1990b) developed a theory to relate rheological properties of cell walls measured with an extensiometer and thermodynamical properties of cell walls measured with psychrometry. They found that the plastic properties of the cell walls determined cell wall extensibility (Nonami and Boyer, 1990b).

Both cell wall extensibility and hydraulic conductance are under metabolic control, and their low levels under water stress indicate that substantial metabolic change occur in the enlarging cells (Nonami and Boyer, 1990a). Wall extensibility is probably determined by the action of wall enzymes on the polymerization of wall constituents and the cross-linking of structural proteins with wall constituents (Cassab and Varner, 1988; Fry, 1986; Theologis, 1986). In seedlings similar to those used here, polyribosome and mRNA levels decreased a few hours after water stress (Mason et al., 1988). Increased amounts of proteins were extractable from the walls and a 28 kD protein present in the cytoplasm accumulated in the walls (Bozarth et al., 1987). Translation products of the mRNAs were enhanced in the elongation zone of water-deficient soybean seedlings, and some of them were preferentially associated with membrane-bound polysomes (Mason et al., 1988). These findings suggest that metabolic alterations involving the cell walls and membranes take place after the collapse of the growth-induced water potential associated with the growth inhibition at low water potentials. Such metabolic alterations in cell walls and membranes may be mediated by ABA (Bensen et al., 1988).

The metabolic changes represented by m and L are clearly at the cellular level because m is a property of the cell walls and the changes in L are similar to the changes in hydraulic conductivity observed in individual cells with the miniature pressure probe (Nonami and Boyer, 1990b), which must have involved cell membranes. It is noteworthy in this regard that proton secretion to the cell walls is thought to affect the extensibility of the walls (Cleland, 1971, 1975, 1983; Cleland and Rayle, 1978) and is curtailed by exposure of the plants to water stress (Van Volkenburgh and Boyer, 1985). Thus, it is possible that decreased proton secretion is the manifestation of a metabolic signal that is being transmitted to the walls.

After water stress is applied to plants, growth recovery occurs when the xylem water potential is equilibrated with the medium. This is caused by osmotic adjustment in cells (Nonami and Boyer, 1989, 1990a). Prior to growth recovery, recovery of the growth-induced water potential is observed. The recovery of the growth-induced water potential induces metabolic changes in cells, resulting in increases in wall extensibility and hydraulic conductance (Nonami and Boyer, 1990a, 1990b). Afterward, growth is resumed. Thus, the growth-induced water potential can be considered to be a primary physiological factor regulating growth under water stress.

Because the expanding cells must absorb water from the water source in order to increase their volume, the growing tissue must form a growth-induced water potential field. The concept of the growth-induced water potential field must be applicable to any growing tissue. Boyer (1988) reduced the rate of cell enlargement by lowering the temperature, and measured the growth rates and the water potential gradient with the pressure chamber and the isopiestic psychrometer. He observed that the water potential gradient decreased as the growth decreased, indicating that the reduction of cell expansion was caused by the reduction of the growth-induced water potential. A similar result was obtained when the growth was reduced by exposure to high pressure (Boyer, 1988). Therefore, a collapse of the growth-induced water potential accounts for the inhibition of cell enlargement caused by environmental stresses such as water deficiency, low temperature and high pressure. It is worth noting that the growth-induced water potential is a physicochemical quantity and changes in the growth-induced water potential can be considered as the first signal received by plants when environmental stresses are imposed on them. After growth inhibition occurs due to the collapse of the growth-induced water potential, subsequent metabolic alteration takes place, and properties of cell walls and membranes will be changed (Nonami and Boyer, 1990a, 1990b).

VII. CONCLUSIONS

The process of transpiration creates demand for water on the plant that is far larger than its needs for just growth and metabolism. Most of the water is lost by transpiration and can cause high and variable humidity in controlled environments. Variations in humidity within the controlled environments result in differential development and quality of various plant parts. As a consequence, humidity control with a large capacity for dehumidification is required for reproducible plant growth. The differential development at various humidity levels can be understood with a mass water budget showing that transpiration and growth compete for the same water supply. Gradients in water potential affect the competition. Growth requires a growth-induced water potential field around the vascular supply to provide water at the required rate. Water is transported along the field to all the cells, moving from cell to cell and encountering significant frictional resistances. Water for transpiration also probably requires a potential field between the water source and the evaporating site in leaves.

The competition for vascular water forces growth to operate with a lower potential field when transpiration is faster. The lower field is supported by a lower osmotic potential because of accumulated solute that also acts to maintain pressures inside the growing cells. The quality of the tissue is affected by the increased solute. Thus, depending on the objectives, plant development can be altered in controlled environments by adjusting humidity to either favor rapid growth, which maximizes the size of the product or to slowing growth, which increases the content of sugars and amino acids and thus alters the quality of the product.

REFERENCES

Acevedo, E., Hsiao, T.C., and Henderson, D.W. (1971) Immediate and subsequent growth responses of maize leaves to changes in water status. Plant Physiol., 48, 631-636.

Batchelor, G.K. (1967) An Introduction to Fluid Dynamics. Cambridge University Press. pp615

Bensen, R.J., Boyer, J.S., and Mullet, J.E. (1988) Water deficit-induced changes in abscisic acid, growth, polysomes, and translatable RNA in soybean hypocotyls. Plant Physiol.

88:289–294.

Boyer, J.S. (1966) Isopiestic technique: Measurement of accurate leaf water potentials. Science 154, 1459–1460

Boyer, J.S. (1967) Leaf water potentials measured with a pressure chamber. Plant Physiol., 42, 133–137.

Boyer, J.S. (1968) Relationship of water potential to growth of leaves. Plant Physiol., 43, 1056–1062.

Boyer, J.S. (1974) Water transport in plants: mechanism of apparent changes in resistance during absorption. Planta, 117, 187–207.

Boyer, J.S. (1985) Water transport in plants. Annu. Rev. Plant Physiol., 36, 473–516.

Boyer, J.S. (1988) Cell enlargement and growth–induced water potentials. Physiol. Plantarum, 73, 311–316.

Boyer, J.S., Cavalieri, A.J., and Schulze, E.–D. (1985) Control of cell enlargement: effects of excision, wall relaxation, and growth–induced water potentials. Planta, 163, 527–543.

Boyer, J.S., and Knipling, E.B. (1965) Isopiestic technique for measuring leaf water potentials with a thermocouple psychrometer. Proc. Natl. Acad. Sci. U.S.A. 54, 1044–1051

Boyer, J.S., and Wu, G. (1978) Auxin increases the hydraulic conductivity of auxin–sensitive hypocotyl tissue. Planta, 139, 227–237.

Bozarth, C.S., Mullet, J.E., and Boyer, J.S. (1987) Cell wall proteins at low water potentials. Plant Physiol., 85, 261–267.

Cassab, G.I., and Varner, J.E. (1988) Cell wall proteins. Annu. Rev. Plant Physiol. Plant Mol. Biol., 39, 321–353.

Cavalieri, A.J., and Boyer, J.S. (1982) Water potentials induced by growth in soybean hypocotyls. Plant Physiol., 69, 492–496.

Cleland, R.E. (1959) Effect of osmotic concentration on auxin action and on irreversible and reversible expansion of the Avena coleoptile. Physiol. Plant., 12, 809–825.

Cleland, R. (1971) Cell wall extension. Annu. Rev. Plant Physiol., 22, 197–222.

Cleland, R.E. (1975) Auxin–induced hydrogen ion excretion: Correlation with growth, and control by external pH and water stress. Planta, 127, 233–242

Cleland, R.E. (1983) The capacity for acid–induced wall loosening as a factor in the control of Avena coleoptile cell elongation. J. Exp. Bot., 34, 676–680.

Cleland, R.E., and Rayle, D.L. (1978) Auxin, H+ excretion and cell elongation. Bot. Mag. (Spec. issue), 1, 125–139

Cosgrove, D.J. and Cleland, R.E. (1983) Solutes in the free space of growing stem tissues. Plant Physiol., 72, 326–331.

Cosgrove, D.J., Van Volkenburgh, E., and Cleland, R.E. (1984) Stress relaxation of cell walls and the yield threshold for growth: micropressure probe and psychrometer techniques. Planta, 162, 46–54.

Cowan, I.R. (1977) Stomatal behaviour and environment. Adv. Bot. Res. 4, 117–228

Fiscus, E.L., Klute, A., and Kaufmann, M.R. (1983) An interpretation of some whole plant water transport phenomena. Plant Physiol., 71, 810–817.

Frensch, J., and Schulze, E.-D. (1988) The effect of humidity and light on cellular water relations and diffusion conductance of leaves of *Tradescantia virginiana* L. Planta 173, 554–562

Fry, S.C. (1986) Cross-linking of matrix polymers in the growing cell walls of angiosperms. Annu. Rev. Plant Physiol., 37, 165–186.

Green, P.B., and Cummins, W.R. (1974) Growth rate and turgor pressure: Auxin effect studied with an automated apparatus for single coleoptiles. Plant Physiol., 54, 863–869

Green , P.B., Erickson, R.O., and Buggy, J. (1971) Metabolic and physical control of cell elongation rate. In vivo studies in Nitella. Plant Physiol., 47, 423–430.

Klepper, B., and Kaufmann, M.R. (1966) Removal of salt from xylem sap by leaves and stems of guttating plants. Plant Physiol., 41, 1743–1747.

Kramer, P.J. (1937) The relation between rate of transpiration and rate of absorption of water in plants. Amer. J. Bot., 24, 10–15.

Kramer, P.J. (1938) Root resistance as a cause of the absorption lag. Amer. J. Bot., 25, 110–113.

Maercker, U. (1965) Zur Kenntnis der Transpiration der Schliesszellen. Protoplasma 60, 61–78

Maier-Maercker, U. (1979) "Peristomatal transpiration" and stomatal movement: A controversial view. I. Additional proof of peristomatal transpiration by hygrophotography and a comprehensive discussion in the light of recent results. Z. Pflanzenphysiol. 91, 25–43

Maier-Maercker, U. (1983) The role of peristomatal transpiration in the mechanism of stomatal movement. Plant Cell Environ. 6, 369–380

Mason, H.S., Mullet, J.E., and Boyer, J.S. (1988) Polysomes, messenger RNA and growth in soybean stems during development and water deficit. Plant Physiol., 86, 725–733.

Matyssek, R., Maruyama, S. and Boyer, J.S. (1988) Rapid wall relaxation in elongating tissues. Plant Physiol., 86, 1163–1167.

Meidner, H. (1975) Water supply, evaporation, and vapour diffusion in leaves. J. Exp. Bot. 26, 666–673

Meidner, H. (1976) Vapour loss through stomatal pores with the mesophyll tissue excluded. J. Exp. Bot. 27, 172–174

Meidner, H. (1976) Water vapour loss from a physical model of a substomatal cavity. J. Exp. Bot., 27, 691–694.

Meyer, R.F., and Boyer, J.S. (1981) Osmoregulation, solute distribution, and growth in soybean seedlings having low water potentials. Planta, 151, 482–489.

Michelena, V.A., and Boyer, J.S. (1982) Complete turgor maintenance at low water potentials in the elongating region of maize leaves. Plant Physiol., 69, 1145–1149.

Molz, F.J., and Boyer, J.S. (1978) Growth-induced water potentials in plant cells and tissues. Plant Physiol., 62, 423–429.

Molz, F.J. and Ferrier, J.M. (1982) Mathematical treatment of water movement in plant cells and tissues: a review. Plant Cell Environ., 5, 191–206.

Morgan, J.M. (1984) Osmoregulation and water stress in higher plants. Annu. Rev. Plant Physiol., 35, 299–319.

Moss, D.N., and Rawlins, S.L. (1963) Concentration of carbon dioxide inside leaves. Nature 197, 1320–1321

Nonami, H. and Boyer, J.S. (1987) Origin of growth-induced water potential: solute concentration is low in apoplast of enlarging tissues. Plant Physiol., 83, 596–601.

Nonami, H. and Boyer, J.S. (1989) Turgor and growth at low water potentials. Plant Physiol., 798–804.

Nonami, H. and Boyer, J.S. (1990a) Primary events regulating stem growth at low water potentials. Plant Physiol., 93, 1601–1609.

Nonami, H., and Boyer, J.B. (1990b) Wall extensibility and cell hydraulic conductivity decrease in enlarging stem tissues at low water potentials. Plant Physiol., 93, 1610–1619.

Nonami, H. and Schulze, E.-D. (1989) Cell water potential, osmotic potential, and turgor in the epidermis and mesophyll of transpiring leaves. Planta, 177, 35–46.

Nonami, H., Schulze, E.-D. and Ziegler, H. (1991) Mechanisms of stomatal movement in response to air humidity, irradiance and xylem water potential. Planta, 183, 57-64.

Oechel, W.C., Strain, B.R. and Odening, W.R. (1972) Tissue water potential, photosynthesis, 14C-labeled photosynthate utilization, and growth in the desert shrub Larrea divaricata. Cav. Ecol. Monogr., 42, 127-141.

Probine, M.C. and Preston, R.D. (1962) Cell growth and the structure and mechanical properties of the wall in internodal cells of Nitella opaca. II. Mechanical properties of the walls. J. Exp. Bot. 13, 111-127.

Rand, H.H. (1977) Gaseous diffusion in the leaf interior. Trans. Am. Soc. Agr. Eng. 20, 701-704

Rayan, A. and Matsuda, K. (1988) The relation of anatomy to water movement and cellular response in young barley leaves. Plant Physiol., 87, 853-858.

Scholander, P.F., Hammel, H.T., Bradstreet, E.D. and Hemmingsen, E.A. (1965) Sap pressure in vascular plants. Science, 148, 339-346.

Shackel, K.A., and Brinckmann, E. (1985) In situ measurement of epidermal cell turgor, leaf water potential and gas exchange in *Tradescantia virginiana* (L.). Plant Physiol. 78, 66-70

Sharkey, T.D., Imai, K., Farquhar, G.D., and Cowan, I.R. (1982) A direct confirmation of the standard method of estimating internal partial pressure of CO_2. Plant Physiol. 69, 657-659

Sharp, R.E. and Davies, W.J. (1979) Solute regulation and growth by roots and shoots of water-stressed maize plants. Planta, 147, 43-49.

Sheriff, D.W. and Meidner, H. (1974) Water pathways in leaves of Hedera helix L. and *Tradescantia virginiana* L. J. Exp. Bot., 25, 1147-1156.

Steudle, E. (1989) Water flow in plants and its coupling to other processes: an overview. In S. and B. Fleisher, (Eds.) Methods in Enzymology, Volume 174, Biomembranes Part U, Cellular and Subcellular Transport: Eukaryotic (Nonepithelial) Cells. Academic Press, New York. pp183-225.

Steudle, E. and Boyer, J.S. (1985) Hydraulic resistance to water flow in growing hypocotyl of soybean measured by a new pressure-perfusion technique. Planta, 164, 189-200.

Steudle, E. and Frensch, J. (1989) Osmotic responses of maize roots. Water and solute relations. Planta, 177, 281-295.

Steudle, E. and Jeschke, W.D. (1983) Water transport in barley roots. Planta, 158, 237–248.

Theologis, A. (1986) Rapid gene regulation by auxin. Annu. Rev. Plant Physiol., 37, 407–438.

Tyree, M.T., and Yianoulis, P. (1980) The site of water evaporation from sub-stomatal cavities, liquid path resistances and hydroactive stomatal closure. Ann. Bot. (NS) 46, 175–193

Van Volkenburgh, E and Boyer, J.S. (1985) Inhibitory effects of water deficit on maize leaf elongation. Plant Physiol. 77, 190–194

Westgate, M.E. and Boyer, J.S. (1984) Transpiration- and growth-induced water potentials in maize. Plant Physiol., 74, 882–889.

Westgate, M.E. and Boyer, J.S. (1985) Osmotic adjustment and the sensitivity of leaf, root, stem, and silk growth to low water potentials in maize. Planta, 164, 540–549.

Chapter 13

GROWTH REGULATION IN PLANT FACTORIES AND GREENHOUSES FROM A PHYSIOLOGICAL VIEWPOINT

Hiroshi Nonami

Department of Biomechanical Systems,
College of Agriculture, Ehime University,
Tarumi, Matsuyama 790, Japan

I. INTRODUCTION

Plant growth is one of the most sensitive physiological mechanisms of plants to environmental stresses (Kramer, 1983), and even slight changes in environmental conditions in plant factories and greenhouses can affect the growth of crops. In order to cultivate high quality crops under a controlled environment, methods for adjustment of environmental conditions have been studied. However, most methods that have been developed for regulating climatic conditions for crops have not included a means for evaluating the physiological status of crops. Growth of crops is most commonly measured in terms of dry matter accumulation. However, measurements of dry matter accumulation are not suitable for use as a parameter for control in order to regulate growth of crops in plant factories and greenhouses, as such measurements must be evaluated over a long period of time such as days and weeks. Significant changes in cell volume expansion, on the other hand, occur in a matter of hours in most plants, and thus, are the most appropriate parameter for growth regulation. Because more than 90% of the total molecules in expanding cells of most plants are water molecules, water movement into cells regulates cell expansion rates in most cases. Thus, we have studied plant–water relations of crops under controlled environments, and will suggest new types of growth regulation methods for crops grown under controlled environments.

In order to explain the mechanisms of physiological regulation for plant growth to readers, I will primarily use a simple model system in which the mechanisms of cell elongation in soybean seedlings were studied intensively. Although such a seedling system is not used in practical plant factories and greenhouses, the principle derived from the seedling system is applicable for plants grown in plant factories and greenhouses. To demonstrate the applicability of the principles of growth regulation in plants grown in plant factories and greenhouses, plants were grown hydroponically, and it will be shown that the growth mechanisms found in hydroponically grown plants is the same as those of soybean seedlings. Furthermore, it will be shown that the same principles are applicable to plants grown under tissue-culture conditions.

II. GROWTH INHIBITION AND WATER STATUS

In order to demonstrate the sensitivity of growth under water stress, soybean seedlings (*Glycine max* (L.) Merr.) were grown from seeds at 29 ± 0.5 °C and 100% RH in the dark in water-saturated vermiculite. After 60 h, each seedling was subjected to water stress by being transplanted to a 200 ml beaker containing vermiculite having -0.28 MPa of water potential. When soybean seedlings were transplanted from vermiculite having -0.01 MPa to -0.28 MPa, stem elongation was inhibited immediately (Fig. 1A).

The water potential was measured by the isopiestic technique (Boyer and Knipling, 1965). A thermocouple chamber was coated with melted and resolidified petrolatum (Boyer, 1967) and loaded with stem segments about 15 mm long. Water potential of the elongation zone and the mature zone was measured. The xylem water potential was estimated from the water potential of the mature region, because water potential of the xylem is equilibrated with that of the mature region as shown by Nonami and Boyer (1990a).

Immediately after application of water stress by transplanting, the xylem water potential decreased and became much lower than water potential of the growing medium (i.e., vermiculite having -0.28 MPa of water potential) (Fig. 1B). Water potential of the zone of elongation decreased after water stress was applied (Fig. 1B). However, decreases in water potential of the elongation zone

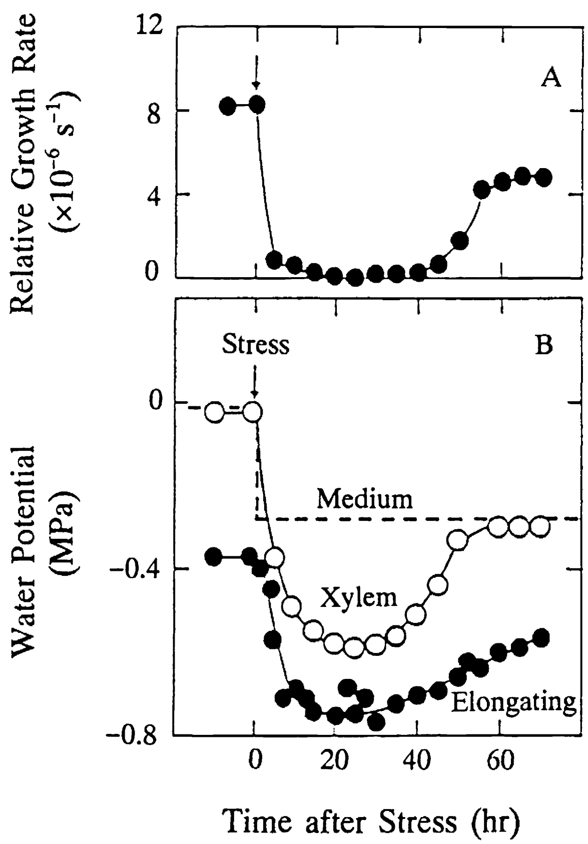


Figure 1. Relative growth rate (A) and water potentials (B) of vermiculite (Medium), the xylem (Open circles) and the zone of elongation (Closed circles).

were slightly slower to occur than decreases in the xylem water potential (Fig. 1B). Eventually, the xylem water potential equilibrated with the water potential of the medium (Fig. 1B), and then, growth recovery was initiated (Fig. 1A).

From this example, it is apparent that a slight change in water potential of the growing medium causes a drastic change in plant growth. I will next introduce theories of cell expansion by using fundamental physical laws to explain growth regulation under environmental stresses.

III. THEORY OF PLANT GROWTH

The theory of water movement through an isotropic tissue is based on the hypothesis that the rate of transfer of a diffusing substance through a unit area of a section is proportional to the concentration gradient measured normal to the section (Fick's law). Because the water content of plant tissue is related to the tissue water potential and because water flux must be proportional to the water potential gradient, the relation between water flux (J) and water potential (Ψw) can be written as follows;

$$J = -D \text{ grad } \Psi w \qquad (1)$$

where D is the diffusion coefficient. It is not necessary to know the microscopic pathway for water movement in order to apply this concept. However, it is likely that water moves from cell to cell along gradients in potentials that are in local equilibrium between the cell wall and the protoplast (Molz, 1976a; Molz and Ikenberry, 1974).

Water moving into plant tissue must either pass out of the tissue or remain within the tissue. This principle, from the law of conservation of mass, can be stated as the divergence theorem according to;

$$\frac{\partial \Psi w}{\partial t} = -\text{div } J \qquad (2)$$

where $\partial\Psi w/\partial t$ indicates the water potential increase per unit time within the cells (hydration) and div J is the difference in influx and efflux per unit volume.

If in addition some portion of the influx is used for cell enlargement, we represent this process by a sink term in the conservation law. The equation can be modified to

$$\frac{\partial \Psi w}{\partial t} = -\text{div } J - S \qquad (3)$$

where S is a sink term used to indicate the increase in cell volume due to growth. The sink term is negative because growth removes a fraction of the water from the component undergoing strictly influx, efflux, and hydration changes. Combining Eqs. 1 and 3 yields;

$$\frac{\partial \Psi w}{\partial t} = D \text{ div grad } \Psi w - S \qquad (4)$$

which states that the change in hydration of a tissue will be determined by the ability of the tissue to transmit water, represented by the diffusivity, and the water potential gradients in the growing tissue.

In the case of modeling stem elongation of plants, it is assumed that stems have a cylindrical shape in the present study. Thus, cylindrical coordinates can be used to simplify the analysis. Expressing Eq. 4 in these coordinates;

$$\frac{\partial \Psi w}{\partial t} = D \frac{\partial^2 \Psi w}{\partial r^2} + \frac{D}{r} \frac{\partial \Psi w}{\partial r} - S \qquad (5)$$

Equation 5 describes the growth-induced water potential field in the zone of elongation of a cylindrical tissue.

In order to apply Eq. 5, it is necessary to obtain a reasonable estimate for D. This can be done by evaluating the theoretical expression for D in terms of protoplast and cell wall hydraulic and elastic properties. One may utilize the expression given by Molz (1976a, 1976b) and Molz and Ikenberry (1974);

$$D = \frac{\Delta x \,(Pa + K\Delta x A/2\,)}{W_v s + V_0/(\varepsilon + \pi)} \qquad (6)$$

where A = cross-sectional area of vacuolar pathway, a = cross-sectional area of cell wall pathway, π_0 = osmotic pressure at zero turgor, x = diameter of cell, K = permeability of cytoplasmic complex separating cell wall from vacuole, V_0 = cell volume at zero turgor pressure, W_v = volume of cell wall/cell, P = hydraulic conductivity of the cell wall, and s = specific water capacity of the cell wall material.

Molz and Boyer (1978) theoretically derived S, assuming local equilibrium between the protoplast and cell wall. Defining G = $(1/V_0)\partial V_0/\partial t$, S is given by

$$S = (\varepsilon + \pi_0)G \qquad (7)$$

Although D and S in Eq. 5 are given, analytical solutions for Eq. 5 are still difficult to obtain. In order to simplify Eq. 5 further, we assume that changes in water potentials with respect to time are much slower than changes in water potentials with respect to position within the elongating stem tissue. This is very likely under natural conditions because influences of changes in the surrounding water status first appear in the root region and gradually move to the elongating region of the stem. If so, $\partial\Psi w/\partial t$ in Eq. 5 can be determined directly from slopes of a curve plotted with water potential changes with respect to time. Thus, Eq. 5 can be modified to

$$\psi_w(t) = D\frac{\partial^2\Psi w}{\partial r^2} + \frac{D}{r}\ \frac{\partial\Psi w}{\partial r} - S \qquad (8)$$

where $\psi_w(t)$ is changes in water potential at a given time and can be replaced by the measured values. If one defines the constant α as follows;

$$\alpha = (\psi_w(t) + S)/D,$$

then Eq. 8 becomes;

$$\frac{d}{dr}\left(\frac{d\Psi w}{dr}\right) + \frac{1}{r}\ \frac{d\Psi w}{dr} = \alpha \qquad (9)$$

Substituting $M = d\Psi w/dr$ reduces Eq. 9 to a linear first order equation in M. Defining $M/r = N$ and differentiating this relationship with respect to r allows further rearrangement of Eq. 9 to give;

$$r\ \frac{dN}{dr} + 2N = \alpha$$

which can be integrated by standard techniques to give the general solution;

$$\Psi w = \frac{\alpha}{4} r^2 - \frac{C_1}{2} \ln r + C_2 \qquad (10)$$

where C_1 and C_2 are constants of integration. First, we will solve this equation in the cortex where it is assumed that $d\Psi w/dr = 0$ at the epidermal surface of the stem and Ψw is equal to the xylem water potential at the region which contains the xylem. For the pith, the approach is similar, but the boundary conditions are $d\Psi w/dr = 0$ at the center of the stem.

IV. GROWTH-INDUCED WATER POTENTIAL FIELDS

When the water potential of the elongation zone was measured with psychrometry, the water potential indicated the average water potential of cells contained in the zone of elongation. If changes in water potential measured with psychrometry are considered to be equivalent to $\psi_w(t)$ in Eq. 8, water potential fields associated with growth can be determined by applying the boundary conditions and using Eq. 10. The xylem water potential was measured in Fig. 1B and $\psi_w(t)$ was determined from the slope of the curve of the water potential of the elongating region in Fig. 1B. Values of biological parameters to calculate D were adopted from Molz and Boyer (1978).

Figure 2 shows water potential fields formed in the elongating region at a given time after application of water stress. Before applying water stress, water potential was lowest next to the epidermal surface and the highest in the xylem region (see a thin line marked with Oh in Fig. 2A). Such a potential field associated with actively growing tissue is called a growth-induced water potential field. Immediately after application of water stress, the xylem water potential decreased, but the water potential near the epidermal surface was unchanged (see the line labelled "5h" in Fig. 2A). Progressively, water potential fields flattened out across the stem, and at the 20th h after application of stress, water potential gradients disappeared completely across the stem (Fig. 2A). Recovery of water potential gradients was initiated by the rise in the xylem water potential. When the growth began to recover at the 45th h after application of stress (Fig. 1A), the growth-induced water potential field was gradually formed (Fig. 2B). At the 65th h, the growth-induced water potential field was formed to support the partial recovery of growth after application of stress (Figs. 1A and 2B).

Cell water potentials could be measured experimentally in the elongating region of soybean hypocotyls. Prior to the measurement, the seedling was coated with petrolatum and further covered with wet tissue paper on the petrolatum to minimize water loss from its surface. Cell turgor was measured with a cell pressure probe according to the procedure described by Nonami et al. (1987). Immediately after cell turgor determination, cell solution was extracted from the same cell with the probe, and transferred to a nanoliter osmometer (Clifton Technical Physics, Hartford, N.Y. U.S.A.) in order to determine cell osmotic potential according to

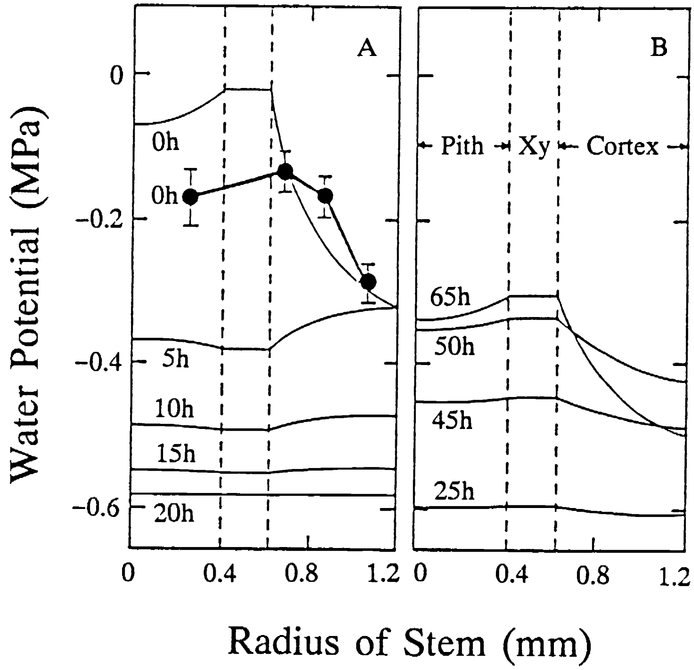


Figure 2. Water potential profiles across the stem of soybean seedlings in the zone of elongation during growth inhibition (A) and growth recovery (B). Water potential profiles are shown in the pith region (Pith), the xylem region (Xy) and the cortical region (Cortex). Potential profiles shown in thin lines were calculated, and time is indicated in hours after the application of stress. Cell water potentials (closed circles) were measured directly, and the vertical bars indicate 95% confidence intervals.

Nonami and Schulze (1989). Cell water potential was obtained by algebraic summation of the turgor and osmotic potential in the same cell (Nonami and Schulze, 1989). All manipulations were conducted under green safelight.

In order to prove the existence of the growth-induced water potential field experimentally, the cell water potential across the stem was measured in actively growing soybean seedlings before applying water stress (closed circles marked with 0h in Fig. 2A). In the cortex region, measured cell water potentials were similar to water potential fields predicted theoretically (0h in Fig. 2A). This indicates that a growth-induced water potential field existed and was extending from the xylem toward the epidermal surface in order to support synchronized cell expansion across the stem. Slight deviations between measured values and theoretically-determined values were observed in the pith region (Fig. 2A).

Because the collapse of the growth-induced water potential field causes immediate cessation of water flow from the xylem to expanding cells, growth of plants can be inhibited when water stress is applied to plants. Although plant growth can be evaluated by measuring the growth-induced water potential field, measurements of the field cannot be made easily in a practical manner. Thus, the simplest way to estimate the growth-induced water potential field will be introduced here.

V. PRACTICAL ESTIMATION OF THE GROWTH-INDUCED WATER POTENTIAL FIELD

Because the growth-induced water potential field is initiated from the xylem water potential and formed by cell water potentials in the growing tissue, water potential difference between the elongating region and xylem may be a good estimate of the magnitude of the growth-induced water potential field. In order to test this hypothesis, kidney bean seedlings were grown hydroponically, and growth and water potentials of the elongating region and mature region were measured simultaneously.

Kidney bean (*Phaseolus vulgaris* L. cv. Shiraginugasa) seedlings were grown hydroponically in nutrient solution (in $mol \cdot m^{-3}$: 0.33 NH_4^+, 4.0 NO_3^-, 0.27 PO_4^{3-}, 4.0 K^+, 0.14 Mg^{2+}, 3.5 Ca^{2+}, 0.62 SO_4^{2-}, 0.014 Fe^{3+}, 0.009 Mn^{2+}: EC = 0.1 $S \cdot m^{-1}$, water potential = −0.07 MPa) at 29 ± 0.5 °C, 40 ± 10% RH and 150 $\mu mol \cdot m^{-2} \cdot s^{-1}$ of photosynthetically active photon flux density. Water stress was applied to the seedlings by increasing concentra ions of the culture solution by 5-fold (−0.35 MPa of water potential) or 7.3-fold (−0.51 MPa of water potential).

When water potential of the culture solution was decreased from −0.07 MPa to −0.35 MPa, growth of kidney bean seedlings was slightly promoted (Fig. 3A). However, immediately after water potential of the culture solution was decreased from −0.07 MPa to −0.51 MPa, growth of kidney bean seedlings was inhibited (Fig. 3B). After 10 h of the water stress, gradual recovery of growth of the seedlings was observed (Fig. 3B).

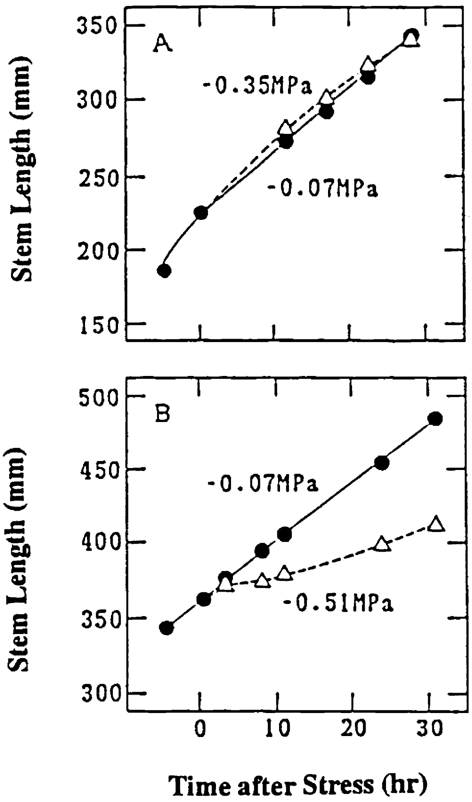


Figure 3. Stem length of kidney bean seedlings when the nutrient solution water potential was increased from -0.07 MPa to -0.35 MPa (A) and from -0.07 MPa to -0.51 MPa (B).

In order to understand the mechanism of growth inhibition by water stress created by increasing the concentration of the nutrient solution, the growth-induced water potential was estimated from the differences between water potentials of the elongation region and mature region. In plants which did not have growth inhibition under -0.35 MPa of the culture solution, the growth-induced water potential was unchanged (Fig. 4B). In plants whose growth was inhibited under -0.51 MPa of the culture solution, the growth-induced water potential decreased immediately after the application of stress, and began to recover prior to growth recovery (Fig. 4D). It is evident that growth of plants is regulated by the growth-induced water potential field.

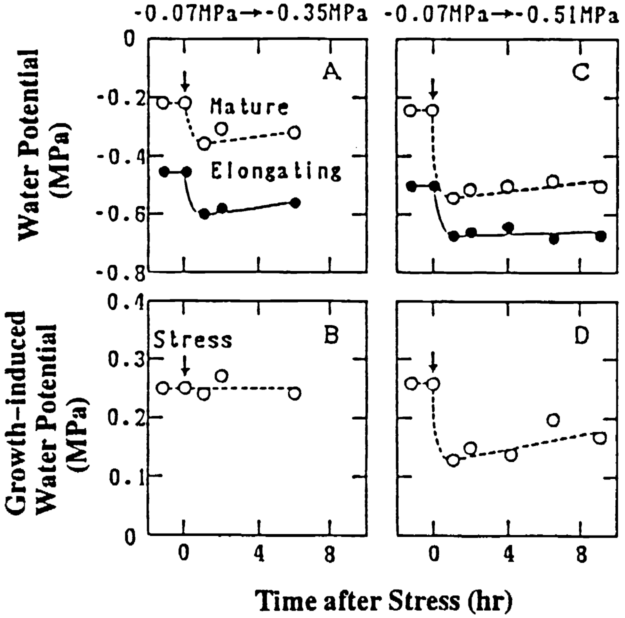


Figure 4. Water potentials of the mature region and the elongation region (A, C) and the growth-induced water potential (B, D) when the nutrient solution water potential was increased from -0.07 MPa to -0.35 MPa (A, B) and from -0.07 MPa to -0.51 MPa (C, D). The growth-induced water potential was estimated from the water potential difference between the elongation region and the mature region.

VI. CELL WALL PROPERTIES AND GROWTH

Once growth is inhibited under water stress, growth is not usually recovered completely as shown in Figs. 1 and 3. This may not be simply caused by changes in the growth-induced water potential field, but also by changes in cell wall properties of plants. In order to evaluate physical properties of cell walls, I will introduce theories of cell wall properties associated with growth. The theories have been derived from fundamental physical laws in rheology, and they let us make physiological measurements in intact growing plant tissues.

Plant tissues are considered to consist of a continuous viscoelastic polymer (cell walls) containing extremely small compartments (cells) with bounding membranes. Water moves throughout the polymer in the walls and cells. Internal pressures exert forces in 3-dimensional directions inside the cells. For experimental purposes, first, force is considered to be applied in 1 direction to the entire tissue.

An external force applied in one direction to plant tissues will cause a deformation in proportion to the force if the force is small (Alfrey, 1948, Wainwright et al., 1976). If the walls consist of mixtures of ideal elastic (Hookian) and ideal nonelastic viscous materials (Newtonian), the Boltzmann superposition principle indicates that the deformation can be split linearly into an elastic component and a plastic component. Energy used for elastic deformation is conserved whereas energy for plastic deformation partly rearranges the molecules and partly is dissipated as heat.

Elastic deformation is instantaneous, reversible, and independent of time and thus,

$$d\gamma = E\, d\sigma \qquad (11)$$

where γ is the strain (relative change in dimensions, unitless), σ is the stress (applied force per unit area, or pressure, i.e., the unit is $N \cdot m^{-2}$ or MPa), and E is the elastic compliance ($m^2 \cdot N^{-1}$ or MPa^{-1}, the inverse of Young's modulus of elasticity).

Plastic deformation is neither instantaneous nor reversible, and it occurs continuously at a rate in proportion to the stress, and thus,

$$d\gamma/dt = M\sigma \qquad (12)$$

where M is the plastic deformability ($m^2 \cdot N^{-1} \cdot s^{-1}$ or $MPa^{-1} \cdot s^{-1}$).

The superposition principle allows the sum of Eqs. 11 and 12 to be obtained after differentiating Eq. 11 by time to give the governing equation as follows;

$$d\gamma = E\, d\sigma/dt + M\sigma \qquad (13)$$

Equation 13 shows that there will be an elastic deformation only when the force changes but a plastic deformation that progresses as long as force is applied. This concept is useful for distinguishing elastic deformation from plastic deformation while growth is occurring in the intact plant.

A plant stem to which a force is applied in one direction parallel to the long axis (z-direction) will stretch instantaneously. Before significant time elapses, plastic deformation is zero so that Eq. 13 reduced to Eq. 11 stated as;

$$dZ/Z = E\,d\sigma = E\,d(F/A) \qquad (14)$$

where Z is the length in the z-direction, F is the force, and A is the cross-sectional area of the tissue.

On the other hand, because cell turgor stretches the wall 3-dimensionally, causing a fractional volume change dV/V, the volume change is related to the length change for isotropic materials by;

$$dV/V = 3(1-2\upsilon)dZ/Z \qquad (15)$$

where υ is Poisson's ratio, a unitless term that corrects unidirectional dimensions for contributions from other directions and is defined by $dX/X = dY/Y = -\upsilon dZ/Z$ (Alfrey, 1948; Feynman et al., 1964; Wainwright et al., 1976). Poisson's ratio varies between 0 and 1/2. For nonisotropic materials, the relation $3(1-2\upsilon)$ must be modified to a more complex form.

Substituting dZ/Z from Eq. 15 into Eq. 14 gives;

$$dV/V = 3(1-2\upsilon)E\,d\Psi p = B\,d\Psi p \qquad (16)$$

where Ψp is F/A operating in 3-dimensions (e.g., the turgor) and

$$B = 3(1-2\upsilon)E \qquad (17)$$

B is the bulk compliance and has units of $m^2 \cdot N^{-1}$ or MPa^{-1} (Wainwright et al., 1976), i.e., B is the inverse of the bulk modulus of elasticity ε (MPa). From Eq. 16, B can be measured with a pressure probe and, from Eq. 14, E can be measured with an extensiometer. Poisson's ratio can be measured from changes in dimension. Thus, Eq. 17 allows 3-dimensional and 1-dimensional elasticity measurements to be compared.

Many materials deform irreversibly only when the applied stress exceeds a minimum necessary for molecular rearrangement. The minimum force is termed the yield threshold and is observed in cell walls that, under a constant stress exceeding the threshold, deform continuously and irreversibly, resulting in cell enlargement. The deformation appears to be linear (Newtonian) with the applied stress (Cleland, 1959; Green et al., 1974; Matthews et al., 1984) and thus can be considered ideal. Over significant time with constant stress, $d\sigma/dt$ is zero (Eq. 13) and the elastic component does not contribute. Thus, for a stress applied in one dimension, Eq. 13 becomes;

$$(1/Z)dZ/dt = M\ (F/A) \qquad (18)$$

In a stem with cross-sectional area A that is essentially constant, 1-dimensional and 3-dimensional strains are related by $(1/Z)(dZ/dt) = (1/AZ)(AdZ/dt) = (1/V)(dV/dt) = G$ (relative growth rate, s−1). Also, because F/A operating in 3 dimensions is the pressure, the F/A for plastic deformation in 3 dimensions is equal to the growth-effective turgor of the cell and given by;

$$(F/A) = (\Psi p - Y) \qquad (19)$$

where Ψp is the cell turgor and Y is the yield threshold which indicates the minimum turgor necessary to expand the cell (Green et al., 1971). Wall extensibility is defined operationally as a coefficient m with the relative growth rate G when the growth-effective turgor $(\Psi p - Y)$ is known in intact plants (Green et al., 1971; Boyer et al., 1985; Nonami and Boyer, 1990a);

$$G = m\ (\Psi p - Y) \qquad (20)$$

where m has units of $MPa^{-1}\cdot s^{-1}$. Combining Eqs. 18, 19 and 20, the plastic deformability measured in 1 dimension is then derived to be equal to the extensibility m measured in 3 dimensions as follows;

$$(1/Z)(dZ/dt)(A/F) = G/(\Psi p - Y) = M = m \qquad (21)$$

From Eq. 20, m can be evaluated with a psychrometer (Boyer et al., 1985; Nonami and Boyer, 1990a), and from Eq. 18, M can be measured with an extensiometer (see the following section), and Eq. 21 allows these 3-dimensional and 1-dimensional measurements to be compared. Although this conclusion can be complicated by the synthesis of new wall material, in principle the effect of synthesis can be tested by determining M when the tissue is synthesizing new wall material and when it is not.

VII. MEASUREMENTS OF CELL WALL PROPERTIES

In order to apply a force to soybean stems (hypocotyls), an extensiometer was made to stretch a section of the intact stem. Soybean stems could be subjected to small extensions by hanging various weights on a nylon thread and leading the thread over a pulley to the stem where it was connected to the hypocotyl hook of the intact plant. Extension occurred upward as the force was applied and was measured with a radial displacement transducer whose arm was attached adjacent to the thread attachment. A rigid reference bar was attached at a position immediately below the elongating region. The diameter of the stem and the distance between the attachments were measured accurately with a caliper. The body of the transducer was mounted in the barrel of a microscope so that calibration could be carried out at any time without disturbing the plant materials. Soybean seedlings could grow without restriction in the vermiculite while attached to the transducer. The experiments were conducted in the growth environment with the entire apparatus and seedling in a dark box with air humidity of nearly 100%.

Application of an upward force caused extension of the stem, and removal of the force caused contraction of the stem (Fig. 5). If significant time elapsed, extension was always greater than contraction, as expected if extension contained an elastic component plus a plastic deformation but contraction contained only an elastic component. Thus, the plastic component could be measured from the difference between the extension and the contraction when a period of time had elapsed between them.

The lengthwise changes required about 3 to 6 min to achieve a new steady slope after a force of 0.196 N was applied (Fig. 5). Initially, the growth rate was not markedly different from that before the stress was applied, but if the force was maintained for longer than about 10 min, the elongation rate accelerated, suggesting that the cells were acclimating to the new stress. To avoid possible changes in wall properties, the measurements were completed in the initial 3 to 6 min. The plastic component of the total extension is defined to be the amount of length increase between the time immediately after force application (immediately after elastic extension) and the time to reach a new steady rate (Fig. 5). The endogenous growth rate was eliminated from the measurement by extrapolating to the time of force application a line having the same slope as before the force application (as in Fig. 5).

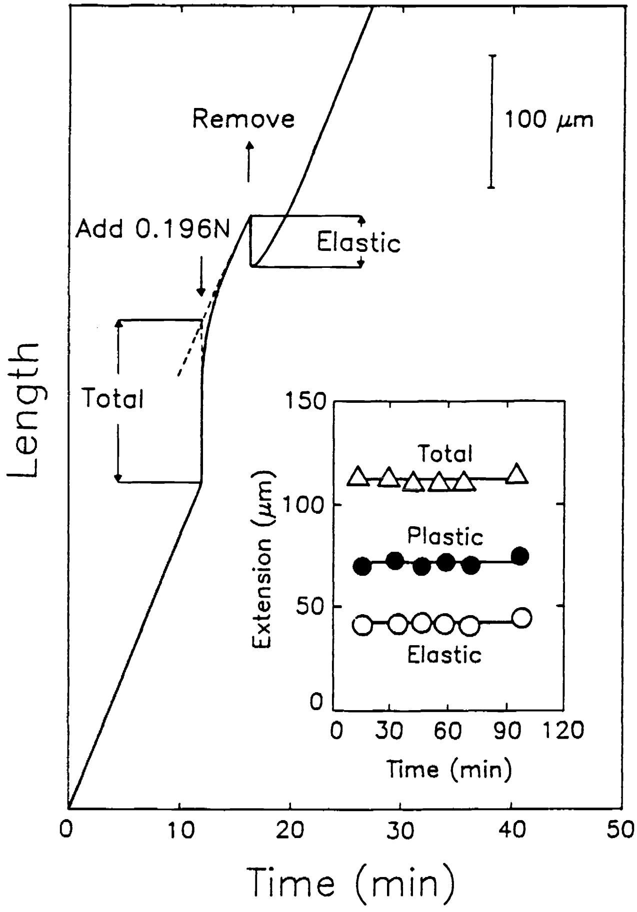


Figure 5. Extension and contraction induced by application and removal of a force of 0.196 N (equivalent to 20 g) to a stem of an intact soybean seedling having a growth rate of 0.47 $\mu m \cdot s^{-1}$. The force was applied to the elongating region at the downward arrow and removed at the upward arrow. Plastic extension was defined as the difference between the total extension (extrapolated at downward arrow) and the elastic contraction (upward arrow). The dashed line having a slope shows the growth rate before force application and, when extrapolated to the time of force application, gives the total extension that occurred in initial 3 to 6 min after force was applied (in this case 3.2 min). Inset: repeatability of total, plastic and elastic extension during repetitive force application (0.196 N) and removal from the same elongating region by using the same seedling.

Elastic extension is defined as the length of contraction within 30 sec after removal of the force (Fig. 5). Because the growth of the plant resumed within 1 to 3 min after the force was removed (Fig. 5), stress relaxation was very rapid in intact plants, confirming earlier observations (Boyer et al., 1985; Matyssek et al., 1988). Stress extension and relaxation could be repeated in the same plant without change (inset of Fig. 5).

In the growth process, cell expansion occurs due to cell wall extension coupled with water influx into cells, and thus, studies of cell wall properties are inseparable from those of water movement. Thus, cell wall extension experiments were conducted by using soybean seedlings grown as in Fig. 1 in order to compare properties of both water movement and cell wall extension during growth under water stress. Soybean seedlings grown in the dark were transplanted to water-saturated (1X) vermiculite having -0.01 MPa of water potential or water-deficient (1/8X) vermiculite having -0.28 MPa of water potential. In order to obtain elastic compliance E and plastic deformability M, various forces were applied to the soybean tissue by using the extensiometer.

When various forces were applied to the same tissue (Fig. 6A), the stress-strain relations were linear for both the elastic and plastic components (Fig. 6B-6D) and there was no evidence of a yield threshold, i.e., the relationship for the plastic component extrapolated to zero when the added stress was zero. This indicates that cell turgor maintained the stress above the yield threshold while the tissue was growing. Because both the elastic and plastic components were linearly related, both quantities could be treated separately according to the superposition principle. Thus, it is apparent that the application of Eq. 13 to this seedling system is valid from a physical viewpoint. Also, I would like to mention that a domain of forces applied here was intentionally chosen so that the linearity of elastic and plastic properties was warranted. When a force was applied to the tissue beyond the proper domain, strain-stress relations became non-linear.

We applied various extending forces to a maximum of 0.343 N to stems that averaged 1.1 x 10^{-3} m in radius. This was equivalent to about 0.9 x 10^{5} $N \cdot m^{-2}$ of stem cross-section or about 20% of the force exerted in the lengthwise direction by the cell turgor of 4.5 x 10^{5} $N \cdot m^{-2}$ = 0.45 MPa. On average, the elongating region of the stems increased to 1.0048x their original length Z and the diameter decreased to 0.998x the original diameter Y when the maximum force of 0.343 N was applied to the elongating tissue instan-

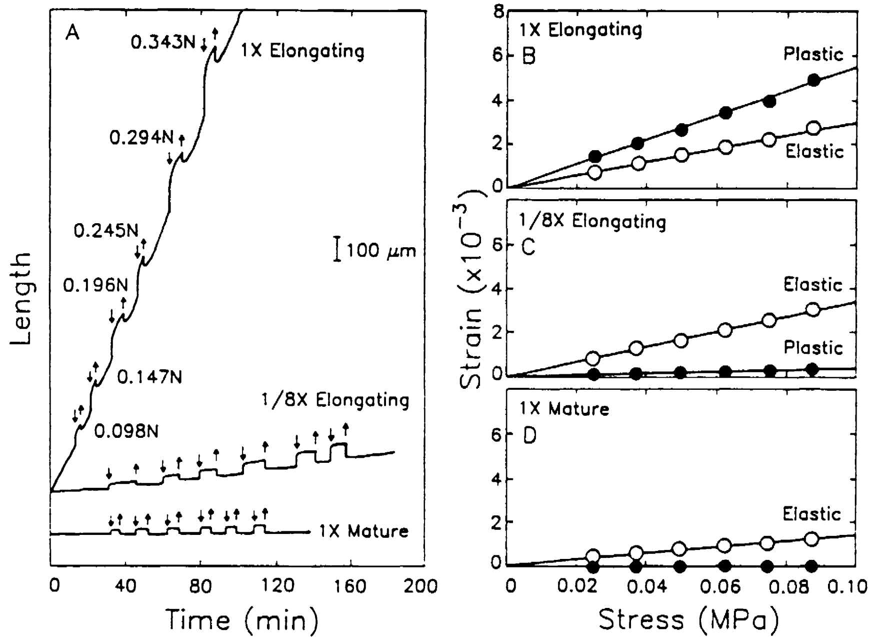


Figure 6. A: Extension and contraction induced by sequential force application and removal (i.e., forces of 0.098 N, 0.147 N, 0.196 N, 0.245 N, 0.294 N and 0.343 N) from the stem of intact soybean seedlings growing in water-saturated (1X) vermiculite having −0.01 MPa of water potential or water-deficient (1/8X) vermiculite having −0.28 MPa of water potential. Measurements were made in the elongating region (1X Elongating and 1/8X Elongating or in the mature region (1X Mature). The forces were applied at downward arrows and removed at upward arrows. In A, 1X Elongating was measured while the stem grew at 0.47 $\mu m \cdot s^{-1}$ in the 17th h after transplanting; 1/8x Elongating was measured while the stem grew at 0.01 $\mu m \cdot s^{-1}$ in the 22nd h after transplanting; and 1X Mature was measured while the stem grew rapidly but the mature region showed zero elongation 21 h after transplanting. B,C,D: Stress-strain relations determined from the length changes in A. Stress (MPa) was calculated from force (N) applied per unit cross-sectional area of stem calculated from stem diameter measured with a caliper. Open circles and closed circles indicate elastic components and plastic components, respectively.

taneously. This gave a Poisson's ratio υ = -(dY/Y)/(dZ/Z) = -(-0.002/1)(0.0048/1) = 0.42 ± 0.02 (95% confidence interval) for the living tissue, which was close enough to 0.5 that any nonideal elastic behavior was ignored.

In the elongating region of rapidly growing stems, the plastic component was larger than the elastic component (Fig. 6B). In stems growing slowly because of exposure of the plants to low water potential, the plastic component was smaller than the elastic component (Fig. 6C). The elastic component was unaffected by growth inhibition (Figs. 6B and 6C). In the mature region, the elastic component was small and there was no plastic component (Fig. 6D).

An average plastic deformability could be obtained from plots like Fig. 6 by determining the slope of the stress-strain relations (dZ/Z)/(F/A) and dividing by the average time to reach the new steady state (1/dt) according to Eq. 18; (dZ/Z)(1/dt)/(F/A) = M. When we used this approach to follow changes in the plastic deformability of seedlings exposed to low water potential, the deformability remained high in the controls (1X) but decreased during exposure of the seedlings to low water potential (1/8X), then increased to an intermediate level (Fig. 7A). The initial decrease was apparent by 5 h after transplanting. These changes took place at about the same time as the wall extensibility changes measured with a psychrometer (Nonami and Boyer, 1990a) and were numerically similar (Nonami and Boyer, 1990a), confirming that plastic deformability was equal to wall extensibility. In the mature tissue, plastic deformability was zero (Fig. 7C).

An average elastic compliance E was measured similarly from the slopes of the stress-strain relations according to Eq. 17; (dZ/Z)/(F/A) = E. The compliance was the same in the controls and treated plants (Fig. 4B). In the mature zone, the compliance of growth-inhibited plants increased slightly over the control. (Fig. 7D).

It was found that the elastic property of cell walls of the elongating tissue is not directly affected by growth inhibition under water stress. Because cell turgor in the zone of elongation could be maintained under water stress due to osmoregulation (Nonami and Boyer, 1989a), cell volume of the tissue in the zone of elongation could be maintained unchanged while growth was inhibited.

When growth is inhibited by water stress, it is apparent that wall extensibility is also altered metabolically, and thus, recovery of growth is affected by changes in the cell wall. This means that

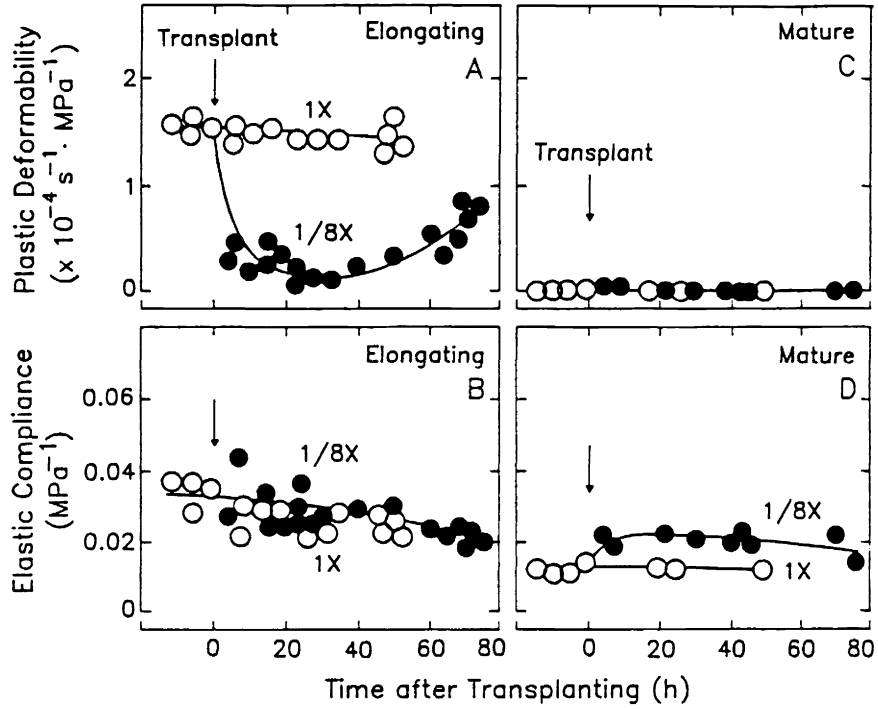


Figure 7. Plastic deformability (A, C) and elastic compliance (B, D) measured in the elongating zone (A, B) and mature zone (C, D) of stems of soybean seedlings transplanted to well-watered vermiculite (1X) or water-deficient vermiculite (1/8X). Measurements were made with the extensiometer. Open circles represent 1X plants and closed circles represent 1/8X plants. Each point is an individual measurement.

growth cannot be recovered completely although the size of the growth-induced water potential is recovered completely after osmoregulation.

VIII. GROWTH REGULATION UNDER SALT STRESS, COLD STRESS AND HEAT STRESS

It is known that growth is inhibited by salt stress, cold stress and heat stress. If the concept of growth regulation presented here is correct, inhibition of cell elongation may be explained by using the concept of growth-induced water potential even if the stress is not caused by water stress.

Increases in cell volume can be considered equivalent to the amount of water uptake by the cell. Because water must flow into cells during cell expansion, water potential gradients must exist between the water source and the expanding cells. When water potentials of individual cells were measured in actively elongating soybean stems, Nonami and Boyer (1989b) found that water potential was lowest near the epidermis and highest near the xylem in the elongation zone of actively elongating soybean stems. Such a water potential field associated with cell expansion is called a growth-induced water potential. The size of the growth-induced water potential can be determined by the difference of water potential between expanding cells and the water source. In the present study, growth rates were altered by differences in nutrient concentrations and growing temperatures. In order to simplify experimental procedure, environmental stresses were applied to embryos which had been removed from seeds to find the regulatory factor most responsible for cell expansion during periods of environmental stress.

Embryos were taken out from kidney beans (*Phaseolus vulgaris* L. cv. Shiraginugasa) under sterile conditions after the seeds were imbibed for one day in running tap water, and transplanted in agar-based tissue culture medium having various concentrations of MS salts (Murashige and Skoog, 1962). The highest concentration of the medium was the standard MS culture medium, diluted to 10% interval concentrations in order to determine the best culture medium concentration to enhance cell expansion of embryos. When embryos were grown in the best culture medium concentration, culture temperature was set between 10 to 40 °C in order to test the effect of culture temperature on cell expansion of embryos.

Water potentials of tissue and medium samples were measured by using the isopiestic psychrometer (Boyer and Knipling, 1965). Eight tissue samples which were 0.5–1.0 cm long were excised from the zone of elongation or maturation of stems. After water potential was measured, the osmotic potential was determined in the same tissue immediately after freezing at −70°C and thawing. Relations of cell water status can be written as follows;

$$\Psi w = \Psi p + \Psi s$$

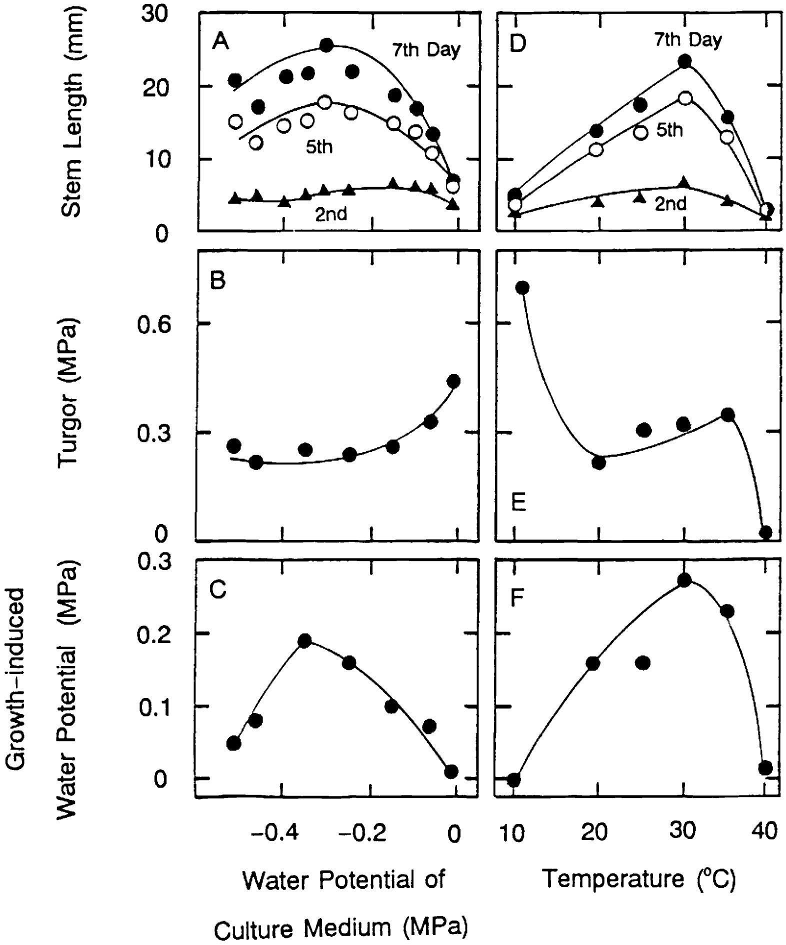


Figure 8. Stem length (A, D), turgor of the zone of elongation (B, E) and growth-induced water potential (C, F) when embryos were grown on tissue culture media having different water potentials (A, B, C) and when embryos were grown under different temperature (D, E. F). Triangles, open circles and closed circles in A and D indicate the average stem length of kidney bean plantlets on the 2nd, 5th and 7th day after embryos were planted on the tissue culture medium.

where Ψw, Ψp and Ψs indicate cell water potential, cell turgor and cell osmotic potential, respectively. This relation was experimentally confirmed by works of Nonami et al. (1987) and Nonami and

Schulze (1989). Thus, turgor was calculated by subtracting the osmotic potential from the water potential.

When embryos were grown on a tissue culture medium having -0.30MPa of water potential, the greatest elongation occurred in stems (Fig. 8A). Turgor of the elongation region of the stem was not related to rates of elongation in stems (Figs. 8A and 8B). Thus, turgor is not regulating the rates of cell expansion in embryos of kidney beans.

In order to correlate water potential gradients between expanding cells and the water source with cell elongation rates, the growth-induced water potential was determined from the differences between the water potential of the zone of elongation and that of maturation. The growth-induced water potential was the smallest when stem elongation was the smallest, and the growth-induced water potential was the largest when stem elongation was the largest (Figs 8A and 8C). It is safe to say that the size of the growth-induced water potential was correlated to rates of cell elongation.

Stem elongation was the largest at 30°C and inhibited at 10 and 40°C (Fig. 8D). Turgor of the elongation zone was the largest at 10°C and almost completely lost at 40°C (Fig. 8E). Turgor of cells in the zone of elongation was not related to rates of cell elongation.

The growth-induced water potential was the largest in plantlets grown at 30°C and almost zero in those grown at 10 and 40°C (Fig. 8F). It is evident that the water potential gradients between elongating cells and the water source regulate growth rates when cell elongation is affected by temperature.

The growth inhibition observed under culture conditions used here implies that cell elongation was inhibited by salt stress, cold stress and heat stress. In all cases, turgor of the elongation zone was unrelated to cell elongation rates, and the growth-induced water potential was related to growth of the plantlets. This indicates that cell elongation is primarily regulated by the size of the water potential gradient between expanding cells and the water source.

IX. DISCUSSION

In the present study, dark-grown soybean seedlings were used to demonstrate the concept of a growth-induced water potential field when the seedlings were subjected to water stress. Because dark-grown soybean seedlings have a restricted growing region in their stems and because the stems have a cylindrical shape, the theoretical analysis could be applied easily. Furthermore, because dark-grown soybean seedlings have a relatively large cell size, water potentials of individual cells could be measured directly and compared with theoretically predicted values. We found that predicted and measured values coincided well in the cortical cells. Deviation of a water potential field in the pith region between theoretically predicted and measured values is probably caused by the assumption that all cells in the elongation zone had the same water permeability, i.e., K in Eq. 6, when the field was calculated theoretically. Because the measured cell water potentials in the pith region were lower than the theoretically predicted values, cells in the pith region must have lower water permeability than the value used in the present study (i.e., the value of K which was taken from the work of Molz and Boyer (1978)).

The xylem water potential cannot be measured with the combination of a pressure probe and a nanoliter osmometer, because the xylem does not have confined cell membranes. Thus, actual xylem potential in Fig. 2 should be represented by values shown in thin lines in the xylem region (Xy in Fig. 2), whose values were measured directly with psychrometry and used as the boundary conditions in calculation. Thus, it is evident that the steepest water potential gradient existed surrounding the xylem in the growth-induced water potential field when growth was occurring actively (0h in Fig. 2).

When a growth-induced water potential field is considered in plants under water stress, physiological phenomena observed under stress can be explained logically. Nonami and Boyer (1989a) observed a decrease in turgor of cells located near the xylem immediately after water stress, but no decrease in turgor of cells located near the stem surface. Because their experimental setups were similar to our theoretical analysis in the soybean stem, both results can be compared. When the water stress was applied to the root region of soybean seedlings, water potentials of the xylem and its surrounding cells decreased but not in cells located near the stem surface as shown in Fig. 2A (see the line marked with 5h). Because accumulation of solute in cells does not seem to occur drastically within such short intervals, osmotic potential of cells can be considered to be constant. Thus, when the water potential in the root region was altered, the initial changes in water potentials of cells in the elongation zone of the stem must have coincided with changes in cell turgor. Because a drop in water potential of cells located near the xylem was about 0.2 MPa within 5h after stress according to Fig. 2A, a decrease in turgor of those cells must have been about 0.2 MPa. Nonami and Boyer (1989a) observed about 0.2 MPa of decrease in turgor of cells located near the xylem, supporting our theoretical analysis.

Although we have restricted our calculations and measurements of a growth-induced water potential field to the zone of elongation of dark-grown soybean seedlings, the same concept must be applicable to any growing tissue. Because the expanding cells must absorb water from the water source in order to increase their volume, the growing tissue must form a growth-induced water potential field. For practical purposes, the magnitude of a growth-induced water potential can be estimated from water potential differences between the elongating tissue and the water source. When the growth-induced water potential was measured by using hydroponically grown kidney bean seedlings, it was found that growth of the seedlings was correlated with changes in the growth-induced water potential. Furthermore, growth of kidney bean embryos was inhibited by salt stress, cold stress and heat stress, and it was also found that growth of the embryos was correlated with changes in the growth-induced water potential.

Nonami and Boyer (1990a) studied the kinetics of physiological parameters regulating plant growth under water stress by using the recently developed guillotine psychrometer. They found that the growth-induced water potential was the first of the physiological parameters to decrease to a growth-limiting level (Nonami and Boyer, 1990a). The decreased water potential gradient within the growing tissue was caused by a decreased water potential of the xylem. This was followed by a similar decrease in cell wall extensibility and tissue conductance for water (Nonami and Boyer, 1990a).

In order to check the observations which indicate that the inhibition of stem growth at low water potentials is accompanied by decreases in cell wall extensibility and tissue hydraulic conductance to water, Nonami and Boyer (1990b) developed a theory to relate rheological properties of cell walls measured with an extensiometer and thermodynamical properties of cell walls measured with psychrometry. They found that the plastic properties of the cell walls determined cell wall extensibility (Nonami and Boyer, 1990b).

Both cell wall extensibility and hydraulic conductance are under metabolic control, and their low levels under water stress indicate that substantial metabolic change occur in the enlarging cells (Nonami and Boyer, 1990a). Wall extensibility is probably determined by the action of wall enzymes on the polymerization of wall constituents and the cross-linking of structural proteins with wall constituents (Cassab and Varner, 1988; Fry, 1986; Theologis, 1986). In seedlings similar to those used here, polyribosome and mRNA levels decreased a few hours after water stress (Mason et al., 1988). Increased amounts of proteins were extractable from the walls and a 28 kD protein present in the cytoplasm accumulated in the walls (Bozarth et al., 1987).

Growth recovery occurred when the xylem water potential was equilibrated with the medium (Fig. 1). This was caused by osmotic adjustment in cells (Nonami and Boyer, 1989a, 1990a). Prior to growth recovery, recovery of growth-induced water potential was observed (Figs. 2 and 4, Nonami and Boyer, 1990a). The recovery of growth-induced water potential induced metabolic changes in cells, resulting in increases in wall extensibility and hydraulic conductance (Nonami and Boyer, 1990a, 1990b). Afterward, growth was resumed. Thus, growth-induced water potential can be considered to be a primary physiological factor regulating growth under water stress.

Because growth conditions of these kidney bean seedlings simulate those of most crops grown under controlled environments and because the estimated growth-induced water potential agreed well with the growth of kidney bean seedlings, it is evident that the concept of growth-induced water potential can be applicable for the regulation of crop growth cultivated in plant factories and greenhouses.

X. CONCLUSION

The enlarging plant tissue displayed significant water potential disequilibria between the water source and the enlarging cells. This disequilibria formed a water potential field (i.e., potential field of chemical potential of water) in the enlarging tissue, which is called a growth-induced water potential field. Theoretically, the growth-induced water potential field can be expressed by an equation derived from a combination of Fick's law and the conservation of mass, and can be used to determine the direction of water flow in the expanding tissues. Experimentally, cell water potentials of cells in the zone of elongation were measured with combinations of a cell pressure probe and a nanoliter osmometer. Both measured and theoretically predicted water potentials coincided well, showing that the growth-induced water potential field existed in the elongating tissue. The average magnitude of the growth-induced water potential field was estimated from the water potential difference between the water source and the expanding cells by using psychrometry. Prior to growth recovery by acclimation to environmental stress by plants, the growth-induced water potential recovered. Thus, we found that if the growth-induced water potential was measured, both growth recovery and inhibition could be predicted. Because changes in water status in plants occur prior to changes in cell expansion, if the system for water status measurements is established in plant factories and greenhouses, it may be possible to regulate crop growth more effectively.

Practically speaking, the magnitude of a growth-induced water potential field can be estimated from differences between water potentials of the elongating region and the water source. Because the growth-induced water potential field recovers prior to growth recovery, if the growth-induced water potential field can be determined accurately, growth behavior of crops can be predicted under controlled environments.

REFERENCES

Alfrey, T. (1948) Mechanical Behavior of High Polymers. Interscience Publishers, Inc., New York.

Boyer, J.S. (1967) Leaf water potentials measured with a pressure chamber. Plant Physiol., 42, 133–137.

Boyer, J.S., Cavalieri, A.J. and Schulze, E.-D. (1985) Control of the rate of cell enlargement: Excision, wall relaxation, and growth-induced water potentials. Planta 163:527–543.

Boyer, J.S., and E.B. Knipling (1965) Isopiestic technique for measuring leaf water potentials with a thermocouple psychrometer. Proc. Natl Acad. Sci., 54, 1044–1051.

Bozarth, C.S., J.E. Mullet, and J.S. Boyer (1987) Cell wall proteins at low water potentials. Plant Physiol., 85, 261–267.

Cassab, G.I., and J.E. Varner (1988) Cell wall proteins. Annu.

Rev. Plant Physiol. Plant Mol. Biol., 39, 321–353.

Cleland, R. (1959) Effect of osmotic concentration on auxin–action and on irreversible and reversible expansion of the *Avena* coleoptile. Physiol. Plant. 12:809–825.

Feynman, R.P., Leighton, R.B. and Sands, M. (1964) Lectures on Physics. Volume II. Addison–Wesley Publishing Company, Reading, Massachusetts.

Fry, S.C. (1986) Cross–linking of matrix polymers in the growing cell walls of angiosperms. Annu. Rev. Plant Physiol., 37, 165–186.

Green, P.B., and Cummins, W.R. (1974) Growth rate and turgor pressure. Auxin effect studied with an automated apparatus for single coleoptiles. Plant Physiol. 54:863–869.

Green, P.B., Erickson, R.O. and Buggy, J. (1971) Metabolic and physical control of cell elongation rate. *In vivo* studies in *Nitella*. Plant Physiol. 47: 423–430.

Kramer, P.J. (1983) Water Relations of Plants. Academic Press, New York.

Mason, H.S., J.E. Mullet, and J.S. Boyer (1988) Polysomes, messenger RNA and growth in soybean stems during development and water deficit. Plant Physiol., 86, 725–733.

Matthews, M.A., Van Volkenburgh, E. and Boyer, J.S. (1984) Acclimation of leaf growth to low water potentials in sunflower. Plant Cell Environ. :199–206.

Matyssek, R., Maruyama, S. and Boyer, J.S. (1988) Rapid wall relaxation in elongating tissues. Plant Physiol. 86: 1163–1167.

Molz, F.J. (1976a) Water transport through plant tissue: the apoplasm and symplasm pathways. J. Theor. Biol., 59, 277–292.

Molz, F.J. (1976b) Water transport in the soil–root system: transient analysis. Water Resource Res., 12, 805–808.

Molz, F.J., and J.S. Boyer (1978) Growth–induced water potentials in plant cells and tissues. Plant Physiol., 62, 423–429.

Molz, F.J., and E. Ikenberry (1974) Water transport through plant cells and cell walls: theoretical development. Soil Sci. Soc. Am. J., 38, 699–704.

Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol. Plant. 15: 473–497.

Nonami, H., and J.S. Boyer (1989a) Turgor and growth at low water potentials. Plant Physiol., 89, 798–804.

Nonami, H., and Boyer, J.S. (1989b) Water potential gradients associated with expanding cells. In the Proceedings of the Yamada Conference XXII, Osaka, Japan, "Plant Water Relations and Growth under Stress," pp428–430.

Nonami, H., and J.S. Boyer (1990a) Primary events regulating stem growth at low water potentials. Plant Physiol., 93, 1601–1609.

Nonami, H., and J.S. Boyer (1990b) Wall extensibility and cell hydraulic conductivity decrease in enlarging stem tissues at low water potentials. Plant Physiol., 93, 1610–1619.

Nonami, H., J.S. Boyer, and E. Steudle (1987) Pressure probe and isopiestic psychrometer measure similar turgor. Plant Physiol., 83, 592–595.

Nonami, H., and E.-D. Schulze (1989) Cell water potential, osmotic potential, and turgor in the epidermis and mesophyll of transpiring leaves: Combined measurements with the cell pressure probe and nanoliter osmometer. Planta, 177, 35–46.

Theologis, A. (1986) Rapid gene regulation by auxin. Annu. Rev. Plant Physiol., 37, 407–438.

Wainwright, S.A., Biggs, W.D., Currey, J.D. and Gosline, J.M. (1976) Mechanical Design in Organisms. Edward Arnold (Publishers) Limited, London, pp. 423.

INDEX

A

absorption 98, 101, 203, 269, 273

absorption flux 269

adaptive control 4, 8, 30, 36–39, 41, 46, 178

air temperature 26, 53, 58, 69, 70, 97, 104, 146, 201, 203, 205, 268, 273

artificial intelligence (AI) 146, 178, 180, 182, 188, 191, 197–199, 206, 207, 212, 215, 220, 221, 227

artificial neural network (ANN) 186

automation 75, 77, 96, 111, 112, 177–179, 188, 189, 191, 199, 227, 228

autoregressive moving-average (ARMA) 184, 186

B

biotechnical process 75–77, 79, 81, 83, 92

blue light 246, 253

C

cell wall 244, 249, 256, 257, 270, 288, 291, 294, 295, 306, 307, 313, 315, 317, 319, 321, 327

chlorophyllous tissue 106

clustering of variables 210

CO_2 enrichment 110, 111, 123, 124, 159, 170, 232

cold stress 322, 325, 327

compartmentation 249, 256, 258

computer aided control system design (CACSD) 4, 6, 21

computer integrated manufacture (CIM) 188, 191

computer integrated system 175, 177, 188–191

computer support system 187

computer technology 225, 235, 237
computerization 111
constraint 126, 128, 130, 131, 133–135, 168, 199–201, 205–214, 216, 217, 219–222, 232, 236
constraint satisfaction problem (CSP) 199, 206–208, 214, 216, 217
control algorithm 15, 20, 30, 69, 90, 92, 139, 140, 146, 177, 178
crop model 128, 144, 154, 155, 168, 170, 229, 230
culture medium 97, 323–325
culture vessel 96, 99–104, 106, 110
cuticle layer 272, 279, 281, 282

D

decision process 199–201, 204, 205
delta (δ) operator 4, 6–9, 11, 12, 19, 20, 24, 41, 42, 45
direct digital control (DDC) 3, 5
discrete differential operator 7
distributed–parameter system 80

E

elastic deformation 314
embryo 98, 323–325, 327
empirical knowledge 119, 197–200, 205
environmental control 25, 95, 96, 98, 112, 140, 144, 146, 148, 168, 176– 179, 182, 188, 191
environmental signal 243, 244, 252, 258
enzyme 244, 245, 247–249, 251–258, 260–262, 294, 328
ethylene 104, 111
expert system 145, 146, 148, 180, 188, 220, 230, 233, 236, 237
explant 96, 97, 106, 111
extensibility 291–295, 316, 321, 327, 328
extensiometer 294, 315–317, 319, 322, 327

F

factory automation (FA) 177, 179, 188, 189, 191
filtering 182, 207, 210, 212–216, 218

flower 51, 126, 202, 203, 244, 258, 260–262
fuzzy control 178

G

gas exchange 101, 110, 273, 274, 279
gene 245, 246, 249, 251, 254, 258–262
genetic regulation 243, 258
graphical user interface (GUI) 21
gravity 64, 245, 251
greenhouse automation 178, 179
greenhouse environment 95
greenhouse production system 197, 198
group method of data handling (GMDH) 185, 186
growth flux 269
growth inhibition 294, 295, 304, 310, 312, 321, 325
growth–development 198
growth–induced water potential 270, 285, 288–290, 292–296, 307, 309–313, 322–329
growth–induced water potential field 295, 296, 307, 309–313, 325–329

H

Hamiltonian 131, 132, 156, 161, 172
heat stress 322, 325, 327
heating 25–27, 29, 30, 53, 65, 66, 92, 104, 119, 124, 155, 159, 161, 164, 169, 198, 201, 202, 204–206, 215, 216, 233
heterotrophy 106
horticultural operation 176, 187–189, 191
hydration 250, 251, 269, 272, 306, 307
hydration flux 269
hydraulic conductance 292, 294, 295, 327, 328

I

identifiability of parameter 90
image analysis 226–229, 235
information content of measurement 80, 82

information matrix 82, 83
information technology 232, 233, 236

K
knowledge processing 180–182, 189, 191
knowledge representation 206, 208, 216, 221
knowledge–based system 199, 200, 219, 220, 228

L
local area network (LAN) 142, 189–191

M
mass water budget 268, 285, 296
maximum principle 153, 156
meristem 260, 262
micropropagation 95–99, 107, 110–112, 178, 179, 191, 228
model identification 26
model predictive control (MPC) 154
mushroom harvesting 229

N
nanoliter osmometer 272, 274, 309, 326, 329
net photosynthetic rate (NPR) 106–108, 110
neural network 154, 182, 186, 187, 228
neural–net 133, 134, 165
node 98, 110, 156, 165, 166, 172, 207, 228
non–minimum state–space (NMSS) 4, 14, 15, 18–20, 26, 27, 39, 40, 45, 46
nutrient film technique (NFT) 4, 25, 33–40, 45, 139

O
object–oriented 199, 216, 220
objective 5, 13, 15, 21, 26, 120, 140, 198, 199, 204, 205, 216, 219, 221, 222, 296
observer 15, 20, 77, 79

operator 19, 20, 24, 26, 41, 42, 45, 46, 184, 189, 226, 228, 253

optimal control 5, 23, 54, 79, 121, 139, 140, 143, 146, 154, 157, 168, 175, 178, 231, 232

optimization 52, 69, 71, 79, 81, 82, 130–132, 134, 140, 143–146, 148, 154, 165, 168, 170, 179, 189, 191, 198, 200

osmometer 272, 274, 309, 326, 329

P

parameter estimation 8, 36, 38, 145, 148, 178

parameter identification 75, 79, 90, 92

performance criterion 155, 156, 159, 167, 169, 170, 172

pH 76, 176, 178, 245, 248, 249, 255, 256, 261

photomixotrophy 106

photoperiod 104–106, 110

photosynthesis 54, 56, 96–98, 106, 107, 123, 126, 128, 129, 132, 133, 135, 144, 153, 180–182, 202–205, 229, 246, 247, 249, 250, 267, 268, 270, 290

photosynthetic photon flux density (PPFD) 97, 99, 104–111

photosynthetically active radiation (PAR) 56

phytochrome 245, 246

PID algorithm 177, 178

plant factory 175–177, 179, 180, 191

plant factory automation 179

plant response 148, 179–182

plantlets 96–104, 106–112, 324, 325

plastic deformability 314, 316, 319, 321

plastic deformation 314–317

Poisson's ratio 315, 321

Pontryagin 153

practical knowledge 200

predictive control 79, 91, 92, 154

pressure probe 272, 274, 294, 309, 315, 326, 329

preventive management decision 202

price 76, 102, 122, 124, 130, 133, 134, 140, 143, 153, 155–157, 159, 161, 163–166, 168–170, 203, 232, 233

process automation (PA) 77, 188, 191, 227

process control 176, 177, 179, 180, 188, 189
proportional-integral-derivative (PID) 3, 5, 45, 146, 177, 178, 188
proportional-integral-plus (PIP) 4, 5, 14, 16, 19-21, 23-26, 28-33, 37, 39, 40-42, 45, 46, 146
protein 79, 245-261, 294, 328
psychrometer 274, 295, 316, 321, 323, 327

Q

quantitative approach 198, 200

R

reasoning process 198, 201, 204, 206, 208
receptor 243, 252-254, 258, 262
recursive least squares (RLS) 11, 39, 40
red light 99, 245, 253
relative humidity 97, 101-103, 128, 129, 132, 177, 179
relaxation 11, 291, 319
resolution mechanism 206-208, 213, 214
robot 177, 191, 228, 229
robot control 228
robotization 96, 111, 112

S

salt stress 322, 325, 327
saturation deficit 201-203, 205
second messenger 253-255
selective field operation 233
self-adaptive control 4, 8, 30, 37-39, 41, 46
self-tuning 4, 8, 25, 36-41, 46
sensor allocation 87, 89
SERRISTE 198-200, 205-208, 212, 213, 215-217, 219-223
set-point 126, 132, 133, 177, 188
setpoint 25, 140, 143, 154-159, 161-167, 169, 170, 175, 176, 197-202, 205-209, 212, 215, 218, 220, 221, 233
simplified refined instrumental variable (SRIV) 8-11, 42, 46

simulation 4, 33, 36, 37, 42, 97, 145, 156, 159, 167, 198, 200, 229, 230
sink term 306
soil temperature 201, 203, 205, 218
solar radiation 55, 56, 67, 129, 146, 156–159, 162, 165, 201–205, 209, 216, 219
speaking plant 144, 179, 180, 182, 189, 231
speaking plant approach (SPA) 179, 180, 183, 188, 189
stages of growth 204
state variable feedback (SVF) 4, 5, 15, 17, 18, 20, 25, 26, 30, 40
state-variable 4, 5, 12, 15, 20, 25, 54, 68, 69, 129, 131–134, 153–157, 161, 162, 166, 168–170, 172, 230
stomata 62, 108, 144, 181, 203, 268, 271, 283
stomatal aperture 231, 274–279, 281, 283, 285
storage 66, 95, 125, 129, 168, 170, 257, 269, 272
support system 127, 187, 189, 191, 220
system identification 79, 178, 182

T

tactical management 198
tissue culture 95–97, 178, 179, 323–325
tomato 33, 126, 130, 153, 154–159, 161, 163–166, 169, 188, 198, 200, 201, 203, 204, 208, 215, 216, 222
tower reactor 83–85
trajectory 144, 154, 155, 162, 166, 169, 170
transfer function (TF) 4, 6, 8, 14, 15, 45
transpiration 52, 54, 56, 61, 62, 67–69, 97, 126, 128, 129, 144, 181, 202–204, 267–269, 271–277, 279–283, 285, 290, 296
transpiration flux 269
true digital control (TDC) 3–5, 6, 9, 21–24, 36, 39, 42–45
turgor 128, 246, 270, 272, 279, 280, 283, 285, 289–291, 293, 307, 309, 310, 315, 316, 319, 321, 324–328

V

vapor pressure difference (VPD) 272–281, 283, 285,
variable 30, 32, 37, 39, 46, 54, 64, 78, 82, 97, 128, 129, 131–134, 144, 145, 147, 153–157, 161, 162, 166, 168–170, 175, 182, 186,

199, 201, 202, 204–220, 230, 231, 233, 296
ventilation 25, 52, 53, 59–61, 63, 68, 69, 104, 108, 110, 111, 129, 155, 156, 159, 232, 233
vessel closure 101, 104

W

wall extensibility 292–295, 316, 321
water potential 102, 126, 181, 250, 269–276, 279–281, 283, 285–296, 304–313, 319–329
water status 101,107,181,250, 271–274, 279, 281–283, 285, 304, 308, 323, 329
water stress 126, 181, 231, 291, 292, 294, 295, 304, 309–313, 319, 321,323, 326–328
weather 53, 71, 119, 122, 127, 131–135, 143, 148, 153, 155–159, 162, 163, 165–170, 198–200, 203–205, 215, 217, 219, 220, 232, 233, 236, 247

Z

z^{-1} operator 5–9, 14, 19, 20, 26, 37, 42, 45