

What can Systems and Control Theory do for Agricultural Science?

UDK 681.5:631
IFAC 2.0:5.1

Original scientific paper

While many professionals with a background in agricultural and bio-resource sciences work with models, only few have been exposed to systems and control theory. The purpose of this paper is to elucidate a selection of methods from systems theory that can be beneficial to quantitative agricultural science.

The state space representation of a dynamical system is the corner stone in the mainstream of systems theory. It is not well known in agro-modelling that linearization followed by evaluation of eigenvalues and eigenvectors of the system matrix is useful to obtain dominant time constants and dominant directions in state space, and offers opportunities for science-based model reduction. The continuous state space description is also useful in deriving truly equivalent discrete time models, and clearly shows that parameters obtained with discrete models must be interpreted with care when transferred to another model code environment. Sensitivity analysis of dynamic models reveals that sensitivity is time and input dependent. Identifiability and sensitivity are essential notions in the design of informative experiments, and the idea of persistent excitation, leading to dynamic experiments rather than the usual static experiments can be very beneficial.

A special branch of systems theory is control theory. Obviously, control plays an important part in agricultural and bio-systems engineering, but it is argued that also agronomists can profit from notions from the world of control, even if practical control options are restricted to alleviating growth limiting conditions, rather than true crop control. The most important is the idea of reducing uncertainty via feed-back.

On the other hand, the systems and control community is challenged to do more to address the problems of real life, such as spatial variability, measurement delays, lacking data, environmental stochasticity, parameter variability, unavoidable model uncertainty, discrete phenomena, variable system structures, the interaction of technical and living systems, and, indeed, the study of the functioning of life itself.

Key words: systems theory, agronomics, state space modelling, linearization, sensitivity analysis, uncertainty, control theory

1 INTRODUCTION

Webster's New International Dictionary describes a system as »an aggregation or assemblage of objects joined in regular interaction or interdependence; a set of units combined by nature or art to form an integral organic or organized whole; an orderly working totality ...«. So, we may say that a system is characterized by individual parts, units or things, which work together as an entity, a whole, to serve a certain purpose. According to Wikipedia (<http://en.wikipedia.org>), systems theory goes back to the biologist von Bertalanffy, who in the 1930s noticed striking similarities in the functioning of systems from various scientific disciplines. Systems theory as a discipline emerged in the years directly after World War II. It aims at providing a generalized framework by which one can analyze and describe any group of objects that

work in concert to produce some result. It is based on the isomorphy of concepts, laws and models (<http://iss.org>).

A significant branch in systems theory deals with the study of dynamical systems, i.e. systems where relevant variables are varying in time. In fact, static systems can be seen as a special case. Within the domain of systems theory, control theory forms a subset that deals with the science to make a system behave in a specified way.

Although – interestingly enough – the history of system theory originates from biology, its materialization in recent years as a discipline with a strong mathematical orientation seems to have enlarged the gap with the main stream of activities in the life sciences, to which agricultural and food science belongs. A brief literature survey on the combination of selected elements from systems theory

Table 1 Cross sections of systems notions with agriculture and horticulture in literature in Scopus™ over the years 1996–2007

| Search term | Agriculture or agronomy | Horticulture | Remarks |
|-----------------------|-------------------------|--------------|-----------------------------|
| Notion | | | |
| Modelling | 12 575 | 546 | |
| Simulation | 10 788 | 476 | |
| System Theory | 57 | 4 | |
| System Dynamics | 119 | 5 | |
| Linearisation | 48 | 0 | |
| Similarity Transform | 0 | 0 | |
| Identifiability | 10 | 1 | almost all related to water |
| Parameter Estimation | 856 | 34 | |
| System Identification | 24 | 4 | |
| Sensitivity Analysis | 1 471 | 66 | 87 reviews |
| Uncertainty Analysis | 260 | 4 | |
| Control Theory | 51 | 5 | all engineering |
| Control System | 555 | 40 | |

NB. Control as search term gives almost 60 000 hits in conjunction with agriculture; this is due to the colloquial use of the term related to manipulation and experimentation.

and agricultural sciences reveals that there is little interaction between these fields. Table 1 summarizes the number of papers found with the Scopus™ system (<http://www.scopus.com>) using the designated areas indicated in the table. The notions in the left column appearing in title, abstract or keywords were combined with agriculture or agronomy, and with horticulture as search words for all fields in the publication records, leading to the figures indicated in the table. Manual inspection of the references revealed that specific systems and control terminology entries, such as »system identification«, were in majority related to engineering, and of these many were dealing with control engineering, which, of course, has found some application. It appears that in the non-engineering agricultural science area little use is made of systems and control theory, despite the wide use of modelling and simulation. Among the most applied techniques are sensitivity analysis and parameter estimation. These topics have some basis in statistics that has always played a major role in agriculture, but even then it is not sure that elements from dynamic systems theory are fully exploited.

The purpose of this paper is two-fold. The first part briefly summarizes important results from systems and control theory, with the idea that these

results might be relevant to researchers and modellers in the agricultural sciences. The second part discusses potential reasons for the lack of penetration of systems theory, thus providing a mirror to the systems and control community, which hopefully will help to bridge the gap. A similar paper about control theory and management of ecosystems has triggered the idea for this paper [8].

2 APPLYING SYSTEMS THEORY

Whereas the advancement of systems theory hinges on the further development of generalised concepts, it is the application of it to a specific discipline that makes it of interest to that discipline. In the application of systems theory to a specific discipline, roughly the following activities can be distinguished:

- a) Modelling
- b) Analysis
- c) Synthesis.

2.1 Modelling

A model is a simplified representation of reality, that encapsulates the significant aspects of the real system for the intended purpose. In the behaviour-

al view of Willems [11] a (mathematical) model is seen as a pair (U, B) , where U is a set representing the universum, and B a subset of U representing the possible outcomes, called the behaviour. In this way, a model is seen as a mathematical formalism, often in the form of mathematical equations, to restrict the space of possible outcomes. Models are obtained either by first principles modelling or by system identification methods. In the first principle approach models are set up by assembling concepts, conservation principles and »laws« known from previous experience. The outcome of this process is a model structure. Next, parameter estimation is used to derive parameters that were not known in advance, by comparison with observation data. The system identification approach starts from available data, and tries to deduce both structure and parameters from these data. The set of possible structures can be limited by prior knowledge.

Once a model is available, it must be put to the test by checking its performance on a set of independent data, not used before. This process is known as model validation. In the philosophy of Popper the task of science is to steadily try to falsify the model, hence detecting shortcomings and lacks in knowledge. In engineering, a model that has not (yet) been falsified is useful, as long as the application remains within the validity domain. In passing, it should be noticed that the assessment of the validity domain is not trivial. The least one could say is that the new application should not have new structural elements – thus violating the system boundaries –, and that the amplitude and frequency spectrum of the new inputs are comparable to those for which the model was originally built.

2.2 Analysis

In the analysis stage the purpose is to derive general properties of the system by analysing properties of its model. This encompasses, among other things, stability analysis, sensitivity analysis, time scale analysis, response analysis, and notions like observability, controllability, reachability, not discussed further here. In the analysis stage the full arsenal of linear system theory is available, which can also be applied locally to non-linear system models by linearization, whereas non-linear methods are currently emerging.

2.3 Synthesis

In the synthesis stage, the systems model is used for the intended purpose. This can pertain to the

design of new or improved systems, to the management and handling of a system, to remedy certain unwanted properties of the system, or to design a control system to force a particular desirable system behaviour. In the latter case, the system plus its controller constitutes a system again, and is thus amenable to analysis and synthesis using the same tools indicated above. However, we now have control over certain parts of the system by design, which allows us to seek for the most suitable solutions or to solutions that are optimal in some sense. Hence, system optimisation is another component of systems theory, and, indeed, dynamic optimisation is one of the most powerful techniques that deserves to find wide application.

In the sequel, I will briefly discuss some results from systems theory that I found to be particularly useful, and therefore, in my opinion, deserve to be more widely known in the food and agricultural scientific community. The discussion is mainly dealing with dynamical models, i.e. models that describe behaviour as functions of time.

3 A SELECTION OF METHODS AND TOOLS

3.1 State Space

In systems and control theory the dominant modelling paradigm is the celebrated state space approach. In this approach a clear distinction is made between signals that act as independent forces (inputs $u(t)$), signals that are of interest and can be observed or calculated (outputs $y(t)$) and variables that represent the state of the system $x(t)$, such that knowledge of the current state plus information of future inputs is enough to be able to compute the future evolution of the output. The states thus encompass the memory of what happened in the past, so that having the current state no information about the past is needed any more. Mathematically, a continuous time state space model is represented by

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t), p) \\ y(t) &= g(x(t), u(t), p)\end{aligned}\quad (1)$$

where the dimensions of the vectors u , x and y are n_u , n_x and n_y , respectively, and the vector p is a n_p dimensional constant parameter vector. In a number of publications the parameter p is allowed to be time varying as well. We prefer to see such parameters as just a set of additional input signals. The function f represents n_x equations to describe the rate of change of x , and the function g is a n_y

dimensional function, sometimes referred to as read-out function.

A discrete time model representation in state space form is given by

$$\begin{aligned} x(t_{k+1}) &= F(x(t_k), u(t_k), p) \\ y(t_{k+1}) &= G(x(t_{k+1}), u(t_{k+1}), p) \end{aligned} \quad (2)$$

where t_k is a sample instant, and $t_{k+1} = t_k + T$ with T the sampling interval.

3.2 Time Scales

An important characteristic of a dynamical system are the time constants of the system. The time constants are related to the response of the system states to a disturbance, as explained below.

Linearization. In general, system (1) is non-linear. The state equation (1a) is called linear if $\partial f/\partial x$ and $\partial f/\partial u$ are not a function of x and u ; if not, the system is non-linear. A non-linear system can be converted into a linear system locally by linearizing around the actual trajectory point $(x^o(t), u^o(t))$. This is done by Taylor series expansion of f :

$$\begin{aligned} f &= f^o + \left. \frac{\partial f}{\partial x} \right|_{x^o, u^o} (x - x^o) + \\ &+ \left. \frac{\partial f}{\partial u} \right|_{x^o, u^o} (u - u^o) + \text{h.o.t.} \end{aligned} \quad (3)$$

Here, $\partial f/\partial x = A$ is a $n_x \times n_x$ Jacobian matrix, with on row i the vector $[\partial f_i/\partial x_1, \partial f_i/\partial x_2, \dots, \partial f_i/\partial x_{n_x}]$, and similarly for the $n_x \times n_u$ dimensional matrix $\partial f/\partial u = B$. The abbreviation h.o.t. encompasses higher order terms that are ignored, which is allowed in regions close to the linearization point. Defining the deviation variables $\tilde{x} = x - x^o$ and $\tilde{u} = u - u^o$, and substituting (3) into (1a) leads to the linearized state equation

$$\dot{\tilde{x}}(t) = A^o(p)\tilde{x}(t) + B^o(p)\tilde{u}(t) \quad (4)$$

The superscript o reminds us to the fact that the matrices A and B depend upon the local point; it is shorthand for $A(x^o(t), u^o(t), p)$. Also note that the elements of these matrices depend upon the original parameters, and are therefore mutually correlated.

As this notation is a bit awkward, in the sequel we simply write

$$\dot{x} = Ax + Bu \quad (5)$$

and remember the special meaning of x , u , and A , B . In particular, note that if the original inputs and state are on their nominal values, the new u and x are zero.

Similarity transform and time constants. Now we look at the evolvment of x (the deviated state) after a small perturbation of the state away from zero, in the absence of input deviations, i.e the solution of the autonomous system

$$\dot{x} = Ax \quad x(0) = x_o \quad (6)$$

Intuitively, this should give information about the »speed« of the components of the system. An eigenvalue of the matrix A is a scalar λ such that

$$Av = v\lambda \quad (7)$$

The vector v for which this holds is called an eigenvector of A . There are n_x eigenvalues and associated eigenvectors. If we form a diagonal matrix Λ with the eigenvalues on the diagonals and zero's elsewhere, and a matrix V where each column is an eigenvector of A , then we can write

$$AV = V\Lambda \rightarrow A = V\Lambda V^{-1} \quad (8)$$

provided that the inverse of V exists. (In the case of non-distinct eigenvalues, V is singular. In that case one has to resort to the more generic singular value decomposition). The crucial role of the eigenvalues and eigenvectors in characterising the dynamics can be seen by introducing new states by linear combination of the old states according to

$$z = V^{-1}x \quad (9)$$

Substituting (8) into (6) gives

$$\dot{x} = V\Lambda V^{-1}x \rightarrow V^{-1}\dot{x} = \Lambda V^{-1}x \rightarrow \dot{z} = \Lambda z \quad (10)$$

and we can see that the individual elements of the transformed state z behave independently of each other, as solutions of (10) have the form

$$z_i(t) = z_{i,o} \exp(\lambda_i t) \quad (11)$$

It is clear that the »speed« of the response depends upon the eigenvalues, which underlines the important role of the eigenvalues. If the eigenvalues are complex numbers, there are oscillations in

the response. In addition, we can see that system returns to the zero state only if the real part of the all eigenvalues are strictly negative. The time constants of the system are defined as

$$\tau_i = \frac{1}{|\lambda_i|} \tag{12}$$

Eigenvalues and eigenvectors are easily obtained by mathematical packages such as Matlab, and several simulation packages, e.g. SIMULINK, have facilities to obtain the linearized model automatically.

Relevance to practice. What is the relevance of all this for the modeller of agricultural systems? We briefly list them here:

- 1 Knowing the time constants helps to determine the necessary detail of the input signals. In particular, if the system is slow, it does not make much sense to use high frequent input data.
- 2 Knowing the various time constants helps to design sampling programs for model calibration, i.e the decision of the required sample interval. This can save a lot of money.
- 3 If the time constants are wide apart, the system is called »stiff« and numerical simulation becomes slow.
- 4 If the system is stiff, it may be desirable to reduce the equations by assuming that the fast parts are reaching equilibrium in infinitely short time. These parts then become algebraic equations, and the order of the system is reduced. If the algebraic equations can be solved explicitly, simulation of a reduced model is faster than for the full model.
- 5 Knowing how the eigenvalues evolve along the trajectory of a non-linear model gives insight in possible instability regions. This is particular relevant in systems with inherent or constructed feed-back.

Example

Consider the greenhouse system with no heating input, given by

$$\begin{aligned} \dot{T}_g &= \frac{1}{C_g} (U_{mg} (T_m - T_g) - U_{ge} (T_g - T_e)) \\ \dot{T}_m &= \frac{1}{C_m} (-U_{mg} (T_m - T_g) + I) \end{aligned}$$

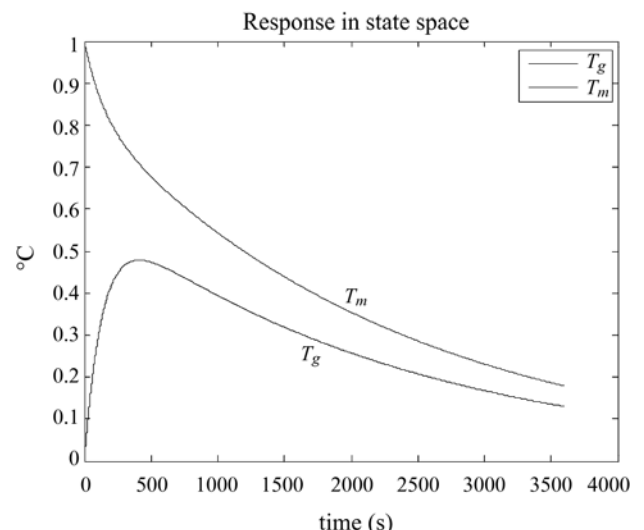
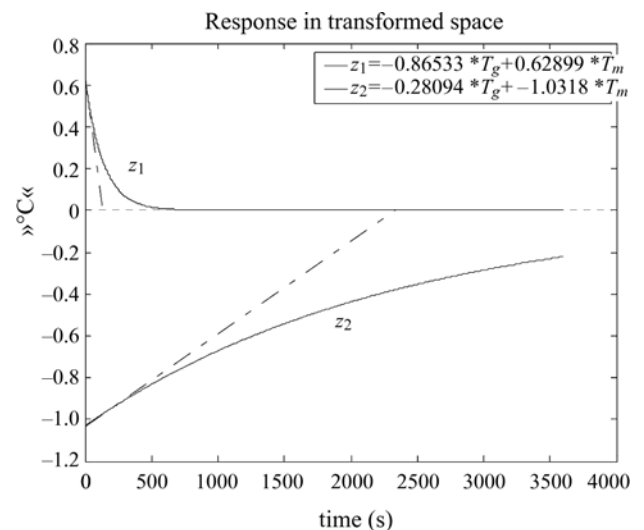
where the states T_g , T_m are the greenhouse air temperature and solid materials temperature, respectively, the inputs T_e and I are the outside tempera-

ture and the solar radiation, respectively. The heat capacities of the greenhouse air is $C_g = 6000 \text{ J m}^{-2} \text{ K}^{-1}$, of the solid materials $C_m = 16000 \text{ J m}^{-2} \text{ K}^{-1}$, and the overall heat transfer coefficient between air and environment is $U_{ge} = 12 \text{ W m}^{-2} \text{ K}^{-1}$, and between air and solid materials $U_{mg} = 25 \text{ W m}^{-2} \text{ K}^{-1}$. With steady nominal inputs $I_{ss} = 100 \text{ W m}^{-2}$ and $T_{e,ss} = 10 \text{ }^\circ\text{C}$ the steady state temperatures are $T_{g,ss} = 18.3 \text{ }^\circ\text{C}$, $T_{m,ss} = 22.3 \text{ }^\circ\text{C}$.

The eigenvalues of the greenhouse system matrix are easily found to be $-0,0073$ and $-0,0004 \text{ (s}^{-1}\text{)}$ so that the time constants are 137 and 2336 s. The associated (decoupled) transformed variables are

$$\begin{aligned} z_1 &= -0.86533 \cdot T_g + 0.62829 \cdot T_m, \text{ and} \\ z_2 &= -0.28094 \cdot T_g + -1.0318 \cdot T_m \end{aligned}$$

We see that z_2 is largely dominated by T_m , and its dynamics is therefore representative for the dy-



namics of T_m . The behaviour of the other state is more complicated and will show clear interaction with T_g . The top figure shows the decoupled behaviour of z , expressed as deviations from steady state, after a disturbance of the solid material temperature of 1 °C. The bottom figure, showing the original states, supports the statements above.

This illustrates among other things:

- the system is stiff, thus a stiff solver is the best choice for simulations
- in order to accurately know the behaviour of the greenhouse temperature, the sampling interval must not be longer than a minute or so,
- at longer time scales the greenhouse air temperature may be considered to be in pseudo-equilibrium, i.e. immediately coupled to the solid material temperature. In such reduced model, the sampling interval for calibration can be much longer (half an hour for instance)

* * *

3.3 Discrete Time Modelling

In agronomics, it is very customary to use discrete time models to describe, for instance, field crop growth. Such models typically produce output on a daily basis. Discrete time models have an intuitive attractive appeal, as they simply state that what we have tomorrow is what we have today plus the gain (positive or negative) of today. As no integration is involved, computer code also runs fast. The use of discrete time models for continuous systems is, however, not without certain risks if the underlying principles are only remotely understood.

Let us consider, for the sake of the argument, again the linear dynamic system (5). Starting from initial condition $x(0)=x_0$, the solution in time is given, formally, by

$$x(t) = e^{A(t-t_0)}x(t_0) + \int_{\tau=t_0}^{\tau=t} e^{A(t-\tau)}Bu(\tau)d\tau \quad (13)$$

The matrix e^{At} is called the state transition matrix, and can be obtained by evaluating the series

$$e^{At} = 1 + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \dots \quad (14)$$

From (13) it can be seen that the solution consists of an autonomous part and a part that depends upon the input signal over the interval from t_0 to t .

Equation (13) can be used to derive the equivalent linear discrete time model. Let $t_{k+1}=t_k+T$, where T is the sample interval, and define $x_k=x(t_k)$, we can then write

$$x_{k+1} = x_k + (e^{AT} - 1)x_k + \int_{\tau=t_k}^{\tau=t_k+T} e^{A(t-\tau)}Bu(\tau)d\tau \quad (15)$$

Compare this with the popular Euler approximation of the solution of (5), being

$$x_{k+1} = x_k + ATx_k + BTu_k \quad (16)$$

and it is immediately clear that the two solutions are not identical. We see that in order to get rid of the integral in (15) to arrive anywhere near (16), an assumption is required about the inter-sample behaviour of the input. Working this out in the scalar case $\dot{x}=ax+bu$, with u piece-wise constant, i.e. $u(t)=u_k t_k \leq t < t_{k+1}$, we obtain from (15) the exact discrete time equivalent model for this input

$$x_{k+1} = x_k + (e^{aT} - 1)x_k + bu_k \left(\frac{1 - \exp(-aT)}{a} \right) \quad (17)$$

to which the Euler approximation (16) still is not equivalent, although the solutions become close if.

Of course, all this is well known, and it is the very reason why more advanced numerical integration methods than Euler have been developed and are widely available in simulation packages. Also, when the system is linear, the conversion from continuous time to discrete time and vice-versa is easily done, e.g. in Matlab, provided the user specifies the inter-sample behaviour by selecting from a set of standard approximations. The point is, that in agronomical practice (and, indeed, in many other fields as well) the discrete time models are usually not set up starting from the continuous model, but rather by directly formulating it as a discrete model. This implies that parameters obtained in the discrete time setting, being dependent upon the chosen sample interval, may be biased when obtained from experiments with another sampling time, and similarly will depend upon the trustworthiness of the assumed inter-sample behaviour. Sure enough, in the non-linear case, the situation may be even more tricky.

Example

Let lettuce growth be modeled by the linear exponential growth model $dW/dt=k_gW$, where W is the plant biomass expressed in g dry weight. As-

sume $k_g = 0.09 \text{ d}^{-1}$. Let measured values be $W = 1 \text{ g[dw]}$ on day 20 and $W = 1.094 \text{ g[dw]}$ on day 21. Then, according to the Euler approximation $k_g \Delta t \approx \Delta W / W$ one finds $k_g = 0.094 \text{ d}^{-1}$, which is about 4% higher than the true value. Taking a longer interval with the same method makes things worse: $W(23) = 1.310 \text{ g[dw]}$ and $k_g^{\text{estim}} = 0.103 \text{ d}^{-1}$, i.e. 14% too high. Unthoughtful use of the continuous growth rate estimate in a discrete time model of the type $W(t+1) = W(t) + k_g W(t) \Delta t$ leads to $W(21) = 1.09$, which is 0.004 units too low. This error does not occur if in the discrete time model the discrete time parameter estimate is used. In that case, it is essential to stick to the chosen time interval.

* * *

3.4 Sensitivity Analysis

Sensitivity analysis is useful in various stages of modelling. Sensitivity to model parameters is used to detect to which parameters the model outcome is most sensitive. Once found, these parameters are candidates for calibration or further experimentation. Knowing the sensitivity to input signals is often also useful, to determine the required accuracy of the input sequence, and to guide management, design and control.

A popular and seemingly straight forward way of performing a parameter sensitivity analysis is by varying the parameters and checking the results in simulation. This technique can be awkward when the number of parameters is large. A more elegant method is local sensitivity analysis, using the model equations [15].

Let in the system description (1) p^0 be a nominal parameter vector, then the sensitivity function of the associated nominal state trajectory to variations in parameters is defined by

$$s_{ij}(t) = \frac{\partial x_i(t; p^0)}{\partial p_j} \quad \begin{array}{l} i = 1, \dots, n_x \\ j = 1, \dots, n_p \end{array} \quad (18)$$

Note that s_{ij} is a function of time and, in general, has physical units. By differentiating (1) with respect to each parameter, and interchanging the sequence of differentiation, the following linear sensitivity system can be derived

$$\dot{S}(t) = A(t; p^0)S(t) + M(t; p^0) \quad (19)$$

where

$$S(t) = \begin{pmatrix} s_{11} & \cdots & s_{1n_p} \\ \vdots & \ddots & \vdots \\ s_{n_x 1} & \cdots & s_{n_x n_p} \end{pmatrix} \quad (20)$$

is the matrix of sensitivity functions, and

$$A(t; p^0) = \left. \frac{\partial f(x, u, p)}{\partial x} \right|_{x=x(t; p^0)} \quad (21)$$

$$M(t; p^0) = \left. \frac{\partial f(x, u, p)}{\partial p} \right|_{x=x(t; p^0)}$$

The matrix A is the Jacobian matrix already encountered in (4). The derivation of (19) is possible under the condition that f is continuous in x and p , and has continuous partial derivatives with respect to x and p on the interval $[t_0, t]$. The parameter sensitivity at $t=0$ is the zero matrix.

The interesting feature is that the sensitivity system can be solved in parallel to the original model, and will then yield $n_x \times n_p$ (local) sensitivity trajectories at once. These patterns can be used not only to detect sensitive parameters, but also for experimental design by selecting observation time points where sensitivity is large.

Examples of application to a lettuce model are given in Van Henten and Van Straten [16] and Van Straten et al. [17].

It has to be said that the method above is local in nature, and only holds for small parameter perturbations. It can therefore provide a first impression of the sensitivities, which can then be used to probe further using the numerical method mentioned in the beginning of this section, but now for the sensitive parameters only. The local method also relies on the existence of the differentials in Eqn. (21). Often, non-differentiability can be avoided by rephrasing the model. In crop growth models, for instance, the temperature dependency is often expressed as a trapezoidal function of the temperature, leaving the derivative undefined at the vertices, whereas it is just as easy – and probably more close to reality as well – to use a continuous function.

The method can be expanded to models with a mix of ordinary differential equations and algebraic equations Caracotsios and Stewart [3] This is relevant as there are many instances where variables are related to each other via very fast equilibrium reactions. The problem is tackled by casting it in the descriptor form

$$E\dot{x}(t) = f(x(t), u(t), p) \tag{22}$$

where E is a $n_x \times n_x$ matrix of constants, often, but not always, represented by

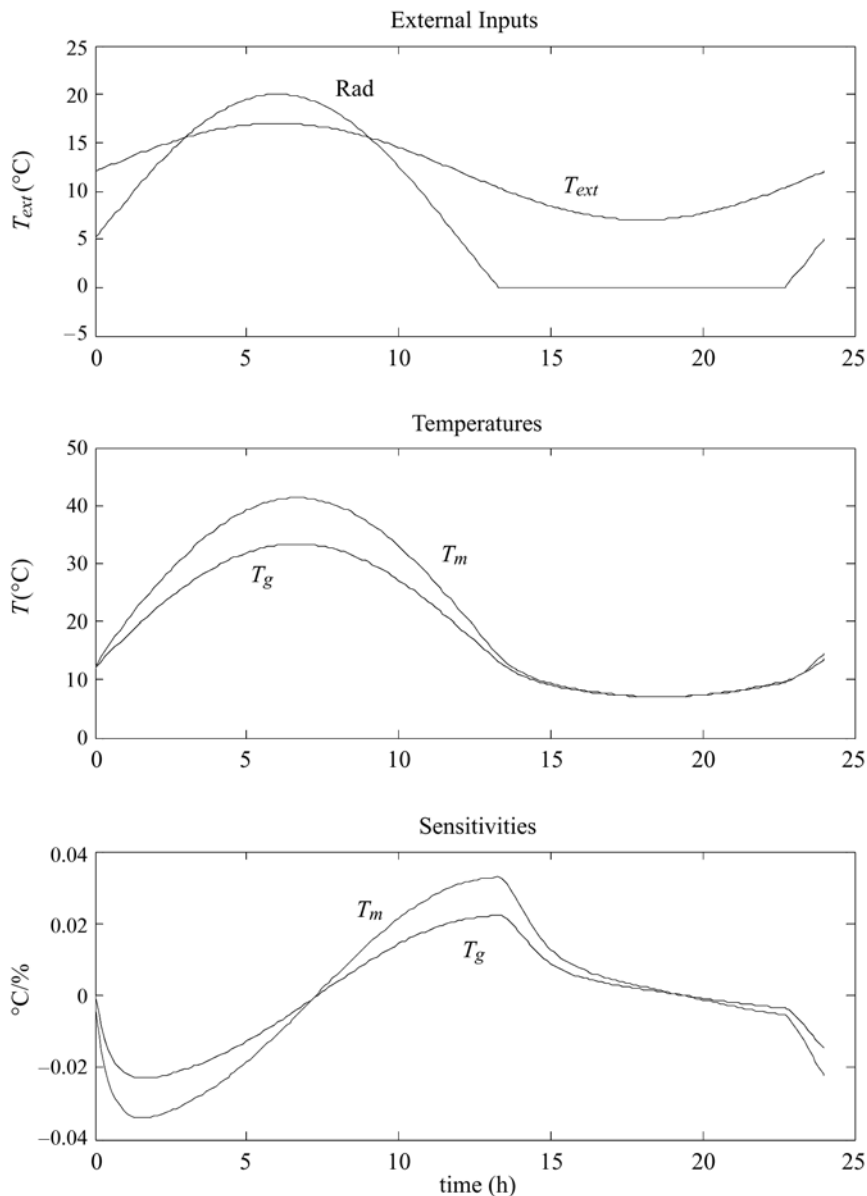
$$E = \begin{pmatrix} I_s & | & 0 \\ \hline 0 & | & 0 \end{pmatrix} \tag{23}$$

with I_s the $s \times s$ identity matrix. If $s < n_x$ there are algebraic relations in the system, which constitutes a challenge to standard simulation packages, if they have no explicit solutions.

In passing we notice that (22) can be cast in the standard state space form only if E is invertible, which is not always the case. This is a rare example of an instance where the standard state space description is not applicable.

Example

The method above was applied to the greenhouse example. The top figure shows the external inputs to the greenhouse (temperature and half-sine wave radiation), the middle figure the resulting greenhouse temperatures, and the bottom figure the sensitivities of the temperatures to a one percent change in the heat capacity parameters of the solid



materials (C_m). The most informative data points for calibration of the model are a few hours after sunrise and a few hours before sunset, while mid-day and night temperatures are non-informative. It appears that for the estimation of the transfer coefficients midday is the most informative period (not shown).

* * *

3.5 Identification, Parameter Estimation and Experimental Design

Calibration of dynamical models is usually done by adjusting the parameters so as to minimise the difference between model output and observed outputs in some sense. Usually, the outputs are observed at specific sampling instants, leading to a set of measurements $y_{obs}(t_k)$, or simply $y_{obs,k}$, $k=1, \dots, N$. A popular criterion to be minimised is the quadratic objective function

$$J(p) = \sum_{k=1}^N (y_k(p) - y_{obs,k})^T Q_k (y_k(p) - y_{obs,k}) \quad (24)$$

where $y_k(p)$ represents the model output vector at t_k , and the $n_y \times n_y$ matrix Q is a matrix that allows the weighting of multiple outputs (with possibly different units) against each other. It may be time dependent. A convenient weighting is to use the reciprocal of the standard error of the observations, while it is usually assumed that the observations are mutually not correlated, which sets the off-diagonal elements equal to zero. In the single output the criterion is just the sum of squares.

Usually, numerical procedures are used to find the minimum. Some methods use the gradients of J to the parameters p which involves once again, among other things, the calculation of the Jacobian matrix already encountered in the linearization and the sensitivity analysis. The relation with sensitivity analysis is by no means accidental. Loosely speaking, a steep gradient of the objective function towards the parameters points in the direction of the most sensitive parameters, and this is the natural direction to go on the way to the bottom of the valley, whereas, conversely, the objective function surface is rather flat in the direction of insensitive parameters, meaning that the system output hardly changes when those parameters are changed.

In practice, search methods have difficulty in finding the minimum if there are many insensitive parameters. Therefore, it makes sense to remove

the insensitive parameters from the calibration set. Moreover, parameter estimates are often correlated, meaning that simultaneous change of two parameters hardly results in changes in objective function. In the case of two parameters, one can draw contour lines with equal sum of squares, which in the neighbourhood of the minimum take the form of ellipses. Correlation between parameter estimates is indicated when the contour ellipse becomes oblong.

The curvature of the sum of squares surface is related to the Fisher information matrix, which in the sampled data case is given by

$$F = \sum_{k=1}^N \left(\frac{\partial y_k(p)}{\partial p} \right)^T Q_k \left(\frac{\partial y_k(p)}{\partial p} \right) \quad (25)$$

An estimate of the co-variance of the parameter estimates is given by the approximate relation

$$\text{cov}(\hat{p}) \approx \frac{J(\hat{p})}{N - n_p} F^{-1} \quad (26)$$

The axes of the ellipses are in the direction of the eigenvectors of F , and the length is proportional to the inverse of the square root of the associated eigenvalues. The ratio of the largest and smallest eigenvalue is the condition number, and from the above it should now be clear that when this ratio is large, there are strong correlations in the estimation.

This knowledge is applied in slightly modified form by Ioslovich et al. [6] to select parameters that are candidates for calibration in the case of a nitrate in lettuce model. It can also be used to design experiments, where the purpose is to make the ellipses more round by suitable choice of experimental conditions [19].

Note that the Fisher information matrix can be computed along with the sensitivity system. An idea developed by Stigter and Keesman [14] is to derive a feed-back law intended to choose input time trajectories such that the parameter sensitivity and hence the ability to estimate the parameters from data is maximised.

Example

The procedure above was applied to the greenhouse example. It is assumed that only measurement data of the greenhouse air temperature are available, with a fixed sampling interval of 10 s. Calculating the inverse of the Fisher information

matrix, and expressing this in the form of a correlation matrix, i.e.

$$c(i, j) = F^{-1}(i, j) / \sqrt{F^{-1}(i, i)F^{-1}(j, j)}$$

| | | | | |
|----------|-------|-------|----------|----------|
| | C_g | C_m | U_{ge} | U_{mg} |
| C_g | 1 | -0.97 | 0.97 | 0.80 |
| C_m | -0.97 | 1 | -1 | -0.62 |
| U_{ge} | 0.97 | -1 | 1 | 0.62 |
| U_{mg} | 0.80 | -0.62 | 0.62 | 1 |

The strong (negative) correlation between the estimates indicates that independent estimation of all 4 parameters is almost impossible. The inverse of the Fisher information matrix is equal to the covariance of the estimates when the standard error of the measurements is 1 °C, and will be quadratically lower with lower measurement error. They also become lower when the sampling interval is made shorter.

The coefficients of variation of the estimates are 69%, 14%, 28%, 25%, respectively. It shows that C_g can be determined less well than C_m from greenhouse temperature measurements alone. Note that this depends upon the dynamics in the input signal (at steady state C_g and C_m cannot be determined at all).

The condition number of the Fisher matrix (ratio of highest to lowest eigenvalues) is about 10^{11} , which is another indication that estimation with these input of all four parameters from just air temperature is not feasible.

* * *

3.6 Control and Optimization

This paper is not the proper place to open up the whole wealth of methods available in control theory. These are obviously of significant interest to engineers. The principles of feed-back to counteract the effects of uncertainty are, however, of general scope, and are certainly often used in management of live stock and open field farming as well. So, a basic understanding of the effects of the choice of measurement variables, sampling, time-delay, feed-back gain on performance and stability of the managed or controlled system is not just important for engineers alone. In fact, in biology, the feed-back methods of living systems are wonderful examples on how nature has solved control problems, and it is quite conceivable that the study of these ingenious mechanisms can also form a source of inspiration for engineers.

A very powerful methodology to achieve relevant controllers and management strategies lies in dynamic optimisation and optimal control. The key point of dynamic optimisation is to find control input trajectories or strategies that generate a system behaviour that is optimal according to a user defined goal function, possibly subject to user defined or physical constraints. Some relevant references here are Bryson [2] and Stengel [13].

3.7 Data driven models

Most of what was written before seemed to be relevant to system models represented by differential or difference equations in state space derived from physical principles. However, if necessary such models can be converted to higher order differential equations in input-output form, which in the single-input – single input (SISO) case reads

$$f\left(\frac{\partial^n y}{\partial t^n}, \frac{\partial^{n-1} y}{\partial t^{n-1}}, \dots, \frac{\partial y}{\partial t}, y\right) = g\left(\frac{\partial^m u}{\partial t^m}, \frac{\partial^{m-1} u}{\partial t^{m-1}}, \dots, \frac{\partial u}{\partial t}, u\right) \quad (27)$$

although there can be intricacies related to the differentiability not discussed here (see [11] for more details and precise mathematical definitions of the delineation of inputs and outputs in the domain of manifest variables).

A special case of these are the linear discrete time autoregressive moving average models (ARMA) of the form (in the SISO case)

$$y_{k+1} = a_0 y_k + a_1 y_{k-1} + \dots + a_n y_{k-n} + b_0 u_k + b_1 u_{k-1} + \dots + b_m u_{k-m} \quad (28)$$

This form or similar is obtained when data driven methods are used to arrive at a systems model, e.g. with the system identification toolbox of Matlab.

It is straight forward to find a state space realisation of such an model, and if found, all the methods above apply. (We are not talking about the problem of finding state space representations in special forms, such as controller canonical or observer canonical forms; this is the topic of realisation theory). In passing it should be noted that the non-linear version of (28) can be written as

$$y_{k+1} = N(y_k, y_{k-1}, \dots, y_{k-n}, u_{k+1}, u_{k-1}, \dots, u_{k-m}) \quad (29)$$

with N a non-linear function mapping independent inputs to dependent outputs. An artificial neural network could be chosen for N . Hence, systems theory methods can be applied equally well to this class of data driven models [10, 4, 12].

4 UNCERTAINTY

In the development above, the suggestion was that the model is an exact representation of the system. It is, however, clear that in reality there will always be uncertainties. Let us discuss the following sources of uncertainty:

- a) Uncertainty in the model parameters.
- b) Uncertainty in the model structure.
- c) Uncertain input sequences, e.g. the weather.

Uncertainty in model parameters arises due to measurement errors in the (output and input) data used to develop the model; this may also pertain to parameter values published in the literature. In the procedure to estimate parameters from data usually an idea about the co-variance of the estimates is obtained. The space enclosed in hyper-contours of equal sum of squares can be viewed as a confidence region, but it is not known what confidence percentage should be adhered to those in the non-linear-in-the-parameters case [5, 9]. In model prediction, Monte Carlo analysis can be used to map the parameter uncertainty to the space of model outcomes, i.e. we do not just get a single trajectory, but rather a family of trajectories.

An alternative to tackle parameter uncertainty comes in mind when it appears that the sequence of residual errors between data and model is correlated with itself or with the input data [7, 4]. If this happens it is tempting to try to model this by considering the parameters to be time-varying. In fact, in that case a data-based sub-model is added to the original model. Recursive estimation techniques can be used to generate time dependent parameter trajectories. In the prediction mode, this trajectories can then be modelled as a stochastic process, and subsequently been simulated using Monte Carlo simulation. This is quite different from the constant but uncertain parameter case before, as the model now essentially becomes a stochastic model, which, in theory, requires special techniques for simulation and analysis.

Time varying parameters, in my view, are in fact an example of structural errors, and instead of data-based modelling, an attempt can be made to clarify underlying structures in an iterative model building process. This, however, can be time consum-

ing, but is, on the long run, the most rewarding step because it provides insight that can be used again and again, and hence contributes to the advancement of science. Another way of tackling these structural uncertainties is to have various models, and calculate possible future behaviours by all models in parallel. This seems to be the dominant approach in global warming studies.

Uncertain future input sequences can be tackled if stochastic distributions are known, by using Monte Carlo studies. Scenario studies arise when the uncertainty in future inputs is related to uncertainty in the development of society due to policies or autonomous phenomena.

5 BARRIERS TO ADOPTION

The above, though just a selection from the plethora of methods in systems and control, has hopefully shown the potential. What, then, could stop its wider dissemination? In my view, the reasons can be grouped into two main categories: (i) ignorance and lack of education, (ii) specific fundamental difficulties associated to systems in the field of food and agriculture.

As to the first point, it is clear that the mathematical skills needed to apply systems and control theory are not normally taught in agro-biology oriented curricula. However, this discrepancy can be remedied by special courses on the level of PhD students, as has been shown in our experience with the two-weeks PhD course »The Art of Modelling« over the past years in Wageningen. Moreover, bio-informatics, theoretical biology, agricultural and bio-systems engineering, hydrology, and theoretical ecology hinge largely on advanced mathematical skills, which may have a positive effect upon the more traditional agricultural sciences. Moreover, the increasing availability of simulation and analysis tools lowers the threshold, and there is, in fact, really no excuse not to use them.

Issues in the second group are:

- a) non-linearity. Most real life systems are non-linear
- b) switching structures and time varying system dimensions. An example is the event that a plant suddenly starts to form fruits
- c) time varying structures due to adaptation and self-organisation
- d) spatial heterogeneity
- e) wide range of scales in time and space; the issue of interconnectivity, tearing and zooming, as

well as emerging properties of complex networks of individual units.

All of these problems are being addressed in the systems and control community, but have not had wide enough appeal to practitioners so far. In the modelling area, catastrophe theory analysis of attractors is quite well developed. In the control community there are several developments in non-linear control theory using Lie algebra concepts, currently mainly of interest to specialists. Switching and time varying structures are being studied in hybrid systems theory, e.g. [1]. Unlike in the linear world, there are, however, very few widely applicable tools available.

Models for systems with a spatial dimension can be forced in the state space form by spatial discretization, although direct segmentation of the space into compartments does not necessarily lead to efficient or accurate numerical algorithms. In contrast, finite element methods are more suitable, but they are not in state space form. There is a tremendous development in what has become known as Computational Fluid Dynamics (CFD), but the link with classical systems theory seems to be missing, and the straight forward use of CFD model for feedback control is rare.

Adaptation, self-organisation and emerging properties have given rise to what is called the »holistic approach«, in contrast to the »reductionistic approach« of classical physics. However, in the end, the holistic approach leads to models of interlinked systems as well, that can be studied very well by systems theory methods. It is interesting to note that by assuming that a plant or micro-organism has some internal objective, optimal control theory can be applied to explain emerging behaviour. The model then contains switches or other heuristic mechanisms that are manipulated via the internal control law, thus avoiding the need for precise mechanistic modelling of such internal processes. An example is Velten and Richter [18].

Interestingly enough, on the cellular and molecular level one currently sees the emerging area of what has become known as »systems biology«, which in its ideal appearance would just be a merger of systems theory, molecular biology and computer science, and as such could be an example for other bio-systems areas of research.

6 CONCLUSION

In this paper a number of useful techniques from systems theory have been presented. Overall, these

techniques are well known to engineers. The point we were trying to make is that in our view it is essential that also non engineers, in fact everybody, who is working with models in agronomy and other agricultural, horticultural or even ecological fields, should be informed about the existence of these methods, and should have been exposed to them at some stage of his professional training.

At the same time it is clear that some special properties of agricultural systems have only marginally been addressed by the systems and control community. Much work is going on, but currently there is a lack of practical translation of findings from non-linear systems theory, hybrid systems theory, non-linear control and optimisation, in a form that is suitable for the practitioner. There is a nice task ahead of us.

REFERENCES

- [1] M. S. Branicky, V. S. Borkar, and S. K. Mitter, **A unified framework for hybrid control: Model and optimal control theory**. IEEE Transactions on Automatic Control 43 (1):31–45, 1998.
- [2] A. E. Bryson, **Dynamic Optimization**: Addison-Wesley, 1999.
- [3] Caracotsios, Makis, and Warren E. Stewart. **Sensitivity analysis of initial value problems with mixed odes and algebraic equations**. Computers & Chemical Engineering 9 (4):359–365, 1985.
- [4] S. S. Chen, **Non-linear system identification using neural networks**. International journal of control 51 (6):1191–1214, 1990.
- [5] N. R. Draper and H. Smith. **Applied Regression Analysis**: Wiley, New York, 1966.
- [6] Ilya Ioslovich, Per-Olof Gutman and Ido Seginer: **Dominant parameter selection in the marginally identifiable case**. Mathematics and Computers in Simulation 65 (1–2):127–136, 2004.
- [7] L. Ljung: **System Identification – Theory for the User**. Prentice-Hall. 1987.
- [8] C. Loehle: **Control theory and the management of ecosystems**. Journal of Applied Ecology 43:957–966, 2006.
- [9] D. C. Montgomery, A. E. Peck, and G. G. Vining: **Introduction to Linear Regression Analysis**. Wiley, New York, 2001.
- [10] K. Narendra and K. Parthasarathy: **Identification and control of dynamical systems using neural networks**. IEEE Transactions on Neural networks 1(1): 4–27, 1990.
- [11] J. W. Polderman and J. C. Willems: **Introduction to Mathematical Systems Theory – A Behavioural Approach**. Springer-Verlag, New York, 1998.

- [12] J. Sjöberg and L. Ljung: **Overtraining, regularization and searching for a minimum, with application to neural networks.** International Journal of Control 62 (6):1391–1407, 1995.
- [13] R. F. Stengel: **Optimal Control and Estimation.** Dover Edition ed: Dover Publ, Mineola, NY, USA. 1994.
- [14] J. D. Stigter and K. J. Keesman: **Optimal parametric sensitivity control of a fed-batch reactor.** Automatica 40 (8):1459–1464, 2004.
- [15] R. Tomovic and M. Vukobratovic: **General Sensitivity Theory.** American Elsevier, New York. 1972.
- [16] E. J. van Henten and G. van Straten: **Sensitivity analysis of a dynamic growth model of lettuce.** Journal of Agricultural Engineering Research 59:19–31, 1994.
- [17] G. van Straten, I. Lopez Cruz, I. Seginer, and F. Buwalda: **Calibration and sensitivity analysis of a dynamic model for control of nitrate in lettuce.** Acta Horticulturae 507:149–156, 1999.
- [18] K. Velten, and O. Richter: **Optimal root/shoot-partitioning of carbohydrates in plants.** Bulletin of Mathematical Biology 57 (1):99–107, 1995.
- [19] Karina J. Versyck, Kristel Bernaerts, Annemie H. Geeraerd and Jan F. Van Impe: **Introducing optimal experimental design in predictive modeling: A motivating example.** International Journal of Food Microbiology 51 (1):39–51, 1999.

Čime mogu teorije sustava i automatskog vođenja pomoći agronomskoj znanosti. Iako mnogi profesionalci s područja agrotehnike i bio-znanosti rade s modelima, tek je nešto malo onih koji su izloženi uporabi teorije sustava i automatskog vođenja. Namjena je ovog rada raščišćavanje izbora metoda s područja teorije sustava koji mogu biti od koristi kvantitativnoj agronomskoj znanosti.

Predstavljanje dinamičkih sustava prostorom stanja temeljno je usmjerenje teorije sustava. U agro-modelima je slabo poznato da je linearizacija, koju slijedi evaluacija svojstvenih vrijednosti i svojstvenih vektora matrice sustava, korisna pri dobivanju dominantnih vremenskih konstanti i dominantnih usmjeravanja u prostoru stanja te pri nudi mogućnosti znanstveno zasnovanog pojednostavljenja modela. Opis kontinuiranog prostora stanja također je koristan u dobivanju stvarno ekvivalentnih diskretno-vremenskih modela te jasno pokazuje da se dobiveni parametri diskretnih modela moraju pažljivo interpretirati kada se prenose u okoliš s drugim kodnim modelom. Analiza osjetljivosti dinamičkih sustava otkriva da je osjetljivost ovisna o vremenu i ulazima. Mogućnost identifikacije i osjetljivost bitni su pojmovi u projektiranju informativnih pokusa, a ideja trajne pobude koja vodi dinamičkim pokusima a ne uobičajenim statičkim pokusima može biti vrlo korisna.

Posebna grana sustavske teorije jest teorija automatskog vođenja. Očigledno, vođenje igra važnu ulogu u agronomskom i bio-sustavskom inženjerstvu, ali se naglašava da agronomi mogu profitirati iz pojmova svijeta automatike, čak i ako je automatizacijska inačica ograničena na olakšavanje uvjeta ograničenja rasta, a ne i pravog vođenja usjeva. Najvažnija je ideja da se uporabom povratne veze smanji neodređenost.

S druge strane zajednica je sustavskih inženjera i automatičara izazvana da se više posveti stvarnim problemima života poput prostorne promjenjivosti, mjernih kašnjenja, nedostataka podataka, stohastičnosti okoliša, varijabilnosti parametara, nezaobilazne neodređenosti modela, fenomena diskretizacije, varijabilnosti sustavskih struktura, međudjelovanja tehničkih i živih sustava, i napose studiju funkcioniranja života kao takvog.

Ključne riječi: teorija sustava, agronomija, modeli prostora stanja, linearizacija, analiza osjetljivosti, teorija vođenja

AUTHORS' ADDRESSES

Gerrit van Straten

E-mail: gerrit.vanstraten@wur.nl

Wageningen University, Systems and Control,
Wageningen, The Netherlands

Received: 2007-12-10

Accepted: 2008-09-10