

Data-Based Mechanistic Modeling, Forecasting, and Control

By Peter Young and Arun Chotai

In the environmental and agricultural sciences, as in many areas of engineering and science, mathematical models are normally formulated as deterministic differential (or equivalent discrete-time) equations. Most often, the structure of these equations is defined by the scientist or engineer in a form that is perceived to be appropriate according to the physical nature of the system and the current scientific paradigms in the area of study. Most of the “simulation” models that emerge from this approach are very large, with many unknown parameters. Consequently, they are difficult, if not impossible, to identify, estimate, and validate in rigorous statistical terms because of problems associated with over-parametrization and the lack of experimental or monitored data. Unlike the situation in engineering, however, natural environmental systems, and many systems in agriculture, are not normally

Young (p.young@lancaster.ac.uk) and Chotai are with the Centre for Research on Environmental Systems and Statistics, Institute of Environmental and Natural Sciences, Lancaster University, Lancaster LA1 4YQ, U.K.

©1991 2181 CENTURY MEDIA

manmade. As a result, their internal physical, biological, and ecological mechanisms are often poorly understood, and planned experiments that might lead to improvements in such understanding are either difficult or, in the case of environmental systems, even impossible to undertake.

Increasingly in recent years, however, the poorly defined nature of many environmental and agricultural systems has been recognized. It is now becoming accepted that a new modeling philosophy and associated methodology is required that acknowledges the need to quantify the uncertainty associated with the model (e.g., [1]-[3]). The Centre for Research on Environmental Systems and Statistics (CRES) at Lancaster University has been prominent in this area of research. This article outlines the major aspects of our approach to stochastic modeling, as well as briefly describing three studies that demonstrate the practical utility of this approach. In particular, the article shows how parametrically efficient (parsimonious), low-order stochastic models can be produced that reflect the dominant modal characteristics of the system and can be interpreted in physically meaningful terms. Within the present context, the most important aspect of these data-based mechanistic (DBM) models is that they provide an appropriate basis for advanced nonstationary or nonlinear signal processing, adaptive forecasting, and multivariable control system design.

Unlike data-based “black-box” models, DBM models are required to have a mechanistic interpretation. Indeed, they are not deemed truly credible in scientific terms unless such an interpretation proves possible. But, unlike the situation with “grey-box” models, such a physically meaningful explanation is not imposed as a hypothesis prior to modeling. Rather, it emerges from an inductive modeling procedure only after the model structure has first been identified parsimoniously from a generic class of models with wide application potential. In this manner, prior scientific prejudice about the nature and complexity of the model is avoided, and the resulting DBM model will normally have a structure and parametrization that is appropriate to the information content of the data. In this article, the generic class of models used in DBM modeling comprises stochastic linear and nonlinear differential or difference equations, or their transfer function equivalents.

DBM modeling methods were developed originally for the analysis of measured time-series data, particularly in relation to the design of signal processing, forecasting, and automatic control systems. In the early stages of modeling, when observational data are either scarce or not available, however, the same methodological tools can also be used to obtain reduced-order versions of the large simulation models. The importance of reduced-order models that reflect the dominant modes of system behavior is being widely recognized in many areas of application within the environmental and agricultural sciences. For example, such models have been used in the simplification of large simulation models used in climate research (e.g., [4]) and in optimal

Table of Abbreviations

AMV	Active mixing volume
AR	Autoregressive (model)
ARC	Adaptive radar calibration (system)
R_r^2	Coefficient of determination based on simulation response error
CRES	Centre for Research on Environmental Systems and Statistics
CV	Control volume
DBM	Data-based mechanistic
DMA	Dominant mode analysis
FIS	Fixed-interval smoothing
GPC	Generalized predictive control
GSA	Generalized sensitivity analysis
HMC	Hybrid-metric-conceptual (model)
IPCC	International Panel on Climate Change
KF	Kalman filter
LQ	Linear quadratic
LQG	Linear quadratic Gaussian
MFD	Matrix fraction description
MCS	Monte Carlo simulation
NMSS	Nonminimal state space
PDF	Probability distribution function
PIP	Proportional-integral-plus (controller)
RLS	Recursive least-squares (method)
RIV	Refined instrumental variable (method)
SRI	Silsoe Research Institute
SDP	State-dependent parameter (model)
SVF	State variable feedback (control)
SRIV	Simplified refined instrumental variable (method)
TF	Transfer function (model)
TVP	Time-variable parameter (model)
UC	Unobserved component (model)

control system design (e.g., [5]). An example of the DBM approach to model reduction is the modeling and control of the microclimate in a large horticultural greenhouse [6] where the reduced-order control model is obtained from a much larger nonlinear simulation model. A similar DBM approach has been used to obtain linear, reduced-order versions of the topically important “global carbon cycle” models [6], [7] used in the studies of the International Panel on Climate Change (IPCC).

This article outlines the main aspects of DBM modeling, as well as the associated methods of signal processing, adaptive forecasting, and multivariable control that exploit such models. This overall generic approach appears to have wide application potential, and its practical utility is illus-

trated by three practical examples. These range from the model reduction and control application mentioned above through the modeling and control of forced ventilation systems in agricultural buildings to the design of adaptive flood forecasting and warning systems.

Methodological Background

The methods used in this article have been developed over the past 20 years, and they cover four main areas: DBM methods for modeling linear and nonlinear stochastic systems; dominant mode analysis (DMA) leading to the combined linearization and simplification of large simulation models; unobserved component (UC) models and their use in the recursive estimation, forecasting, and smoothing of nonstationary time series; and nonminimal state space (NMSS) methods of control system design. All of these methods have been described in detail elsewhere, so they are outlined only briefly in the following subsections.

Data-Based Mechanistic Modeling

Previous publications (e.g., [6]-[10]) illustrate the evolution of the DBM philosophy and its methodological underpinning. This methodological basis is heavily dependent on the exploitation of recursive estimation in all its forms. These include the traditional application to online and offline parameter estimation using recursive least-squares (RLS) and optimal refined instrumental variable (RIV) algorithms; the design of adaptive forecasting systems using the Kalman filter (KF) algorithm; the use of recursive fixed-interval smoothing (FIS) for more statistically efficient, offline time-variable parameter (TVP) estimation for use in signal extraction; and the exploitation of FIS combined with other procedures for state-dependent parameter (SDP) estimation and the nonparametric/parametric identification of nonlinear stochastic systems.

The three major phases in the DBM modeling strategy are as follows:

- *Stochastic Simulation Modeling and Sensitivity Analysis* (e.g., [2], [3], [6], [8], and the references therein). In the initial phases of modeling, observational data may well be scarce, so any major modeling effort will have to be centered on simulation modeling, usually based initially on largely deterministic concepts such as dynamic mass and energy conservation. Recognizing the inherent uncertainty in many environmental and agricultural systems, however, these deterministic simulation equations can be converted into an alternative stochastic form. Here, it is assumed that the associated parameters and inputs can be represented in some suitable stochastic form such as a probability distribution function (PDF) for the parameters and time-series models for the inputs. The subsequent stochastic analysis then exploits Monte Carlo simulation (MCS) in three ways: first, to explore the propagation of uncertainty in the resulting stochastic model;

second, as a mechanism for generalized sensitivity analysis (GSA) to identify the most important parameters leading to a specified model behavior; and third, the use of MCS in stochastic optimization.

- *Dominant Mode Analysis and Simulation Model Simplification* (e.g., [6], [7], and the references therein). The initial exploration of the simulation model in stochastic terms can reveal the relative importance of different parts of the model in explaining the dominant behavioral mechanisms. This understanding of the model is further enhanced by employing a novel method of combined statistical linearization and model order reduction. This is applied to time-series data obtained from planned experimentation, not on the system itself, but on the simulation model that becomes a surrogate for the real system. Such DMA is exploited to develop low-order, dominant mode approximations of the simulation model; approximations that are often able to explain its dynamic response characteristics to a remarkably accurate degree (e.g., coefficients of determination > 0.999 ; i.e., greater than 99.9% of the high-order model output is explained by the low-order model). Conveniently, the statistical methods used for such linearization and order reduction exercises are the same as those used for the next phase in the modeling process.
- *DBM Modeling from Real Data* (see previous references). Time-series data form the basis for stochastic DBM modeling. The DBM models characterize those dominant modes of the system behavior that are clearly identifiable from the time-series data, and unlike simulation models, the efficacy of the DBM models is heavily dependent on the quality of these data. The models derived in this form normally constitute the main vehicles for the subsequent design of forecasting and automatic control systems. In the case of constant-parameter models, the methodological tools used in this final stage of DBM analysis are linear, constant parameter, transfer function (TF) identification and estimation tools, such as optimal instrumental variable methods (SRIV/RIV; e.g., [11] and the references therein). These can be compared with the alternative and better known methods, such as those in the MATLAB Identification Toolbox, which are designed specifically for backward shift operator models and, in most cases, require concurrent estimation of noise model parameters (note that the RIV/SRIV algorithms are different from the IV/IV4 options in the Identification Toolbox). In contrast, RIV/SRIV algorithms are available for the estimation of TF models in *all* the major operators (backward shift, time derivative, and delta) and do not require concurrent estimation of a noise model. If the system under investigation is nonstationary, however, then the constant-parameter estimation is replaced by TVP esti-

mation [12]-[14], whereas if the system is heavily nonlinear, the SDP estimation allows for identification of the nonlinear model structure prior to final nonlinear parametric estimation (see [7], [9], [14]). This latter approach provides an alternative to, or a reinforcement of, other approaches to nonlinear modeling, such as neural network and neuro-fuzzy modeling (e.g., [15]). It has the advantage that the resulting models are normally much less complex, of considerably lower order, and more easily interpretable in mechanistic terms.

DBM modeling is obviously very different from physically based simulation modeling, but it is often confused with grey-box modeling. Grey-box modeling follows the traditional hypothetico-deductive approach. Here the hypothesis is normally in the form of a much-simplified model of the physical system under study, and deduction is based on estimating the parameters that characterize this assumed model structure from measured data. In contrast, DBM modeling is inductive and commences with few prior perceptions of the model form, except that it is a member of a suitable, generic class of models (normally linear/nonlinear differential equations or their discrete-time equivalents). Methods of statistical model structure identification and parameter estimation are then utilized to produce a low-dimensional, black-box model, within this generic class of models, that explains the data unambiguously in a statistically efficient manner. Only then is the question of the underlying, mechanistic interpretation of the model addressed, as illustrated by the examples discussed later. In this manner, a minimally parametrized and statistically well-defined model is obtained, and the dangers of imposing too much confidence in prior assumptions about the physical nature of the system are avoided.

Signal Processing, Forecasting, and Automatic Control

The DBM methods were developed primarily for modeling systems from normal observational time-series data obtained from monitoring exercises (or planned experimentation, if possible) carried out on the real system. Often, such time series require some form of preprocessing, and this is accomplished using methods of nonstationary time-series analysis based on recursive fixed-interval smoothing (e.g., [11], [12], and the references therein). These powerful statistical tools allow for interpolation over gaps in the time series; the detection of outliers; "signal extraction," including the estimation and removal of periodic components (e.g., seasonal adjustment); and time-frequency analysis.

The DBM models are also in a form appropriate for model-based forecasting and control system design. Of course, any available methods can be used for these applications. At Lancaster, however, the stochastic UC modeling approach based on TVP estimation provides the main

vehicle for adaptive forecasting ([12], [13], and the references therein). Model-based automatic control system design is formulated within the NMSS setting, normally resulting in the multivariable proportional-integral-plus (PIP) control algorithm (see [16] and the references therein). The NMSS is the most natural state-space definition for discrete-time transfer function models, and the PIP controller provides a natural multivariable extension of the conventional PI and PID controllers. In addition to proportional and integral action, it includes additional forward path and feedback filters that allow for NMSS state feedback, without the need for state reconstruction, thus enhancing robustness and closed-loop performance. Of course, the PIP controller is just one specific outcome of the NMSS design concept. This concept is not only attractive in its own right, but control algorithms that derive from it are very general in form and thus able to mimic other well-known design procedures, such as minimal state linear quadratic Gaussian (LQG), generalized predictive control (GPC), and Smith predictor control for time-delay systems (see [17] and the references therein).

It should be noted that many of the time-series analysis and modeling procedures mentioned above and used in the examples described below are contained in the MATLAB *CAPTAIN* time-series analysis and forecasting toolbox, currently in the final stages of beta testing (see <http://www.es.lancs.ac.uk/cres/captain/>).

Modeling and Control of a Greenhouse Microclimate

The literature on modeling of the microclimate in greenhouses is very large, with models ranging from high-order, nonlinear simulation models (e.g., [18]) through much simpler but still physically meaningful data-based models (e.g., [19]) to purely black-box models (e.g., [5]). One important aspect of this research has been the development and use of models for the control of greenhouse microclimates (e.g., [20] and the references therein). This example falls into the latter category and is concerned with the modeling and automatic control of the microclimate in the large Venlo horticultural greenhouse at the Silsoe Research Institute (SRI) in Bedford, U.K.

Prior to planned experimentation on the actual greenhouse at SRI, the control model was obtained by combined linearization and order reduction analysis [6], [7], [21]. This was applied to the high-order, nonlinear simulation model shown in Fig. 1, originally derived by scientists at SRI [18] and converted to Simulink form at Lancaster. When planned experimentation became possible, these earlier control models were replaced by the equivalent TF models obtained by identification and estimation applied to the experimental input-output data. In the final stages of the study, the control models were used to design automatic climate control systems that were first evaluated on the full, nonlinear simulation model, prior to implementation in the greenhouse.

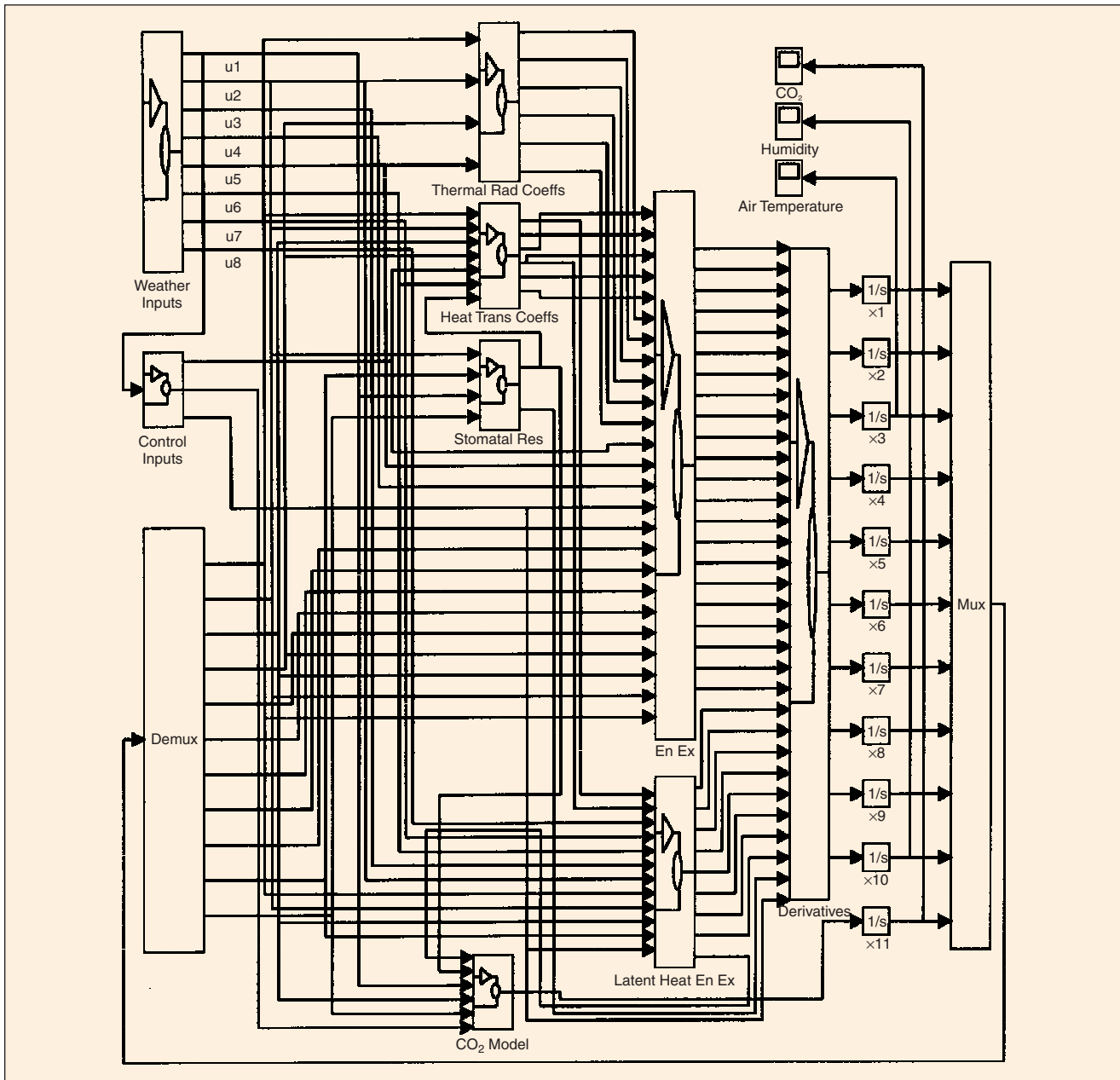


Figure 1. Simulink simulation model of the microclimate in the SRI Venlo greenhouse.

To simplify the model shown in Fig. 1, it was perturbed about a number of different operating points by suitably designed input functions. These were applied to each control input in turn (i.e., the fractional valve aperture of the heating boiler $u_{1,t}$, the input to the mist spraying system $u_{2,t}$, and the CO₂ enrichment input $u_{3,t}$), and they produced climate perturbations defined by the internal air temperature $y_{1,t}$ in °C, the percentage relative humidity $y_{2,t}$, and the CO₂ concentration $y_{3,t}$ in ppm. The input-output perturbational data set obtained in this manner was then used as the basis for the identification and estimation of reduced-order, discrete-time linear models that were directly suitable for model-based PIP system design [21].

At the most important operating condition, the estimated reduced-order model for $\mathbf{y}_t = [y_{1,t} \ y_{2,t} \ y_{3,t}]^T$, in response to $\mathbf{u}_t = [u_{1,t} \ u_{2,t} \ u_{3,t}]^T$, takes the form of the following discrete-time, transfer function matrix model:

$$\mathbf{y}_t = \begin{bmatrix} \frac{0.0143z^{-1}}{1-0.9056z^{-1}} & -0.0712z^{-1} & 0 \\ \frac{-0.0587z^{-1}}{1-0.7906z^{-1}} & 0.7909z^{-1} & 0 \\ 0 & \frac{-0.1791z^{-1}}{1-0.6996z^{-1}} & \frac{83.1294z^{-1}}{1-0.6541z^{-1}} \end{bmatrix} \mathbf{u}_t, \quad (1)$$

This model is well identified, and the parameter estimates are statistically well defined. As might be expected, the major coupling is between the temperature and relative humidity, with CO₂ being almost independent of these variables. Despite its simplicity, the model explains the high-order nonlinear model response very well, with coefficients of determination $R_r^2 > 0.99$ (i.e., over 99% of the high-order model response explained by the low-order model). Clearly, the dominant modal behavior of the simulation model at this operating condition, which is so important for subsequent control system design, has been captured very well. Moreover, this model is much simpler than might be expected on the basis of the full nonlinear model equations.

From the DBM standpoint, it is important that the reduced-order model should make good physical sense, and, as expected from mass and energy transfer considerations, the temperature and humidity dynamics are highly coupled. Indeed, it is possible to associate the model (1) with the differential equations of mass and energy transfer for this system (much simplified versions of the equations used to derive the high-order model in Fig. 1). Since the model is estimated in discrete time, however, this relationship is not too transparent. Consequently, it is better to delay our illustration of such a mechanistic interpretation to the next section, where a related model is estimated directly in continuous-time terms and the physical interpretation is much more obvious.

The high-order simulation model provided a very useful vehicle for initial PIP control system design and evaluation exercises [21], where it functioned as a valuable surrogate for the real system before experimentation was possible. Furthermore, the fact that the reduced-order models were *structurally* identical to those identified later from the real data meant that the initial PIP controller designs were also structurally similar. Note that the inherent stochastic nature of the DBM models is most important in the PIP control system design process, since it allows for evaluation of the controller's robustness to uncertainty using stochastic simulations. These can be either single simulations with stochastic inputs or MCS used to assess the robustness of the control system to uncertainty in the model parameters (e.g., [17] and the references therein).

NMSS control system design is straightforward and flexible in this example [21]. Once the reduced-order TF matrix model (1) has been obtained, it is converted straightforwardly to the left matrix fraction description (MFD), and thence to the following NMSS form:

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{G}\mathbf{u}_t + \mathbf{D}\mathbf{y}_{d,t+1} \quad \mathbf{y}_t = \mathbf{H}\mathbf{x}_t \quad (2)$$

where \mathbf{F} and \mathbf{u}_t are defined as at the bottom of the next page and

$$\mathbf{G} = \begin{bmatrix} 0.0143 & -0.0712 & 0 \\ -0.0587 & 0.7909 & 0 \\ 0 & -0.1791 & 83.1294 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.0143 & 0.0712 & 0 \\ 0.0587 & -0.7909 & 0 \\ 0 & 0.1791 & -83.1294 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T.$$

Here $\mathbf{x}_t = [y_{1,t} \ y_{2,t} \ y_{3,t} \ y_{1,t-1} \ y_{2,t-1} \ y_{3,t-1} \ u_{1,t-1} \ u_{2,t-1} \ u_{3,t-1} \ z_{1,t} \ z_{2,t} \ z_{3,t}]^T$ is the NMSS vector; $\mathbf{u}_t = [u_{1,t} \ u_{2,t} \ u_{3,t}]^T$ is the control input vector; and $\mathbf{y}_{d,t} = [y_{d1,t} \ y_{d2,t} \ y_{d3,t}]^T$ is the command input vector (the user-defined levels of the three climate variables). The variables $z_{i,t}$, $i=1,2,3$ are integral-of-error states, introduced to ensure type 1 servomechanism performance (unity gain to the command inputs and steady-state decoupling: i.e., in the steady state, the controlled variables all reach their demand levels without any effect on the steady-state levels of the other output variables).

PIP control system designs based on this NMSS model (2) provide the basis for any SVF design procedures, such as optimal linear-quadratic (LQ), pole assignment, or risk-sensitive/robust design. In the simplest LQ case, for example, quadratic cost function weighting matrices $\mathbf{Q} = \text{diag}[11111111100525]$ and $\mathbf{R} = \text{diag}[10.10.1]$ produce the SVF control law below (equation \mathbf{u}_t at the bottom of the next page), where the off-diagonals of the last 3×3 block in the matrix have been constrained to zero consistent with the steady-state decoupling requirement. Note that the pair $[\mathbf{F}, \mathbf{G}]$ is stabilizable by the nonminimal state feedback, and, in addition, we can always find a matrix \mathbf{E} such that the Cholesky decomposition $\mathbf{E}\mathbf{E}^T = \mathbf{Q}$, ensuring that the pair $[\mathbf{F}, \mathbf{E}]$ is observable [22]. Consequently, this PIP-controlled system will be stable and achieve the desired LQ optimality.

The results obtained when this controller is applied to the model are shown by the fine lines in Fig. 2. The responses show some coupling between the controlled variables, with a 10% change in humidity resulting in a transient error of 1°C in the internal air temperature. Note

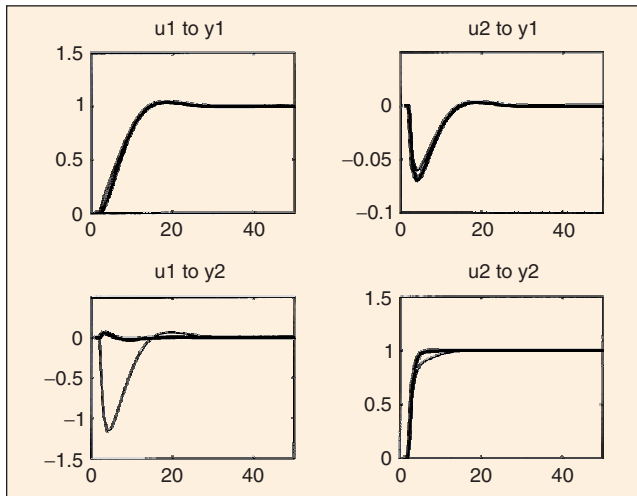


Figure 2. Closed-loop step responses for standard PIP-LQ (thin curves) and decoupling PIP-LQ (thick curves) controllers.

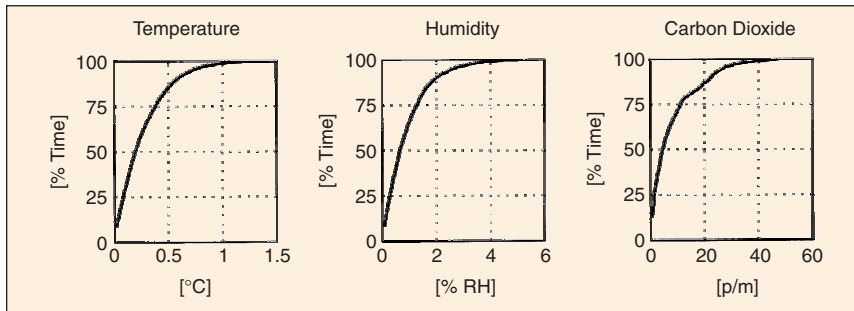


Figure 3. Long-term controller performance for the SRI Venlo greenhouse; presented as percentage of total evaluation period (% time) against absolute control error in relation to the set point (control limit).

that the relative values of the cost function weighting matrix elements associated with the three integral-of-error states have been adjusted here to ensure that the rise

times of each channel have the required control characteristics, while minimizing harsh actuator movements that could cause wear.

The controller designed in the above manner was validated in practical terms over a three-month implementation period during the 1993-1994 growing season with a tomato crop in the greenhouse. Fig. 3 shows the control performance of each climate variable over the entire validation period. Each graph shows the percentage of the validation period that a control variable was inside a certain control limit. For example, air temperature was less than 1°C away from the set point for 98% of the validation period. This performance easily meets the requirements of the growers.

Finally, it should be noted that the results shown in Figs. 2 and 3 were obtained with the simple, constant-gain, multivariable controller defined above, and some improvements in performance would be possible if more complex control systems were used. For example, further decoupling of the temperature and humidity control can be accomplished by exploiting a special form of multiobjective optimization developed at Lancaster (e.g., [16]). Here, the Cholesky factors of the LQ cost function weighting function matrices are optimized so that off-diagonal weightings are introduced into the cost function to ensure satisfactory decoupling (or other multiobjective requirements).

The responses obtained with such an optimized PIP controller are shown as the thick lines in Fig. 2; clearly, almost perfect decoupling has been achieved. Other possible improvements include a stability-guaranteed, neuro-fuzzy gain scheduling system

The responses obtained with such an optimized PIP controller are shown as the thick lines in Fig. 2; clearly, almost perfect decoupling has been achieved. Other possible improvements include a stability-guaranteed, neuro-fuzzy gain scheduling system

The responses obtained with such an optimized PIP controller are shown as the thick lines in Fig. 2; clearly, almost perfect decoupling has been achieved. Other possible improvements include a stability-guaranteed, neuro-fuzzy gain scheduling system

$$\mathbf{F} = \begin{bmatrix} 0.9056 & 0 & 0 & 0 & 0 & 0 & 0.0645 & 0 & 0 & 0 & 0 \\ 0 & 0.7906 & 0 & 0 & 0 & 0 & -0.6253 & 0 & 0 & 0 & 0 \\ 0 & 0 & 13537 & 0 & 0 & -0.4576 & 0 & 0.1171 & -58.1536 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.9056 & 0 & 0 & 0 & 0 & 0 & -0.0645 & 0 & 1 & 0 & 0 & 0 \\ 0 & -0.7906 & 0 & 0 & 0 & 0 & 0.6253 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -13573 & 0 & 0 & 0.4576 & 0 & -0.1171 & 58.1536 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{u}_t = \begin{bmatrix} 22.8406 & 1.2938 & 0 & 0 & 0 & 0 & 0.6036 & 0.0001 & -5.6927 & 0 & 0 \\ 0.6954 & 0.9666 & 0 & 0 & 0 & 0 & -0.7149 & 0.0001 & 0 & -0.7991 & 0 \\ 0.0015 & 0.0021 & 0.0021 & 0 & 0 & -0.0055 & 0 & -0.0001 & -0.6996 & 0 & 0 & -0.0112 \end{bmatrix} \mathbf{x}_t$$

(e.g., [23]) or a fully adaptive PIP control system (e.g., [24]). The latter would be straightforward to implement, since the recursive estimation algorithms used in the modeling stage of the design can be easily implemented online in real time. However, the performance shown in Fig. 3 is perfectly acceptable for most practical purposes, and such added complexity is not warranted in this case.

Modeling and Control of Forced Ventilation Systems

This example is related to the previous one and concerns the analysis of data from planned experiments in a large instrumented chamber in the Laboratory for Agricultural Buildings Research at the Katholieke Universiteit Leuven (for details, see [25]). This chamber, shown in Fig. 4, has a volume of 9 m³ and has been designed to represent a scale model of a livestock building or greenhouse for use in experimental research on forced ventilation and heating in agricultural buildings. The two variable inputs, ventilation rate (120-300 m³/h) and heating element input (0-400 W), determine the dominant airflow pattern within the chamber. An envelope chamber or “buffer zone” is constructed around the test room to minimize disturbance of the airflow by heat conduction from the laboratory. A series of aluminum conductor heat sinks and steam generation from a water reservoir provide the internal heat and moisture production to simulate animal occupants. To gain information about the distribution of mass and energy in a quantitative manner, 36 temperature sensors and 24 humidity sensors are positioned in a three-dimensional (3-D) array within the chamber.

The agricultural engineering literature reports a great deal of research on the modeling of ventilated air spaces, much of it based on deterministic simulation models. More recent data-based modeling has been stimulated by the control volume (CV) concept of Barber and Ogilvie [26]. For example, Berckmans et al. [25] use least-squares methods to estimate the parameters in a simple conceptual CV model of an imperfectly mixed 3-D airspace from variations in ventilation rate and heat supply, whereas Daskalov [27] develops a model for measured temperature and humidity variations in a naturally ventilated pig building using discrete-time TF models. To obtain the DBM model, which can be considered a natural extension of these earlier models, the continuous-time SRIV estimation algorithm is used first to identify the linear TF (ordinary differential equation) model between the measured temperatures at the inlet, T_i , and the outlet, T . In the case of the response to a step increase of the ventilation rate from 80 to 300 m³/h, the estimated model for the change in temperature ΔT from the initial steady levels takes the form

$$\Delta T(t) = \frac{1.959s + 0.089}{s^2 + 2.962s + 0.111} \Delta T_i(t) + \xi_t \quad (3)$$

where s is the differential operator and ξ_t is the residual noise. Fig. 5 compares the output of this model with the mea-

sured change in the outlet temperature, and the associated residual ξ_t is plotted in the lower graph. This unexplained residual has a zero mean value and low variance (0.021 compared with a measured variance of 2.619 for the output ΔT series; $R_T^2 = 0.992$).

The TF model can be decomposed into a parallel or feedback connection of first-order processes, but the latter has more physical significance, as required in DBM modeling. This becomes clear if we invoke the classical theory of heat transfer and formulate the differential equations of dynamic heat transfer, bearing in mind the need to arrive at a second-order, lumped, differential equation with a TF similar to (3). Such analysis [28] yields the following equations for the main chamber and the buffer zone:

$$\text{Main Chamber: } \frac{d(\Delta T)}{dt} = \beta_1 \Delta T_i + K_1 \Delta T_{\text{buff}} - \alpha_1 \Delta T$$

$$\text{Buffer Zone: } \frac{d(\Delta T_{\text{buff}})}{dt} = -K_2 \Delta T_{\text{buff}} + K_3 (\Delta T - \Delta T_{\text{buff}}).$$

In these equations,

$$\beta_1 = \frac{V}{\text{vol}_1}; \quad \alpha_1 = \left(\frac{V}{\text{vol}_1} + \frac{k_1 s f_1}{\text{vol}_1 \gamma_1 c p_1} \right); \quad K_1 = \frac{k_1 s f_1}{\text{vol}_1 \gamma_1 c p_1};$$

$$K_2 = \frac{k_2 s f_2}{\text{vol}_2 \gamma_1 c p_1}$$

and

$$K_3 = \frac{k_1 s f_1}{\text{vol}_2 \gamma_1 c p_1}$$

where T is the temperature (°C) measured at the outlet port but assumed to be representative of the temperature in the active mixing volume (AMV) [10]; T_i is the temperature

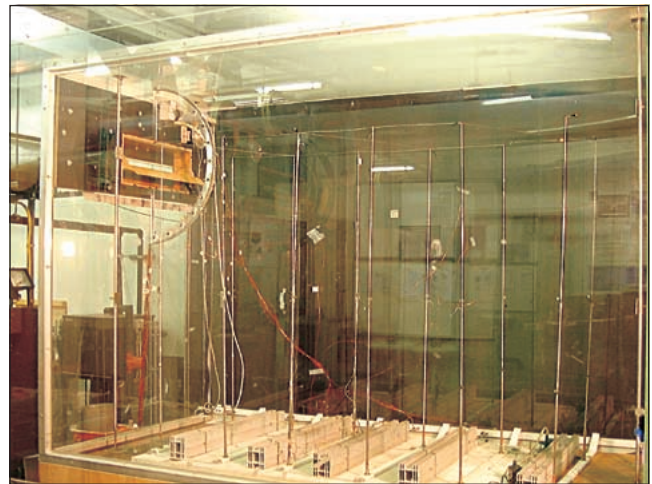


Figure 4. Large instrumented chamber in the laboratory at Katholieke Universiteit Leuven.

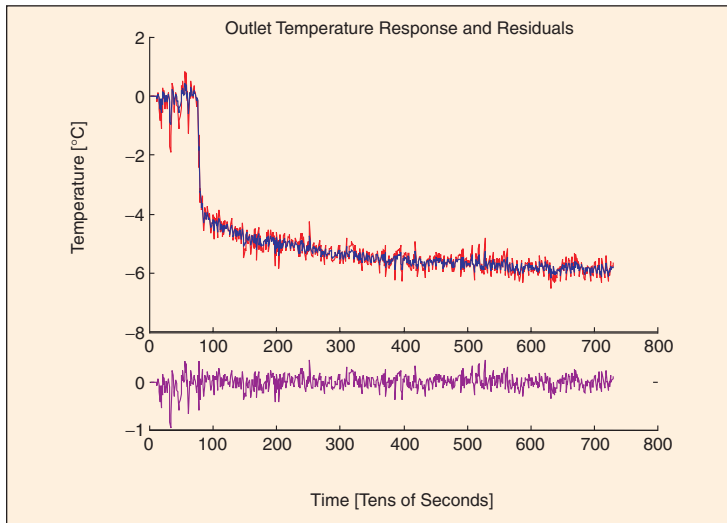


Figure 5. Comparison of the model estimated temperature and the measured temperature at the chamber outlet.

(°C) measured at the chamber inlet; V is the ventilation rate (m^3s^{-1}); γ_1 is the air density of the incoming air and the air in the main chamber (kgm^{-3}); cp_1 is the specific heat capacity of the incoming air ($\text{Jkg}^{-1}\text{°C}^{-1}$) and the air in the main chamber; k_1 is the thermal conduction coefficient through the inner chamber walls ($\text{Wm}^{-2}\text{°C}^{-1}$); sf_1 is the surface area of the AMV for imperfect mixing (m^2); T_{buff} is the temperature in the outer buffer zone (°C); vol_2 is the volume of the AMV associated with the buffer zone (m^3); k_2 is the thermal conduction coefficient through the buffer zone (outer chamber) wall ($\text{Wm}^{-2}\text{°C}^{-1}$); and sf_2 is the surface area of the buffer zone AMV for imperfect mixing (m^2).

The differential equations above may be combined and expressed in TF form to yield the following second-order TF model for the complete chamber-buffer zone system:

$$\Delta T = \frac{b_0 s + b_1}{s^2 + a_1 s + a_2} \Delta T_i \quad (4)$$

where

$$b_0 = \beta_1; \quad b_1 = \beta_1(K_2 + K_3); \quad a_1 = (K_2 + K_3 + \alpha_1); \\ a_2 = (K_2 + K_3)\alpha_1 - K_1 K_3.$$

Clearly, the TF model (4) has exactly the same structural form as the estimated TF model (3), so in DBM modeling terms, it can be considered as one particular physical interpretation of the data-based model that has scientific credibility. The first-order TF numerator ($B_j, j=0,1$) and denominator ($A_j, j=0,1$) parameters in the feedback decomposition can now be calculated as follows:

$$B_0 = \beta_1; \quad A_0 = \alpha_1 = \left(\frac{V}{vol_1} + \frac{k_1 sf_1}{vol_1 \gamma_1 cp_1} \right); \\ B_1 = \frac{K_1 K_3}{\beta_1} = \frac{k_1^2 sf_1^2}{vol_2 \gamma_1^2 cp_1^2 V};$$

$$A_1 = (K_2 + K_3) = \frac{k_2 sf_2}{vol_2 \gamma_1 cp_1} + \frac{k_1 sf_1}{vol_2 \gamma_1 cp_1}.$$

These two first-order TF's have very different time constants: T_{c1} , associated with the A_0 parameter, reflects the fast response within the main chamber:

$$T_{c1} = \frac{1}{A_0} = \frac{1}{\frac{1}{vol_1} \left[V + \frac{k_1 sf_1}{\gamma_1 cp_1} \right]}$$

and T_{c2} , associated with the A_1 parameter and the much slower temperature effect due to feedback conduction from the buffer zone, i.e., $T_{c2} = 1/A_1$.

The above analysis is useful because it shows how the inductive approach of DBM modeling objectively defines the lowest order structure (dynamic order and feedback nature) of the model prior to defining a mechanistic interpretation compatible with

this structure. Note also that the conventionally defined heat transfer parameters used in the above mechanistic formulation of the equations are not identifiable from the above relationships. In contrast, the parameters of the model (4) have direct physical significance, and they can be identified fully from the measurements obtained in the chamber experiments. Consequently, this model is not only useful in subsequent control system design terms, but it also provides an alternative, mechanistic characterization of the heat transfer dynamics that is useful in its own right.

The above model is currently being used in the design of a multivariable PIP controller for a chamber similar to that described in the previous section. However, other associated research has been concerned with the modeling and control of the fans used in the forced ventilation systems. Fig. 6, for example, is a diagram of the fan test installation at Leuven. The model for this system is identified by the SRIV algorithm as a first-order discrete time TF, and this model has been used as the basis for PIP and GPC system design. This study confirms that the model-based control algorithms offer better performance than the conventional PID controller in terms of both the regulation of ventilation rate and the reduction of energy consumption. In particular, for a 450-mm axial fan, the normalized mean square errors are as follows: PIP = 1, GPC = 1.64, PID = 5.59. These results were obtained for an airflow rate of 3000 m^3/h with realistic wind disturbances, but, in order to evaluate their practical robustness, all three controllers were optimized for an airflow rate set point in the neighborhood of 1500 m^3/h , without wind.

Environmental Forecasting

Previous publications from Lancaster illustrate how the modeling and forecasting procedures outlined above have been applied to environmental and other time series (e.g., [13] and the references therein). Here, we will consider the results obtained in the case of two practically important en-

vironmental examples: first, the famous series of monthly atmospheric CO₂ measurements at Mauna Loa in Hawaii, as shown in Fig. 7, which did much to promote the current debate on global warming; and second, the hourly rainfall and river flow series from the River Ribble Catchment in northwest England, as shown in Fig. 8.

Adaptive Forecasting, Backcasting, and Interpolation of the Mauna Loa CO₂ Series

It is clear from Fig. 7 that the variations in atmospheric CO₂ at Mauna Loa are dominated by a long-term, upward trend and annual periodicity. As a result, the identified UC model takes the form

$$y_t = T_t + C_t + e_t \quad (5)$$

where

$$T_t = a_0 + a_1 t + a_2 t^2 + d_t \text{ and } C_t = \sum_{j=1}^2 a_{j,t} \cos(\omega_j t) + b_{j,t} (\sin \omega_j t) \quad (6)$$

and y_t denotes the CO₂. As shown, the trend T_t is identified as a polynomial in time t , with unknown but constant parameters, combined with a residual signal d_t that models the medium-term, stochastic deviations about this polynomial trend. The annual seasonality is modeled by the periodic component C_t : this is of a standard trigonometric (Fourier) form but with stochastically time-varying parameters (TVPs) $a_{j,t}$, $b_{j,t}$, $j=1,2$. This type of relationship will model any periodic behavior defined by the frequencies ω_j , $j=1,2$, and the TVPs allow for the estimation of changes in the amplitude and phase of these seasonal variations. In the present CO₂ example, the frequency $\omega_1 = 2\pi/12$ is the fundamental frequency associated with the 12-month seasonal cycle, and $\omega_2 = 2\pi/6$ is its first harmonic. The spectrum of the data suggests that the other harmonics are insignificant and can be omitted. The TVPs d_t , $a_{j,t}$, $b_{j,t}$, $j=1,2$ in this UC model are identified and modeled as simple random walk processes in an associated set of stochastic state equations (see [13]). The hyper-parameters (e.g., noise variance ratios) in this state-space model are estimated by a special form of optimization in the frequency domain (see [12] and [13]) and the recursive KF and FIS algorithms are then used for forecasting, backcasting, and interpolation of the series.

Typical results are shown in Figs. 9 and 10. For this analysis, the estimation data set consisted of the first 313 samples (1958(1)-1985(5)) of the 504 data set, but with two years of these data between 1971(8) and 1973(8) omitted to show how the algorithms interpolate over such a gap. For valida-

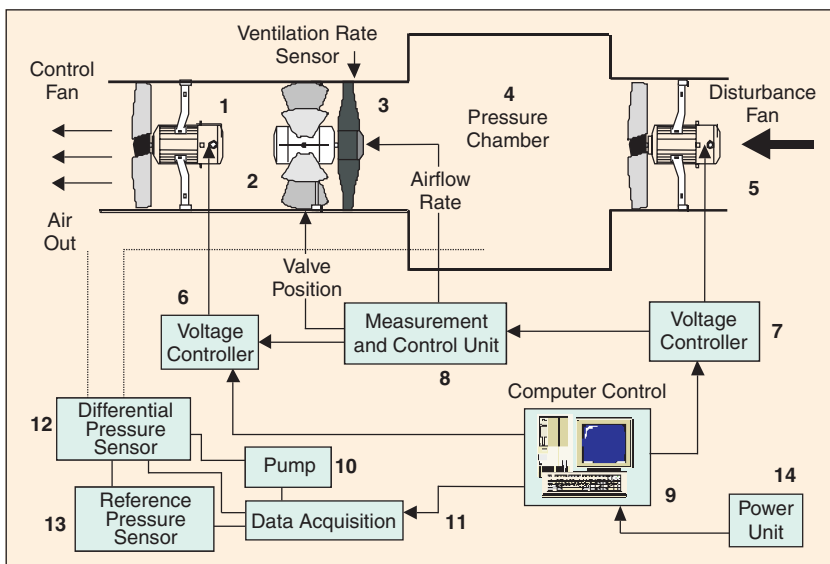


Figure 6. Schematic layout of the test chamber and associated control equipment.

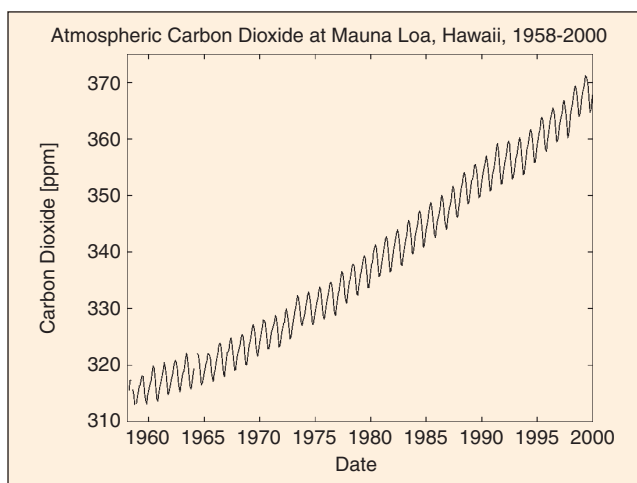


Figure 7. Atmospheric CO₂ measured at Mauna Loa, Hawaii, 1958-2000.

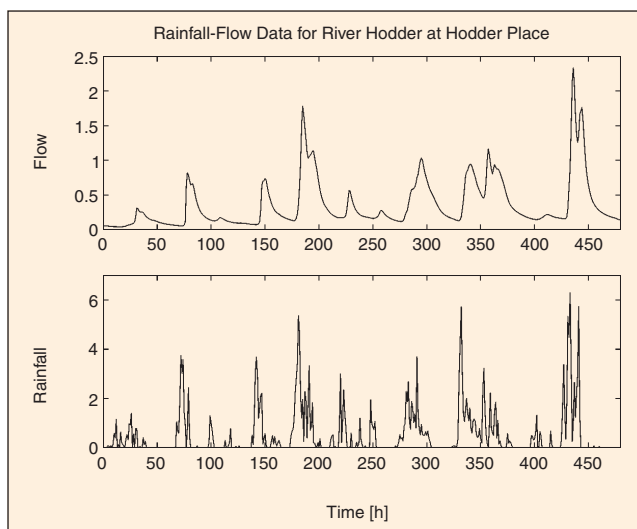


Figure 8. Hourly rainfall (lower panel) and flow (upper panel) series for the River Hodder at Hodder Place in the River Ribble Catchment of northwest England.

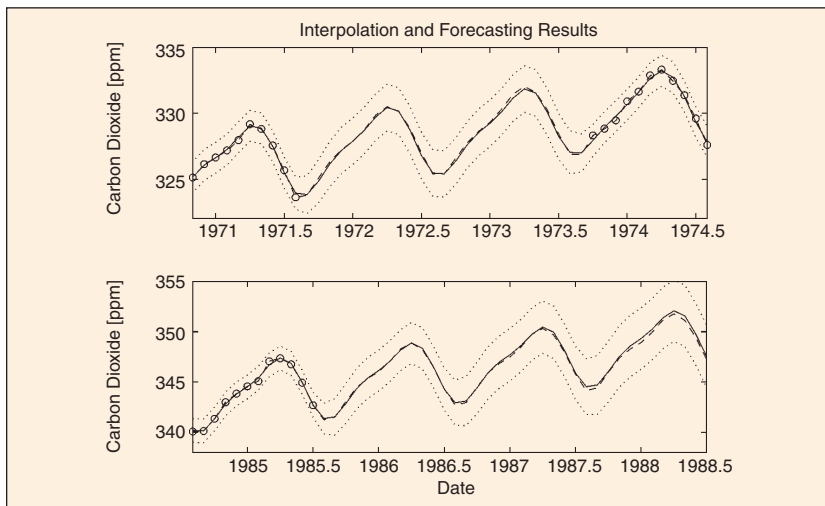


Figure 9. Two-year interpolation (upper panel) and three-year-ahead forecasting (lower panel) results for the Mauna Loa CO₂ series.

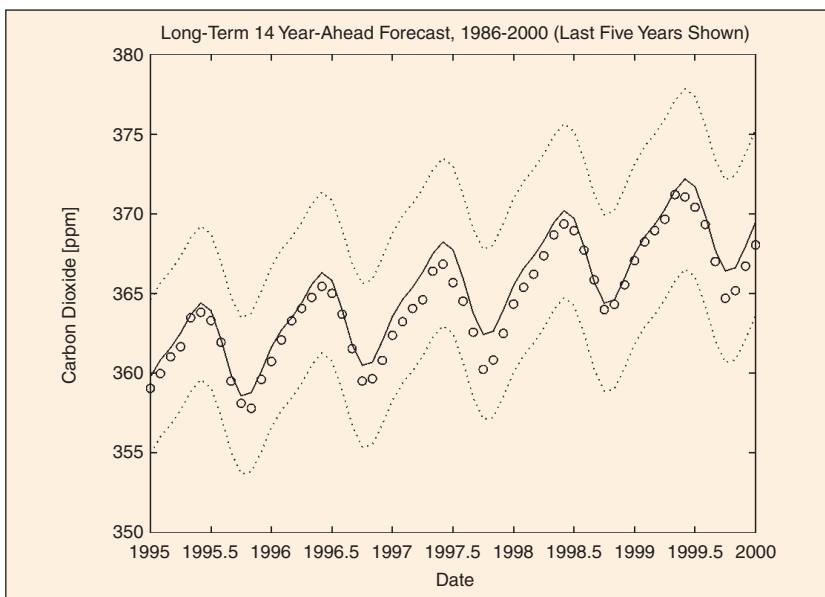


Figure 10. Long-term, 14.4-year-ahead forecast of the Mauna Loa CO₂ series, showing only the last five years for clarity.

tion purposes, true multistep-ahead forecasting is carried out in the short term (three years ahead) and long term (14.4 years ahead to 1999(12)), respectively, based only on the estimation data set, with the model optimized on the basis of these same limited data (306 samples). The top panel in Fig. 9 shows the two-year interpolation results, whereas the lower panel shows the three-year-ahead forecasts. In both cases, the results are excellent, with very small errors between the interpolates/forecasts (full lines) and the data (dashed lines, not used in the analysis). The estimated standard error bounds (95% confidence intervals) are shown as dotted lines. Fig. 10 shows the last 5.5 years of the 14.4-year-ahead forecast. Here, the data (again not used in the estimation or forecasting) are plotted as circular points and the forecast as a solid line, with the standard error

bounds shown dotted. Given the very long forecasting interval, these results are quite remarkable and show how predictable this important series can be if sufficiently powerful forecasting algorithms are used. As far as the authors are aware, these are the best forecasting results produced so far for this series.

Adaptive Flow Forecasting in the Ribble Catchment

Once again, the scientific literature abounds with different models of rainfall-flow processes in river catchments. These vary considerably in complexity from those based on continuum mechanics (solved approximately via finite difference or finite element spatiotemporal discretization methods) [29] through conceptual models such as TOPMODEL [30] to hybrid-metric-conceptual (HMC) grey-box models such as IHACRES (e.g., [31]), which are similar in structure and complexity to the DBM model equivalent [32]. The main aim of the River Ribble study [33] was threefold: first, to obtain DBM models relating rainfall to the river flows measured in the Ribble catchment; second, to evaluate the performance of a state-adaptive, KF-based approach to forecasting, with the state dynamics defined by these DBM models; and finally, to compare this state-adaptive approach with its predecessor at Lancaster, the parameter-adaptive forecasting system developed for the Solway River Purification Board as a flood warning system for the town of Dumfries in Scotland (e.g., [34]). Another, secondary objective following from the development

of the Lancaster adaptive radar calibration (ARC) system for the National Rivers Authority/Environment Agency [35] was to consider how well forecasting performance using weather radar measured rainfall compared with the performance using more conventional ground-based methods.

In the case of a uniform sampling interval of one hour, SRIV identification and estimation yields the following discrete-time TF model between the gauged rainfall r_t and the measured flow y_t , for the River Hodder at Hodder Place in the Ribble catchment during December 1993:

$$y_t = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} g\{r_{t-4}\} + \xi_t. \quad (7)$$

Here, $g\{r_{t-4}\}$ is a nonlinear function of the rainfall delayed by four sampling intervals to introduce the pure (advective) time delay between the occurrence of the rainfall and its first measured effect on the river flow. The noise level on flow data is normally quite large, and the noise term ξ_t is introduced here to reflect the combined effects of measurement noise, unmeasured disturbances, and imperfections in the model.

In this case, SDP estimation shows that a rainfall nonlinearity is present and takes the form

$$u_{t-4} = g\{r_{t-4}\} = c \cdot f\{y_t\} \cdot r_{t-4} \quad (8)$$

where u_t is termed the “effective rainfall” (or “rainfall excess”) and c is a scale factor chosen conventionally so that the volume of the effective rainfall is equal to the total stream flow volume over the estimation period. Equation (8) shows that the rainfall affects the flow via a multiplicative nonlinearity between r_{t-4} and a nonlinear function of flow $f\{y_t\}$, where y_t is acting as a surrogate for the soil moisture, which is difficult to measure [9]. The estimated nonlinear function shows that the SDP is small for low flows (low soil moisture) and increases, but with diminishing slope, for high flows. Consequently, the effective rainfall is lower than the actual rainfall for low soil moisture, since it tends to be absorbed by the dry soil under these conditions. At high flow (high soil moisture), however, the rainfall is much more effective in inducing flow variations.

Based on this SDP identification analysis and subsequent parametric estimation, the rainfall-flow nonlinearity is parametrized using a power law $f\{y_t\} = y_t^\theta$ to approximate $f\{y_t\}$. The SRIV estimates of the associated parameters, as well as the TF parameters, are shown in Table 1. This also includes some parameters that are derived from decomposition of the TF model into the following parallel connection of two first-order processes (e.g., [9]):

$$\frac{b_0 + b_1 z^{-1}}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}} = \frac{\beta_1}{1 + \alpha_1 z^{-1}} + \frac{\beta_2}{1 + \alpha_2 z^{-1}} \quad (9)$$

where α_i and $\beta_i, i=1,2$ are, respectively, the eigenvalues of the TF denominator and the associated residues in the partial fraction expansion of the TF. The latter define the weighting or “partitioning” associated with this decomposition. It is clear from (8) and (9) that the effective rainfall u_{t-4} can be partitioned into two parallel pathways resulting in two component flows that, when added together, yield the total measured river flow. In effect, therefore, the decomposition provides estimates of two “unobserved” (“latent” or “hidden”) flow states and their associated rainfall-flow dynamics.

The estimates in Table 1(b) reveal that the two pathways have very different dynamics. The “quick-flow” pathway has a total travel time ($T_i + \delta$) of 7.5 hr, whereas the “slow-flow” pathway has a total travel time of 77 hr. The associated par-

Table 1. (a) Model parameter estimates (to three decimal places).	
Parameter Estimates	
$\hat{\alpha}_1 = -1.739(0.015); \hat{\alpha}_2 = 0.743(0.014)$	
$\hat{b}_0 = 0.153(0.006); \hat{b}_1 = -0.149(0.005)$	
$\hat{\alpha}_1 = 0.986; \hat{\alpha}_2 = 0.753; \hat{\beta}_1 = 0.006; \hat{\beta}_2 = 0.147$	
$\hat{c} = 0.847; \hat{\vartheta} = 0.160(0.0001)$	

(b) Derived physically meaningful parameters for (9).		
Physically Meaningful Derived Parameters	Slow Flow	Quick Flow
SSG, G_i	0.40	0.60
Residence time, T_i	73 hr	3.5 hr
Advective delay, δ	4 hr	4 hr
Partition %, p_i	40%	60%

tion percentages of 60% and 40%, respectively, suggest that more of the effective rainfall affects the quick-flow pathway than the slow-flow pathway. Note, however, that if the uncertainty in the parameter estimates is taken into account using MCS analysis (see previous discussion and [8]), then some of the derived parameters in Table 1 have wide confidence intervals and non-Gaussian distributions. In particular, and not surprisingly, the slow-flow dynamics are much more poorly defined than the quick-flow dynamics.

As we have stressed, an important aspect of DBM modeling is the physical interpretation of the model. Given the derived parameters in Table 1, the most obvious physical interpretation of the model (9) is that the effective rainfall affects the river flow nonlinearly via two main pathways. First, the initial rapid rise in the hydrograph, following an instantaneous rise in effective rainfall, derives from the quick-flow pathway, probably as the aggregate result of the many surface processes active in the catchment. The subsequent long tail in the recession of the hydrograph is associated mainly with the slow-flow component, probably as the result of water displacement within the groundwater system.

For flow forecasting purposes, the complete catchment model for the River Ribble is formulated in terms of rainfall-flow models such as (7), together with flow-flow TF models that route the flow down the river system. All of these TF models are then combined into a set of stochastic state equations with a state vector \mathbf{x}_t . In the case of the rainfall-flow models, such as (7), the associated state variables are defined as the latent states defined by decompositions such as (9), and the rainfall inputs are the effective rainfall series, such as u_t . The estimate $\hat{\mathbf{x}}_t$ of the state vector \mathbf{x}_t and, from this, the recursive estimate \hat{y}_t of the river flow y_t can

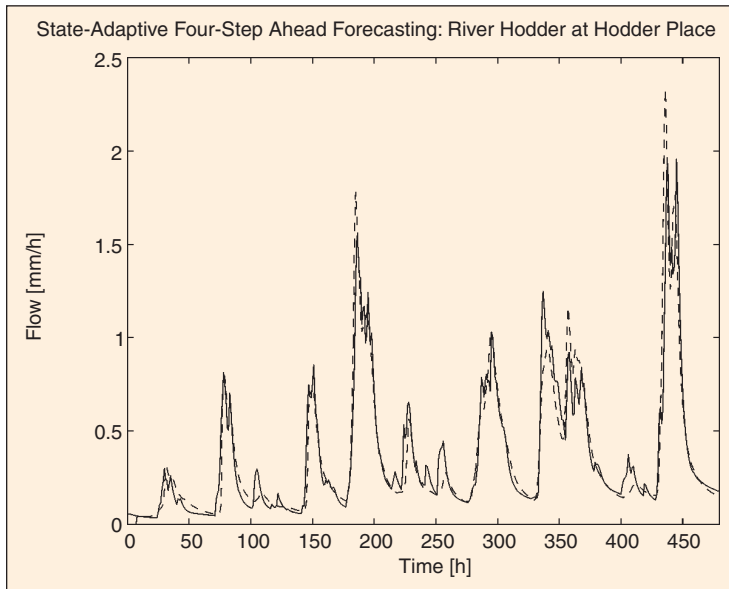


Figure 11. Four-step-ahead forecasting results for the River Hodder at Hodder Place in December 1993: forecasts (solid line) compared with measured flow series (dashed).

then be obtained from the associated KF algorithm (see [11], [33], and [36]), whose stochastic hyper-parameters (variance ratios) are estimated by maximum likelihood optimization based on prediction error decomposition [37]. The resulting four-step-ahead forecasting results are obtained from the prediction step in the KF update equations and are shown graphically in Fig. 11.

The forecasting performance shown in Fig. 11 can be improved a little if the noise ξ_t in (7) is modeled as an AR process and introduced into the state-space description (thus increasing its dimension to include the additional stochastic states). Even without this improvement, however, the performance is good, and this is reflected in the statistical properties of the forecast errors, which have a mean value near zero ($0.0032 \text{ mm}\cdot\text{hr}^{-1}$) and a standard deviation $\sigma_{e_4} = 0.132 \text{ mm}\cdot\text{hr}^{-1}$. The coefficient of determination based on the variance of the four-step-ahead prediction errors is $R_4^2 = 0.871$, compared with $R_4^2 = 0.484$ for the naïve (persistence) forecast, where the predicted four-hour-ahead flow is equal to current flow.

Conclusions

This article has briefly reviewed the main aspects of the generic DBM approach to modeling stochastic dynamic systems and shown how it is being applied to the analysis, forecasting, and control of environmental and agricultural systems. The advantages of this inductive approach to modeling lie in its wide range of applicability. It can be used to model linear, nonstationary, and nonlinear stochastic systems, and its exploitation of recursive estimation means that the modeling results are useful for both online and offline applications. To demonstrate the practical utility of the various methodological tools that underpin the DBM ap-

proach, the article also outlined several typical, practical examples in the area of environmental and agricultural systems analysis, where DBM models have formed the basis for simulation model reduction, control system design, and forecasting. However, the same methods have been used in many other applications in diverse areas of science and social science. Recent applications include:

- *Engineering*: the delta operator modeling and autostabilization of the Harrier VSTOL aircraft in its most difficult transitional mode between hovering and normal flight [7], [16], the modeling and control of interurban traffic systems [38], and both model reduction and control of an industrial gasifier system [39].
- *Ecology*: the modeling of limit cycling behavior in blowfly population dynamics [7], [14].
- *Biology*: DBM modeling has revealed new aspects of the dynamics associated with stomatal behavior in plants [40]; the chaotic electrical activity in the axon of a squid has also been modeled [41], [42].
- *Business Data Analysis and Forecasting*: UC modeling and forecasting of microeconomic and business data for the credit card company Barclaycard, U.K. [43]
- *Macroeconomics*: where DBM modeling yields some thought-provoking, and rather topical, insights into the relationship between government spending, private capital expenditure, and unemployment in the United States between 1948 and 1998 [44].

Acknowledgments

The authors are grateful to the United Kingdom Biotechnology and Biological Sciences Research Council for its support, through Research Grants 89/E06813 and 89/MMI09731, together with the Ph.D. studentship 98/B1/E/04226. The authors would also like to thank all the past and present members of the Environmental Systems and Control Group in CRES at Lancaster University who have contributed to the research described in this article, as well as Prof. Daniel Berckmans and his staff in the Laboratory for Agricultural Buildings Research at the Katholieke Universiteit Leuven, Belgium.

References

- [1] K.J. Beven, "Prophecy, reality and uncertainty in distributed hydrological modelling," *Adv. Water Resources*, vol. 16, pp. 41-51, 1993.
- [2] K. Keesman and G. Van Straten, "Set membership approach to identification and prediction of lake eutrophication," *Water Resources Research*, vol. 26, pp. 2643-2652, 1990.
- [3] S. Parkinson and P.C. Young, "Uncertainty and sensitivity in global carbon cycle modelling," *Climate Res.*, vol. 9, pp. 157-174, 1998.
- [4] K. Hasselmann, "Climate-change research after Kyoto," *Nature*, vol. 390, pp. 225-226, 1997.
- [5] I. Seginer, "Some artificial neural network applications to greenhouse environmental control," *Comput. Electron. Agriculture*, vol. 18, pp. 167-186, 1997.

- [6] P.C. Young, S.D. Parkinson, and M. Lees, "Simplicity out of complexity in environmental systems: Occam's Razor revisited," *J. Appl. Stat.*, vol. 23, no. 2&3, pp. 165-210, 1996.
- [7] P.C. Young, "Data-based mechanistic modelling of environmental, ecological, economic and engineering systems," *Environ. Modeling Software*, vol. 13, pp. 105-122, 1998.
- [8] P.C. Young, "Data-based mechanistic modelling, generalized sensitivity and dominant mode analysis," *Comput. Phys. Commun.*, vol. 115, pp. 1-17, 1999.
- [9] P.C. Young, "Time variable and state dependent parameter modelling of nonstationary and nonlinear time series," in *Developments in Time Series*, T.S. Rao, Ed. London: Chapman and Hall, 1993, pp. 374-413.
- [10] P.C. Young and M.J. Lees, "The active mixing volume (AMV): A new concept in modelling environmental systems," in *Statistics for the Environment*, V. Barnett and K.F. Turkman, Eds. Chichester: Wiley, 1993, pp. 3-44.
- [11] P.C. Young, *Recursive Estimation and Time-Series Analysis*. Berlin: Springer-Verlag, 1984.
- [12] P.C. Young, "Nonstationary time series analysis and forecasting," *Progress Environ. Sci.*, vol. 1, pp. 3-48, 1999.
- [13] P.C. Young, D. Pedregal, and W. Tych, "Dynamic harmonic regression," *J. Forecasting*, vol. 18, pp. 369-394, 1999.
- [14] P.C. Young, "Stochastic, dynamic modelling and signal processing: Time variable and state dependent parameter estimation," in *Nonstationary and Nonlinear Signal Processing*, W.J. Fitzgerald, A. Walden, R. Smith, and P.C. Young, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2000, pp. 74-114.
- [15] J.-S.R. Jang, C.-T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Upper Saddle River, NJ: Prentice-Hall, 1997.
- [16] A. Chotai, P.C. Young, P. McKenna, and W. Tych, "Proportional-integral-plus design for delta operator systems: Part 2, MIMO systems," *Int. J. Contr.*, vol. 70, no. 1, pp. 149-168, 1998.
- [17] C.J. Taylor, A. Chotai, and P.C. Young, "State space control system design based on nonminimal state-variable feedback: Further generalisation and unification results," *Int. J. Contr.*, vol. 73, no. 14, pp. 1329-1345, 2000.
- [18] Z.S. Chalabi and B.J. Bailey, "Simulation of the energy balance in a greenhouse," Div. Note, DN1516. AFRC Silsoe Research Institute, Bedford, 1989.
- [19] A.J. Udink ten Cate, "Modelling and simulation in greenhouse climate control," *Acta Horticulturae*, vol. 174, pp. 461-467, 1985.
- [20] G. Van Straten, H. Challa, and F. Buwalda, "Towards user accepted optimal control of greenhouse climate," *Comput. Electron. Agriculture*, vol. 26, pp. 221-238, 2000.
- [21] M. Lees, P.C. Young, A. Chotai, and W. Tych, "A nonminimal state variable feedback approach to multivariable control of glasshouse climate," *Trans. Inst. Measure. Control*, vol. 17, pp. 200-211, 1995.
- [22] B.C. Kuo, *Digital Control Systems*. New York: Holt, Rinehart and Winston, 1980.
- [23] S.G. Cao, N.W. Rees, and G. Feng, "Stability analysis and design for a class of continuous-time fuzzy control systems," *Int. J. Contr.*, vol. 64, pp. 1069-1087, 1996.
- [24] A. Chotai, P.C. Young, and M.A. Behzadi, "Self-adaptive design of a nonlinear temperature control system," *Proc. Inst. Elec. Eng.*, vol. 138, pt. D, pp. 41-49, 1991.
- [25] D. Berckmans and M. De Moor, "Analysis of the control of livestock environment by mathematical identification on measured data," in *Proc. Int. Winter Meeting Ameri. Soc. Agric. Eng.*, Chicago, IL, 1993, paper no. 934574.
- [26] E.M. Barber and J.R. Ogilvie, "Incomplete mixing in ventilated spaces: Part 1. theoretical considerations," *Canadian Agric. Eng.*, vol. 24, pp. 25-29; Part 2. scale model study, vol. 26, pp. 189-196, 1984.
- [27] P.I. Daskalov, "Prediction of temperature and humidity in a naturally ventilated pig building," *J. Agric. Eng. Res.*, vol. 68, pp. 329-339, 1997.
- [28] L.E. Price, P.C. Young, D. Berckmans, K. Janssens, and C.J. Taylor, "Data-based mechanistic modelling and control of mass and energy transfer in agricultural buildings," *Annu. Rev. Contr.*, vol. 23, pp. 71-82, 1999.
- [29] M.B. Abbott, J.C. Bathurst, J.A. Cunge, P.E. O'Connell, and J.L. Rasmussen, "An introduction to the European Hydrology System (SHE). 2: Structure of a physically based distributed modelling system," *J. Hydrology*, vol. 87, pp. 61-77, 1986.
- [30] K.J. Beven and M.J. Kirkby "A physically-based variable contributing area model of basin hydrology," *Hydrological Sci. J.*, vol. 24, pp. 43-69, 1979.
- [31] A.J. Jakeman and G.M. Hornberger, "How much complexity is warranted in a rainfall-runoff model?" *Water Resources Res.*, vol. 29, pp. 2637-2649, 1993.
- [32] P.C. Young, "Data-based mechanistic modelling and validation of rainfall-flow processes," in *Model Validation: Perspectives in Hydrological Science*, M.G. Anderson and P.D. Bates, Eds. Chichester: Wiley, 2001, pp. 117-161.
- [33] P.C. Young and C. Tomlin, "Data-based mechanistic modelling and adaptive flow forecasting," in *Flood Forecasting: What Does Current Research Offer the Practitioner?* M.J. Lees and P. Walsh, Eds. British Hydrological Society (BHS) Occasional Paper No. 12, Wallingford, U.K., 2000, pp. 26-40.
- [34] M.J. Lees, P.C. Young, K.J. Beven, S. Ferguson, and J. Burns, "An adaptive flood warning system for the River Nith at Dumfries," in *River Flood Hydraulics*, W.R. White and J. Watts, Eds. Wallingford, U.K.: Institute of Hydrology, 1994, pp. 65-75.
- [35] M.E. Lord, P.C. Young, and R.C. Goodhew, "Adaptive radar calibration using rain gauge data," in *Hydrological Uses of Weather Radar*, British Hydrological Society Occasional Paper No.5. London: British Hydrological Society, 1995.
- [36] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, Journal Basic Eng.*, vol. 80, pt. D, pp. 468-478, 1960.
- [37] F. Schwegge, "Evaluation of likelihood functions for Gaussian signals," *IEEE Trans. Inform. Theory*, vol. IT-11, no. 1, pp. 61-70, 1965.
- [38] C.J. Taylor, P.C. Young, A. Chotai, and J. Whittaker, "Nonminimal state space approach to multivariable ramp metering control of motorway bottlenecks," *Proc. Inst. Elec. Eng.*, vol. 145, no. 6, pp. 568-574, 1998.
- [39] C.J. Taylor, A.P. McCabe, P.C. Young, and A. Chotai, "Proportional-integral-plus (PIP) control of the ALSTOM gasifier problem," *Proc. Inst. Mech. Eng., Journal of Systems and Control Engineering*, vol. 214, pp. 469-480, 2000.
- [40] A.J. Jarvis, P.C. Young, C.J. Taylor, and W.J. Davies, "An analysis of the dynamic response of stomatal conductance to a reduction in humidity over leaves of *Cedrella odorata*," *Plant Cell and Environment*, vol. 22, pp. 913-924, 1999.
- [41] P.C. Young, "The identification and estimation of nonlinear stochastic systems," in *Nonlinear Dynamics and Statistics*, A.I. Mees, Ed. Boston: Birkhauser, 2001, pp. 127-166.
- [42] P.C. Young, P. McKenna, and J. Bruun, "Identification of nonlinear stochastic systems by state dependent parameter estimation," *Int. J. Contr.*, special issue on Nonlinear Identification and Estimation, in press, 2001.
- [43] W. Tych, D.J. Pedregal, and P.C. Young, "An unobserved component model for multi-rate forecasting of telephone call demand," *Int. J. Forecasting*, 2001.
- [44] P.C. Young and D. Pedregal, "Macro-economic relativity: Government spending, private investment and unemployment in the USA 1948-1998," *Structural Change and Economic Dynamics*, vol. 10, pp. 359-380, 1999.

Peter Young is Professor of Environmental Systems and Director of the Centre for Research on Environmental Systems and Statistics at Lancaster University. He is also Adjoint Professor of Environmental Systems at the Australian National University, Canberra. He obtained the B.Tech. and M.Sc. degrees at Loughborough University and the Ph.D. degree at Cambridge University, where he was a Fellow of Clare Hall. His most recent research has been concerned with data-based modeling, forecasting, signal processing, and control for nonstationary and nonlinear stochastic systems.

Arun Chotai is a Senior Lecturer in the Environmental Science Department at Lancaster University and joint Head of the Systems and Control Group in the Centre for Research on Environmental Systems and Statistics. He holds a Ph.D. in systems and control and a B.Sc. in mathematics, both from the University of Bath. His research interests are in modeling and control design, including self-adaptive, self-tuning, and multivariable control of nonstationary and nonlinear systems.