

Nonlinear input/output modelling

Ronald K. Pearson

E. I. du Pont de Nemours and Co., PO Box 80101, Wilmington, DE 19880-0101, USA

This paper gives a highly abbreviated overview of some of the key issues in empirical nonlinear modelling for chemical process applications. This task is complicated by the inherent nature of nonlinearity: the term describes a class of systems by the one feature they lack. In fact, this division – linear vs. nonlinear – suggests a ‘unity’ or ‘homogeneity’ of the class of nonlinear systems that does not exist. Consequently, this review will focus on specific sub-classes of nonlinear models that have analytically useful structural characteristics, and comparisons will be made both between these classes and with the more familiar linear models. Length limitations restrict these discussions somewhat, but it is hoped that the range of examples will be great enough to demonstrate how nonlinear model identification is both similar to and different from linear model identification. The general conclusion of this paper is that nonlinear input/output modelling is a vitally important practical art with many unresolved issues; the principal objective of this paper is to elucidate some of these issues.

The title of this paper immediately raises two questions:

- why nonlinear models?
- why input/output models?

The motivation to explore nonlinear models comes from the unavoidable nonlinearity of the dynamics of many chemical processes. Indeed, several of the references cited in this paper deal with the nonlinearity of two of the most important chemical process unit operations: reactions and separations. Turning to the second question, while it is true that fundamental (i.e., ‘first principles’) models generally give us more complete process understanding than empirical models do, they are also generally much more complex and require correspondingly longer to develop. Thus, while models like that developed by Congalidis *et al.*¹ for the co-polymerization of methyl methacrylate and vinyl acetate could, in principle, be used for applications like model predictive control, in practice, they are not. Also, note that besides providing simpler models, analysis of input/output data can also give us useful process insights that can be used in subsequently developing or refining fundamental models. In particular, all fundamental models are based on assumptions (e.g. these effects are important, those are negligible) and these assumptions may be wrong: empirical models can help us uncover such ‘surprises’.

As many can attest, even linear empirical modelling can be quite challenging. The objective of this paper is to give some general insights into the additional complications that can arise in nonlinear empirical model-

ling. Consequently, the next section defines both the class of linear ARMAX models and several classes of nonlinear dynamic models. The following section illustrates the general nature of nonlinearity with two specific examples of nonlinear behaviour: asymmetric responses to symmetric input changes and sub-harmonic generation in response to periodic input sequences. Building on this background, the next section discusses some of the practical considerations involved in selecting a nonlinear model structure and the following section discusses some of the important connections between nonlinear models and higher-order statistics. A brief overview follows of the important practical issue of ‘robustness’, as the term is used in statistics – the degree to which computed results are sensitive to a small fraction of ‘anomalous’ data in our complete dataset. The notions of identifiability and input sequence design are then taken up and the next section illustrates many of the ideas presented in this paper with a process example that is simple enough to be comprehensible, but complex enough to illustrate the nature of the problems involved. Finally, a brief summary of the main ideas presented is given; for a more detailed discussion of some of these ideas, refer to the forthcoming survey by Pearson and Ogunnaike².

Model structures

The following paragraphs describe a few of the more common nonlinear dynamic model structures appearing in the literature. Because many of these model struc-

tures are generalizations of linear models, the discussions will begin with the linear case. This discussion will be restricted to discrete-time formulations for three reasons: the fact that process data is typically available only at discrete time instants, the fact that model-based control algorithms like Model Predictive Control are usually formulated in discrete time, and because space constraints do not permit discussion of both continuous-time and discrete-time models. In the discrete-time formulation, one of the most popular linear representations is the ARMAX (AutoRegressive Moving Average with exogenous inputs) model discussed by Ljung³:

$$y(k) = \sum_{j=1}^p a_j y(k-j) + \sum_{j=0}^q b_j u(k-j) + \sum_{j=0}^r c_j e(k-j) \quad (1)$$

In a typical process application, this model would relate the process response sequence $\{y(k)\}$ to the input sequence $\{u(k)\}$, called an exogenous input in the statistical time-series literature. In the applications considered here, this input can be either deterministic or stochastic and represents the time-varying value of a manipulated process variable. The 'disturbance sequence' or 'noise sequence' $\{e(k)\}$ represents the combined effects of measurement noise, modelling errors (e.g. neglected nonlinearities), unmeasured disturbances etc. Probably the most common approximation is to take $\{e(k)\}$ as an independent, identically distributed (i.i.d.) sequence, setting $r = 0$. For the $r > 0$, the coefficients $\{c_j\}$ permit the description of correlated disturbance sequences. In either case, note that since $\{e(k)\}$ is a stochastic process, $\{y(k)\}$ is also a stochastic process, regardless of the nature of the input sequence $\{u(k)\}$; if the input sequence is also taken as a stochastic process, it is generally assumed to be independent of the error sequence $\{e(k)\}$. Note that this model extends directly to multiple-input ARMAX models by adding terms to the second sum, corresponding to the lagged values of additional exogenous inputs. This observation is generic and also applies to all of the nonlinear models introduced below.

Two special cases of this linear model are particularly significant. The case $p = 0$, $q = \infty$ corresponds to the convolution model with $\{b_j\}$ representing the impulse response of the system. By Wold's decomposition theorem, any stationary Gaussian stochastic process $\{y(k)\}$ may be represented in this form with $\{u(k)\}$ an i.i.d. Gaussian sequence and $\{e(k)\}$ identically zero⁴. Conversely, the case $p = \infty$, $q = 0$ represents an infinite-order autoregressive model. As in the previous case, it follows from a theorem of Kolmogorov that any stationary Gaussian process may also be represented in this form⁴. The practical implication of this result is that some linear systems are best approximated by 'low-order autoregressive models', while other systems are best approximated by 'low-order moving average models'. Conversely, it will be demonstrated in that this result does not extend to nonlinear systems: some nonlinear phenomena are 'inherently autoregressive' in nature and do not possess a 'moving average' representation.

Probably the best known class of nonlinear systems that do possess moving average representations is the Volterra model:

$$\begin{aligned} y(k) = & y_0 + \sum_{j=0}^{\infty} a_j u(k-j) \\ & + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} b_{i,j} u(k-i)u(k-j) \\ & + \sum_{l=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{l,i,j} u(k-l)u(k-i)u(k-j) \\ & + \dots \end{aligned} \quad (2)$$

This class of systems is very broad⁵, but to be practically useful, the sums must be truncated to some finite upper limit M and the number of sums included must also be made finite. As a simple example of this model class, the following second-order Volterra model will be considered repeatedly in the following sections⁶:

$$y(k) = u(k) + bu(k-1)u(k-2) \quad (3)$$

Even when M is relatively small and only a few sums are retained, objection is frequently raised to the number of parameters required to specify Volterra models, motivating interest in various structurally-constrained Volterra models.

A particularly useful subset of the class of Volterra models is the class of 'block-oriented' nonlinear models. Probably the most popular member of this class is the Hammerstein model⁷⁻¹², shown in *Figure 1* and consisting of a static nonlinearity $f(\cdot)$ followed by a linear dynamical system $H(z)$. If the function $f(\cdot)$ is analytic, it is not difficult to show that this model can be represented by the Volterra series (2) with 'diagonal' model coefficients (i.e. $b_{i,j} = 0$ unless $i = j$, $c_{i,j,k} = 0$ unless $i = j = k$, etc.). If the order of the blocks is reversed, i.e. if the static nonlinearity follows the linear dynamics, we obtain the Wiener model^{7,8,10,13} shown in *Figure 2*. As subsequent examples will illustrate, this model is not equivalent to the Hammerstein model. Both of these nonlinear models are special cases of the more general 'sandwich model' considered by Brillinger¹⁴ and Greblicki and Pawlak¹⁰, in which the static nonlinearity is 'sandwiched' between two linear dynamic models. More general 'block-oriented' nonlinear models have been investigated involving both series and parallel connections of static nonlinearities and linear dynamics^{7,8,10}. As a specific example, the Uryson model consists of several

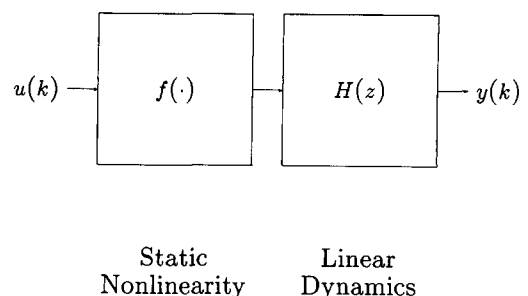


Figure 1 Hammerstein model structure

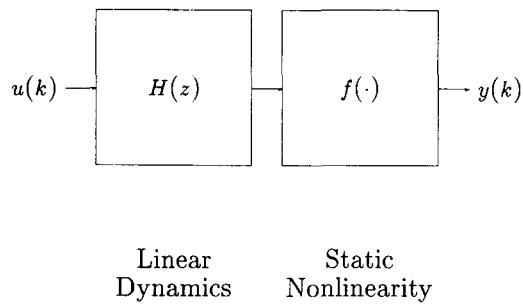


Figure 2 Wiener model structure

Hammerstein models connected in parallel with their outputs summed, but driven by a common input⁷.

The 'autogressive' family of nonlinear dynamic models is represented by the NARMAX (Nonlinear ARMAX) models studied by Billings and Voon¹⁵:

$$y(k) = F(y(k-1), y(k-2), \dots, y(k-p), u(k), u(k-1), \dots, u(k-q), e(k-1), e(k-2), \dots, e(k-r)) + e(k) \quad (4)$$

where $F(\cdot)$ is a nonlinear function of the $p + q + r + 1$ variables indicated. Usually, this nonlinear function is taken to be a low-order polynomial, but if we extend this class to more general functions, the range of nonlinear behaviour we can represent expands correspondingly. In particular, note that most of the nonlinear autoregressive models discussed by Tong⁵ are special cases of the NARMAX model with non-polynomial functions $F(\cdot)$. The advantage of restricting $F(\cdot)$ to multivariable polynomials is that it permits the use of linear regression techniques like stepwise regression for model identification¹⁵. Alternatively, Chen and Tsay¹⁶ consider a different structural restriction that permits the use of nonparametric regression methods (see later for a brief discussion). Specifically, the NAARX (Nonlinear Additive AutoRegressive models with eXogenous inputs) are defined as:

$$y(k) = f_1(y(k-1)) + f_2(y(k-2)) + \dots + f_p(y(k-p)) + g_0(u(k)) + g_1(u(k-1)) + \dots + g_q(u(k-q)) + e(k) \quad (5)$$

Note that the linear ARMAX model with i.i.d. modelling errors discussed above ($r = 0$ in Equation (1)) represents an important special case of the NAARX family. Similarly, Hammerstein models may be obtained by taking $f_i(x) = a_i x$ and $g_j(x) = b_j g_0(x)$ where $g_0(x)$ is the static nonlinearity defining the Hammerstein model. Further, note that the logistic model $y(k) = a y(k-1)(1 - y(k-1))$, perhaps the standard chaotic discrete-time model⁵, is also a special case of the NAARX model.

Artificial neural networks are mathematical models motivated by and loosely modelled on biological neural models proposed originally by McCulloch¹⁷. Probably the most popular artificial neural network is the feedforward network¹⁸, a multi-layer structure that imple-

ments a static nonlinear map $T : R^m \rightarrow R^n$ between an m -dimensional input space and an n -dimensional output space. In particular, if $u \in R^m$ and $y \in R^n$, a two layer feedforward network implements the transformation:

$$y_i = s\left(\sum_{j=1}^m w_{ij} u_j - \theta_i\right) \quad (6)$$

where u_i and y_i are the i th components of the input and output vectors, respectively, $\{w_{ij}\}$ is an $n \times m$ matrix of 'synaptic weights' and $\{\theta_j\}$ is an m -vector of 'bias weights'. The function $s(\cdot)$ is a saturation nonlinearity (typically, $s(x) = 1/(1 + e^{-x})$) called the 'squashing function'. Multi-layer networks are constructed by connecting two-layer networks in cascade and it has been shown¹⁹ that a three-layer network is flexible enough to approximate any continuous transformation T with arbitrary accuracy on compact subsets of R^n and R^m , although Kramer²⁰ and others have argued that more than three layers are desirable for certain applications.

This flexibility may be used in developing 'neural versions' of either block-oriented nonlinear models or NARMAX models. For example, applications of neural network-based Hammerstein models to chemical processes are discussed by Su and McAvoy¹². Similarly, 'Model IV' discussed by Narendra in Parthasarathy²¹ is basically the NARMAX model described above with $F(\cdot)$ implemented as a feedforward network. These networks feed delayed responses from the output layer back to the input layer and are called recurrent feedforward networks. In multi-layer networks, it is possible to feed delayed 'hidden-layer' responses back to the hidden layer itself, introducing dynamics that are entirely internal to the network. Networks of this form were proposed by Elman²² and have been considered for chemical process model applications because of their similarity to nonlinear state-space models. In particular, these models have the general structure:

$$\begin{aligned} \mathbf{x}(k) &= \mathbf{F}(\mathbf{x}(k-1), \mathbf{u}(k)) \\ \mathbf{y}(k) &= \mathbf{G}(\mathbf{x}(k)) \end{aligned} \quad (7)$$

where $\mathbf{F}(\cdot)$ and $\mathbf{G}(\cdot)$ represent vector transformations like those discussed above. One advantage of this formulation is that the general structure parallels that of typical first-principles models, raising the possibility of both simplified model identification and physical model interpretation²³.

Closely related to the feedforward networks just described are networks based on radial basis functions (RBF)²⁴. These networks also implement multivariable nonlinear static mappings $F(\cdot)$, defined by:

$$F(\mathbf{x}) = \sum_{i=1}^N a_i \theta(\|\mathbf{x} - \mathbf{y}_i\|) \quad (8)$$

where $\|\mathbf{x}\|$ is the Euclidean norm of the vector \mathbf{x} , $\theta(\cdot)$ is a scalar function, and $\{a_i\}$ is a set of unknown model coefficients to be determined from the available data.

The function $\phi(z)$ typically exhibits a local minimum or maximum at $z = 0$, so the vectors $\{y_i\}$ are called the centres of the basis functions. As a specific example, Pottmann and Seborg²⁴ consider the function $\phi(z) = z^2 + \beta)^{-1/2}$, which exhibits a maximum at $z = 0$. The advantage of radial basis functions is that, once the width parameter β and the centres $\{y_i\}$ are specified, identification of the model coefficients $\{a_i\}$ reduces to a linear regression problem; the harder part of the identification problem is the initial specification of β and $\{y_i\}$. The use of RBF networks in dynamic modelling is similar to the use of feedforward neural networks: the static mapping $F(\cdot)$ can implement either the static nonlinearity appearing in the NARMAX model or the static nonlinearities appearing in block-oriented models.

The nonlinear models just described represent a very incomplete list, but they are representative of the types that have been considered for chemical process applications. The following sections will attempt to give some useful insights into the general nature of these models, their important differences, and the practical issues involved in using them to model real-world processes.

Model behaviour

To illustrate some of the important differences between linear and nonlinear model behaviour, the following paragraphs briefly consider two inherently nonlinear phenomena: asymmetric responses to symmetric input changes and subharmonic generation in response to periodic input sequences. This second example is not included because subharmonic generation is a problem of great practical significance in chemical process applications, but rather because it clearly illustrates one profound difference between two broad classes of nonlinear systems: 'nonlinear moving average models' and 'nonlinear autoregressive models'. Other inherently nonlinear phenomena include harmonic generation, jump phenomena, synchronization, chaos and a wide variety of amplitude-dependent responses, as discussed by Tong⁵ and Nayfeh and Mook²⁵.

First, consider the second-order Volterra model defined in Equation (3) with $u(k)$ and $y(k)$ representing deviations from a nominal (e.g. steady-state) operating condition so that both positive and negative values make physical sense for both variables. For a linear system, note that if $y(k)$ is the response to an input change $u(k)$, then $-y(k)$ is the response to an input change $-u(k)$. In contrast, the response of the second-order Volterra model (3) to a step input of amplitude A is:

$$y(k) = \begin{cases} 0 & k < 0 \\ A & k = 0, 1 \\ A(1 + Ab) & k \geq 2 \end{cases} \quad (9)$$

Specifically, assume $b > 0$ and note that the 'steady-state gain' inferred from this step response is $K = 1 +$

Ab , which is greater than 1 for steps of positive amplitude and less than 1 for steps of negative amplitude.

Chien and Ogunnaiké²⁶ illustrate similar asymmetric step responses in their analysis of high-purity distillation columns. In particular, changes in reflux ratio or steam flow to the reboiler that move toward lower purity have greater effects on composition than changes that move toward higher purity. Responses of this general type – moves away from a fundamental constraint (e.g. the 100% purity limit) being easier than moves toward it – can be expected in many systems if they are operated close enough to the constraint. An even more pronounced example of asymmetric response is considered in a later section.

The phenomenon of subharmonic generation illustrates the difference between 'inherently autoregressive' nonlinear behaviour and 'inherently moving average' behaviour. Specifically, note that if an input sequence $\{u(k)\}$ is periodic with period L , it satisfies the condition $u(k + L) = u(k)$ for all k . If $y(k) = g(u(k))$ for any static nonlinear function $g(\cdot)$, it follows immediately that $y(k + L) = g(u(k + L)) = g(u(k)) = y(k)$ for all k ; that is, if the input sequence is periodic with period L , the output sequence is also periodic with period L . Since the generation of the '1/nth-subharmonic' corresponds to lengthening the fundamental period of oscillation from L to nL , it follows that static nonlinearities cannot generate subharmonics. Since linear systems also preserve periodicity, it also follows that block-oriented models – Hammerstein, Wiener, 'sandwich', Uryson, etc – cannot generate sub-harmonics. Similar arguments show that Volterra and other 'moving average' nonlinear models cannot generate sub-harmonics, either.

In contrast, Tong shows explicitly that the following threshold autoregressive model can generate sub-harmonics⁵:

$$y(k) = \begin{cases} 2y(k-1) + u(k) & |y(k-1)| \leq 2 \\ u(k) & |y(k-1)| > 2 \end{cases} \quad (10)$$

Specifically, in response to the input sequence $u(k) = (-1)^k$ with period 2, the output sequence has period 6, representing a '1/3rd-subharmonic' of the input. This observation suggests that the key feature required in nonlinear models capable of generating sub-harmonics is recursion: the present output must depend on previous outputs and not just on the past history of the inputs. Applying this conclusion to neural networks, it follows that recurrent networks are necessary to generate sub-harmonics. In particular, it follows that either the 'neural NARMAX' models or the Elman networks discussed above can generate sub-harmonics (note that Equation (10) is a special case of Equation (7)), but the 'neural Hammerstein' and other block-oriented neural models cannot. The same conclusions apply to nonlinear models based on radial basis function networks.

Structure selection

The first step in constructing any empirical process model is the selection of an appropriate model structure (e.g., Volterra, Hammerstein, Wiener, NARMAX etc.). This selection problem may be approached from either of two basic philosophies: exploratory or confirmatory. The exploratory philosophy seeks to extract as much structural information from the available data as possible, while the confirmatory philosophy selects a candidate model structure and asks the data for a confirmation or rejection of its appropriateness. Closely related to these philosophies are the notions of parametric, nonparametric and semi-parametric modelling. In parametric modelling, the structure of the model is fixed, reducing the model identification problem to one of determining a finite set of model parameters for which the model predictions best match the available data. In nonparametric modelling, a particular model structure is sought from the available data. For example, in nonparametric regression analysis²⁷ a model $y(k) = f(x(k))$ is sought without specifying the form of $f(\cdot)$. In semi-parametric modelling, part of the model structure is completely specified while part is only loosely specified. In process applications, we are generally interested in obtaining a parametric model, but nonparametric or semi-parametric procedures can be extremely useful in suggesting model forms for subsequent parametric identification.

These ideas are illustrated in *Figure 3* for the NARMAX modelling problem. In its most general form, this problem involves selection of a functional form for the multivariable nonlinearity $F(\cdot)$, selection of the input variable or variables $\{u(k)\}$, and specification of the order parameters p, q, r in Equation (4). Once we have specified the model inputs and order parameters, it would be possible, in theory to use nonparametric regression techniques to determine the form of $F(\cdot)$, as indicated at the top of *Figure 3*. In practice, however, the data requirements for nonparametric regression in many variables are extreme as discussed by *Hardle*²⁷. Three more practical alternatives are shown in the sec-

ond tier of *Figure 3*: neural NARMAX modelling, polynomial NARMAX modelling and nonparametric NAARX modelling. Other possibilities exist (e.g. RBF networks instead of neural networks), but the three given here are representative. First, the neural NARMAX alternative discussed earlier replaces the nonparametric NARMAX problem with a parametric one, implementing the multivariable static nonlinearity $F(\cdot)$ as a neural network. Next, probably the most popular procedure in practice is the use of polynomial NARMAX models: the function $F(\cdot)$ is taken to be a low-order multivariable polynomial in all of the lagged output, input and model error variables. The number of terms in this polynomial (hence the number of model parameters to be identified) grows rapidly with increasing polynomial model order. The use of stepwise regression techniques has been discussed by various authors^{15,28,29} and it provides a systematic approach to reducing this complexity by retaining only those terms in the polynomial that significantly improve the goodness of fit. Finally, the general procedure for nonparametric regression in many variables advocated by *Hardle*²⁷ is the use of additive models, replacing a single N -dimensional smoothing problem with N one-dimensional smoothing problems. This idea was *Chen and Tsay's* motivation for investigating nonparametric NAARX models¹⁶.

The model identification problem may be simplified further by imposing additional structural restrictions on the NARMAX model, as illustrated in the third tier of *Figure 3*. For example, if the cross-terms in the polynomial NARMAX model are eliminated entirely, the result is a polynomial NAARX model. Alternatively, recall that Hammerstein models are special cases of the NAARX model with $f_i(x) = a_i x$ and $g_j(x) = b_j g_0(x)$ where $g_0(x)$ is the static Hammerstein nonlinearity. Thus, we can adopt a semi-parametric model identification procedure by using nonparametric smoothing to estimate the scalar function $g_0(\cdot)$ and estimating the linear model parameters $\{a_i\}$ and $\{b_j\}$. *Greblicki and Pawlak*⁹ discuss nonparametric identification of this nonlinearity in detail. This model identification problem can be further reduced to a fully parametric procedure like that employed by *Eskinat et al.*¹¹. There, the authors developed Hammerstein models for distillation columns, representing the linear dynamics by a low-order ARMAX model and the static nonlinearity by a low-order polynomial. The authors conclude that the identified Hammerstein models represent the column dynamics better than linear models do, but that model performance degrades unacceptably for high-purity columns. It remains an open question, however, whether this conclusion represents an inherent limitation of the class of Hammerstein models or a limitation of the parametric sub-class considered by the authors. Note that the semi-parametric Hammerstein modelling procedure described above could give useful insight into this question. In particular, note that nonparametric estimation of the static nonlinearity could be used to identify functions that are not well approximated by

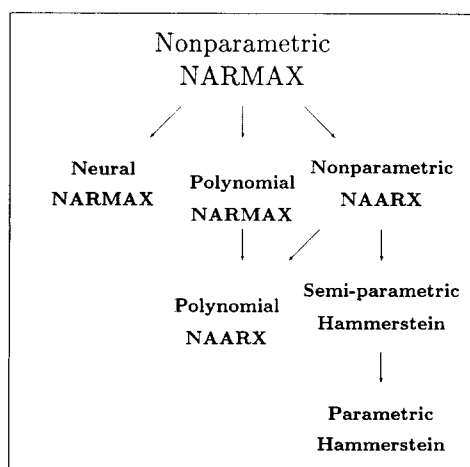


Figure 3 NARMAX identification procedures

low-order polynomials (e.g., saturation nonlinearities). As a specific example of this difficulty, Pottmann *et al.*³⁰ consider a pH control problem in which the pH curve must be approximated by a piecewise-polynomial function to achieve an adequate fit; this approach is closely related to the use of spline functions for nonparametric regression²⁷.

In taking a regression approach to model structure selection, the primary focus is on goodness of fit. In practice, it is important to note that other criteria may be of equal or greater importance. For example, an extremely important related issue is the sensitivity of the model prediction errors to changes in the problem formulation. This issue has two aspects: the sensitivity of the algorithms used for model identification to errors in the data, and the region of validity of the model \mathcal{M} as an approximation of some real-world process \mathcal{P} . The issue of sensitivity to errors in the data is the domain of robust statistics^{31,32} and will be discussed further in a later section. Note also that this sensitivity to data quality will generally depend on both the model structure chosen and the input variables selected for inclusion in the model. In particular, note that the stepwise regression procedure discussed above for polynomial NARMAX model development considers each possible model term for inclusion or exclusion on the basis of goodness of fit alone; Leger and Altman³³ give a detailed discussion of the influence of such variable selections on the robustness of the resulting model fit. Specifically, they note that the terms included in a multivariable model can profoundly affect the sensitivity of the model predictions to outliers in the data.

The region of validity of the model \mathcal{M} is an extremely important topic, to be discussed further later. One way of describing this region of validity is in terms of a set \mathcal{U} of input sequences $\{u(k)\}$ for which the model predictions $\{\hat{y}(k)\}$ are 'sufficiently close' to their 'true' values $\{y(k)\}$. An important extension of this concept is to consider the qualitative behaviour of the model \mathcal{M} over the set \mathcal{U} : does the model exhibit multiple steady states, chaotic regimes, etc? Note that these considerations are both more difficult and more important for nonlinear models than for linear models because the qualitative behaviour of nonlinear models can depend very strongly on the specific input sequence driving them. The example discussed later illustrates this point in detail.

Finally, note that other practical considerations may dictate that certain model structures are preferred over others, even at the expense of goodness of fit or some of the other adequacy criteria just discussed. In particular, differences in overall qualitative behaviour between different models with comparable goodness of fit can be profound, as a later section will demonstrate. For example, it has been argued³⁴ that non-minimum phase dynamics in linear models limit achievable control system performance because they do not exhibit well-defined (e.g., stable and causal) inverse. Analogous difficulties can arise in nonlinear models, where the corresponding notion is that of unstable zero dynamics³⁵. A

different, though similar, difficulty can arise in block-oriented models if the static nonlinearity $f(\cdot)$ fails to have an inverse. Specifically, note that a continuous function $f(\cdot)$ defined on a compact support set is invertible if and only if it is strictly monotonic³⁶. In a later section several models are considered for the same system, including both a non-invertible Wiener model and a NARMAX model (Model 1) for which the inverse is given explicitly. The point is, even though both models may yield comparable goodness of fit to the available input/output data, control strategies developed from these models are apt to be quite different in both structure and performance.

Higher-order statistics

When a linear system is excited by a Gaussian input sequence $\{u(k)\}$, the resulting output sequence $\{y(k)\}$ is also Gaussian. When a nonlinear system is excited by a Gaussian input sequence, the resulting output is generally non-Gaussian, establishing a strong connection between non-Gaussian statistics and nonlinear systems. As a practical example of this connection, Brillinger¹⁴ considered the problem of identifying the 'sandwich model' discussed earlier. He showed that the static nonlinearity and both linear dynamic subsystems can be identified from the power spectrum of the input sequence, the cross-spectrum of the input and output sequences, and a cross-bispectrum between these sequences. Note that the bispectrum and other higher-order statistics are identically zero for Gaussian sequences⁶ and are therefore quantitative measures of 'non-Gaussianity'. Brillinger's results assume a Gaussian input sequence $\{u(k)\}$, are nonparametric, and require the static nonlinearity $G(\cdot)$ to be asymmetric (i.e. neither even nor odd). Qualitatively similar results for low-order Volterra model identification have been derived by Koh and Powers³⁷ and Pearson *et al.*³⁸. In particular, the matrix B of second-order Volterra coefficients $\{b_{ij}\}$ is given by:

$$B = \frac{1}{2} R_{uu}^{-1} T_{yu} R_{uu}^{-1} \quad (11)$$

where R_{uu} is the autocorrelation matrix for the Gaussian input sequence $\{u(k)\}$ and T_{yu} is a cross-bicorrelation matrix between the input and output sequences.

Brillinger's 'sandwich model' identification results¹⁴ are based on cumulant expressions involving (generally multivariable) static nonlinearities and multivariate Gaussian random variables. Applying these results to the Wiener model shown in *Figure 2* yields the following cross-correlation expression:

$$R_{uy}(\tau) = \left[\frac{E\{xf(x)\}}{E\{x^2\}} \right] R_{ux}(\tau) \quad (12)$$

where $\{x(k)\}$ is the output sequence generated by the linear subsystem $H(z)$ in response to the input sequence

$\{u(k)\}$. In the special case where $H(z) = 1$, $x(k) = u(k)$ and Equation (12) reduces to Bussgang's theorem³⁹. Taking the Fourier transform of Equation (12) yields the nonparametric 'equivalent linear system identification' expression:

$$\begin{aligned} \hat{H}(\omega) &= \frac{S_{uy}(\omega)}{S_{uu}(\omega)} \\ &= \left[\frac{E\{xf(x)\}}{E\{x^2\}} \right] H(\omega) \end{aligned} \quad (13)$$

Here, use has been made of the fact that $S_{ux}(\omega) = H(\omega)S_{uu}(\omega)$. Note that if $f(x) = ax$, this expression reduces to the correct linear result $\hat{H}(\omega) = aH(\omega)$. However, if $f(x) = ax + \theta(x)$ where $\theta(x)$ is any even function (i.e. $\theta(-x) = \theta(x)$), it follows that $E\{x\theta(x)\} = 0$, so again $\hat{H}(\omega) = aH(\omega)$. Thus, nonparametric system identification with Gaussian inputs is blind to arbitrary even perturbations $\theta(x)$. To detect the presence of this nonlinearity, it is necessary to consider higher-order statistics. Conversely, if $f(x) = ax + bx^3$, the 'equivalent linear model' identified is $\hat{H}(\omega) = (a + 3b\sigma^2)H(\omega)$, permitting the ratio b/a to be estimated from the dependence of $\hat{H}(\omega)$ on the input signal intensity σ^2 .

The second-order Volterra model considered earlier gives another illustration of the need for higher-order statistics in nonlinear system identification. Specifically if the model in Equation (3) is excited by a zero-mean, i.i.d. Gaussian sequence with variance σ^2 , the model response will also be a zero mean sequence that is uncorrelated, i.e.:

$$R_{yy}(k) = R_{uu}(k) = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \quad (14)$$

For Gaussian processes, lack of correlation implies statistical independence, but this result does not hold for non-Gaussian processes. As in the Wiener model example just considered, to see the nonlinear nature of this system it is necessary to consider higher-order statistics. Here, however, the effects are extremely subtle because the skewness $\gamma = E\{x^3\}/\sigma^3$ is also zero. A complete dynamic third-moment characterization reveals that the third cumulant $c_3(1, 2) = E\{y(k)y(k-1)y(k-2)\}$ is nonzero, reflecting the 'dynamically asymmetric' non-Gaussian nature of the sequence $\{y(k)\}$. In fact, the nonlinear model structure correctly follows from Equation (11) since this system is a second-order Volterra model driven by a Gaussian input sequence. Equivalently, this model may be identified by nonparametric frequency-domain techniques based on cross-bispectra like those considered by Brillinger¹⁴.

Outliers and robust statistics

Unfortunately, length considerations do not permit a detailed discussion of robustness here, but this topic is extremely important in practice, as the following exam-

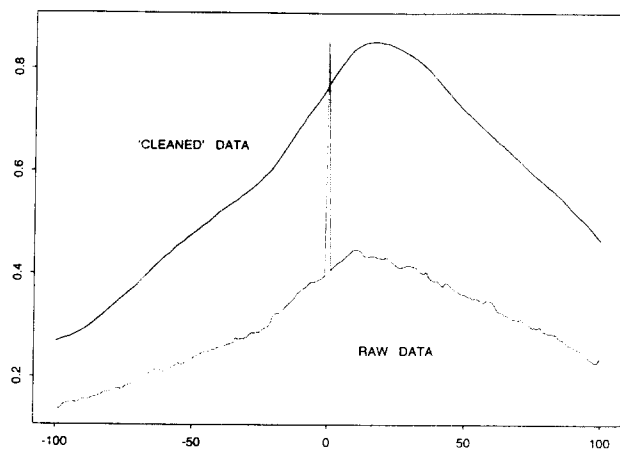


Figure 4 Two cross-covariance functions

ple illustrates. Specifically, Figure 4 shows two normalized cross-covariance functions, estimated from on-line physical property measurements made at the inlet and outlet of a storage tank as part of an attempt to characterize the flow through the tank. In both cases, a standard moment-based estimator was applied to data records of length $N = 1024$. The difference between the two plots reflects the influence of a single pair of data values. Specifically, the function with the sharp peak at $\tau = 0$ was computed for the original dataset, in which a single missing data record resulted in a pair of coincident spurious values about 30 standard deviations below the median value. The other cross-covariance function was computed for the same dataset, but with these spurious data values replaced by the median value for each sequence. While this pair of points represents a 'gross outlier', it amounts to only 0.1% of the dataset, yet completely dominates the computed results.

A useful approach to this problem in linear system identification is to first 'clean' the data using a procedure based on robust statistics like that outlined by Martin and Thomson⁴⁰. This approach effectively identifies dynamic outliers in the data and replaces them with estimates based on neighbouring data points. An inherent working assumption in this approach is that the underlying 'clean' data are well approximated by a Gaussian random variable, so the resulting 'cleaned' data sequence is made 'more Gaussian' than the original. The difficulty in extending this philosophy to nonlinear system identification is that the 'clean' nonlinear model response to a Gaussian input sequence is non-Gaussian. Further, it is precisely the 'deviations from Gaussianity' of the output sequence that contain the information needed for model identification. Thus, the task of distinguishing 'outliers' from 'legitimate nonlinear model responses' is difficult and represents an important area for continued research activity.

It is easy to show that outlier problems generally become more severe with increasing moment order. As a specific example, consider the effects of a single 'outlier' of magnitude $+8\sigma$ contaminating a sequence of 1024 i.i.d. Gaussian random variables with mean zero and variance σ^2 . Estimated values for the mean, vari-

Table 1 Effects of +8s outlier on estimated moments

Outlier	Average	Variance	Skewness	Kurtosis
No	-0.006	1.000	-0.002	0.553
Yes	+0.002	1.065	+0.454	3.894

ance, skewness and kurtosis⁴¹ with and without this outlier are given in *Table 1*. The presence of this one 'bad' data point (i.e. a large error in 0.1% of the data) has negligible effect on the estimated mean, increases the variance slightly, but inflates both the estimated skewness and kurtosis enormously: note that both of these quantities should be zero. Analogous outlier sensitivities can be expected for estimated bispectra, trispectra and other higher-order statistics. Indeed, the sensitivity is worse for these dynamic statistics in the sense that the effects of the outlier can be distributed unpredictably over the entire frequency space on which these higher-order spectra are defined.

Identifiability and input sequence design

In theory, linear systems may be identified from their responses to a wide variety of input sequences: impulses, steps, pseudo-random binary sequences (PRBS), i.i.d. Gaussian sequences (i.e. Gaussian 'white noise') or correlated Gaussian sequences (e.g. Gaussian autoregressive processes). In practice, some input sequences are more effective than others, but input sequence design procedures for linear system identification typically only specify the power spectrum of the sequence³. In particular, note that the random telegraph wave (a particular PRBS discussed in Papoulis⁴²) and the first-order Gaussian autoregressive process have the same correlation structure (and thus the same power spectrum). Consequently, these sequences would be equally effective in linear system identification. The choice between these (or other) sequences will ultimately depend on additional practical considerations, as discussed in Ljung³.

In contrast, specification of the power spectrum of the input sequence is not sufficient for nonlinear system identification. Specifically, it was noted by Pearson *et al.*³⁸ that no PRBS input can be used to identify the diagonal elements of the B matrix in second-order Volterra models. However, the B matrix can be identified using the Gaussian autocorrelation sequence with the same autocorrelation matrix; this result is given explicitly by Equation (11). In general, PRBS inputs are a poor choice for nonlinear system identification because they do not exercise the process over a wide enough range of input values. Note, however, that simply expanding the range of input values is not sufficient, either. For example, note that the impulse response of the second-order Volterra model defined in Equation (3) is $y(k) = u(k) = \alpha\delta(k)$, for arbitrary α . Thus, the impulse response of this system gives no indication of its nonlinearity, regardless of the input amplitude α or

the strength of the nonlinearity b . More generally, note that the impulse response of a Volterra model depends only on the diagonal model coefficients: $y(k) = y_0 + a_k + b_{k,k} + c_{k,k,k} + \dots$.

From a theoretical perspective, i.i.d. input sequences are often very effective for nonlinear system identification^{38,43,44}. Because so many results are available for Gaussian stochastic processes, i.i.d. Gaussian sequences are particularly popular for both developing and investigating nonlinear system identification algorithms. In some cases, however, better performance may be achieved by using non-Gaussian sequences tailored to the model structure under consideration. For example, it was shown by Pearson *et al.*³⁸ that leptokurtic distributions⁴¹ are desirable for identifying the diagonal elements of the B matrix in second-order Volterra models, while platykurtic distributions⁴¹ are desirable for identifying the off-diagonal elements. As in linear modelling, the dependence structure of the input sequence is important as well, although much less is known about specification of 'good' dependence structures for nonlinear system identification.

As in the case of linear system identification, identifiability also depends on the structure of the model sought. For example, note that the nonlinear static map implemented by a feedforward neural network is invariant under the permutation of processing units in the hidden layers. Thus, nonlinear models based on feedforward networks are only identifiable within the limits of this permutation invariance. If the nonlinear 'squashing functions' are removed from a feedforward network, this permutation invariance expands to an invariance under a much broader class of transformations. In particular, the (linear) map implemented by such a network is invariant under post-multiplication of the weights in layer i by an arbitrary nonsingular matrix W and pre-multiplication of the weights in layer $i + 1$ by its inverse W^{-1} . This invariance is analogous to the invariance of state-space models under the appropriate similarity transformations of the matrices defining the model. Specifically, consider the linear state-space model:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}u_k \\ y_k &= \mathbf{C}\mathbf{x}_k \end{aligned} \quad (15)$$

and note that the input/output response of this model is invariant under the transformation $\mathbf{A} \rightarrow \mathbf{W}^{-1}\mathbf{A}\mathbf{W}$, $\mathbf{B} \rightarrow \mathbf{W}^{-1}\mathbf{B}$, and $\mathbf{C} \rightarrow \mathbf{C}\mathbf{W}$. In ARMAX models, this invariance corresponds to non-uniqueness under pole-zero cancellation. For example, consider the standard linear 'first order plus deadtime' model:

$$y(k) = ay(k-1) + bu(k-d) \quad (16)$$

Using this equation to represent $y(k-1)$ in terms of $y(k-2)$ and $u(k-d-1)$ and substituting the results back into Equation (16) then yields the fully equivalent second-order model:

$$y(k) = a^2 y(k-2) + bu(k-d) + abu(k-d-1) \quad (17)$$

The same non-uniqueness arises in NARMAX models: simply use the original model to derive expressions for the variables $y(k-j)$ required in the model and substitute them back into the original equation. Since this result will always lead to a higher-order model, Occam's razor can be invoked to select the simpler, original model. In practice, however, since these models are only approximations of a more complicated 'truth' we will generally not identify either of these equivalent models, but rather a model that is 'near' one or the other, depending on our initial choice of model order parameters, the available data, the details of the identification algorithm used etc. Further, as the following example will illustrate, the same general qualitative behaviour can often be obtained from very different model structures: in such cases, it may not be obvious which model is 'simpler', or whether there are any 'near equivalences' between these models.

Example

Many of the nonlinear modelling issues discussed in the preceding sections are illustrated by the following example. A single irreversible, exothermic reaction, $A \rightarrow B$, is assumed to occur in a CSTR. Detailed descriptions of this problem are given by Nahas *et al.*⁴⁵ and Pottmann and Seborg²⁴; the dynamics are described by two coupled nonlinear ordinary differential equations relating the concentration c_A of species A in the reactor effluent to the reactor temperature T and the flow rate q_c of coolant to the reactor cooling jacket. Like the high-purity distillation column example discussed earlier, this system exhibits a pronounced asymmetry in its response to changes in the manipulated variable q_c . Specifically, the change in c_A for a 10% step decrease in q_c is a monotonic decrease that looks very much like a first-order linear system response. In contrast, the change in c_A for a 10% step increase in q_c is a decaying oscillation, very much like the response of an underdamped second-order linear system. The following paragraphs will not attempt to develop a specific empirical model for this CSTR example, but they will explore the model structure selection problem in some detail. In particular, this section will focus on the question: which classes of models are rich enough to exhibit the qualitative behaviour just described for the CSTR? The input sequence considered here will be:

$$u(k) = \begin{cases} A & k \geq 0 \\ 0 & k < 0 \end{cases} \quad (18)$$

where A may be either positive or negative.

First, consider the class of truncated Volterra models. If each infinite sum is truncated to M terms, the model responses reaches steady-state after M time steps, i.e.:

$$y(k) = y_0 + A \sum_{i=0}^M a_i + A^2 \sum_{i=0}^M \sum_{j=0}^M b_{i,j} + \dots \quad (19)$$

for $k > M$. Thus, if we are to approximate the CSTR response with a truncated Volterra model, we must take M large enough to allow the oscillatory positive step response to approach its steady-state value. Generally, this value of M will be large, resulting in an unacceptably large number of model parameters to identify. For example, if it takes $M = 30$ samples for the oscillation to decay, even restricting consideration to second-order Volterra models will require identification of approximately 450 quadratic parameters $b_{i,j}$.

As noted earlier, block-oriented models are Volterra models with restricted structures that can be characterized by many fewer parameters. For example, it was noted that the Hammerstein model is a Volterra model in which only the 'diagonal' model coefficients are non-zero. Applied to the example just discussed, this restriction reduces the number of quadratic model parameters from about 450 to 30; in practice, Hammerstein models may be approximated by many fewer parameters than this by exploiting the block structure of the model rather than using the Volterra representation. Unfortunately, the Hammerstein model is incapable of describing the qualitative behaviour exhibited by the CSTR example considered here. In particular, note that the static nonlinearity $f(\cdot)$ preceding the linear dynamics in the Hammerstein model changes a step of amplitude A into a step of amplitude $f(A)$. Thus, if the qualitative nature of the linear subsystem's step response is inherently monotone, the overall Hammerstein model's step response will also be monotone. Conversely, if the linear system exhibits an oscillatory step response, the Hammerstein will also exhibit an oscillatory step response.

In contrast, the Wiener model is sufficiently flexible to exhibit step responses like the CSTR example. In particular, consider the Wiener model with first-order linear dynamics followed by the static nonlinearity $f(x) = g(h(x))$ where:

$$h(x) = -\text{sgn}(x) \ln(1 - |x|) \\ g(x) = \begin{cases} 1 - e^{-x/\tau} \sin(2\pi f x) & x \geq 0 \\ 1 - e^{x/\tau} & x < 0 \end{cases} \quad (20)$$

Here, $\text{sgn}(x)$ is the algebraic sign function, defined as +1 if $x > 0$, -1 if $x < 0$ and 0 if $x = 0$. This nonlinearity was chosen so that $y(t) = 1 - e^{-t/\tau} \sin(2\pi f t + \theta)$ for a step input of amplitude $A = +1$ and $y(t) = e^{-t/\tau} - 1$ for a step input of amplitude $A = -1$. The response for intermediate amplitude steps is shown in Figure 5. For comparison, the step responses for the corresponding Hammerstein model - i.e. the same static nonlinearity followed by first-order linear dynamics - are shown in Figure 6. As noted above, the step response of the Hammerstein model exhibits the qualitative character of the first-order linear subsystem, with appropriately trans-

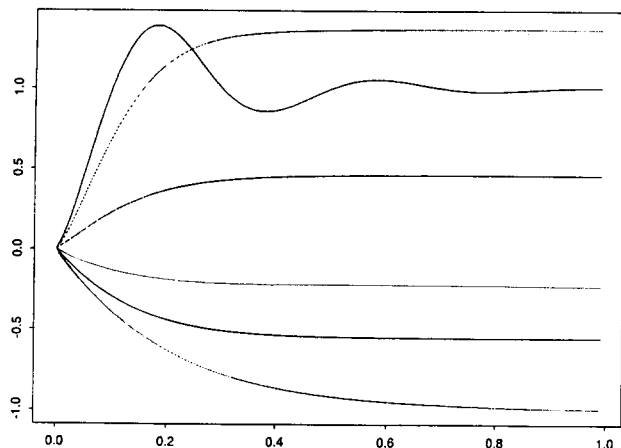


Figure 5 Wiener model step responses

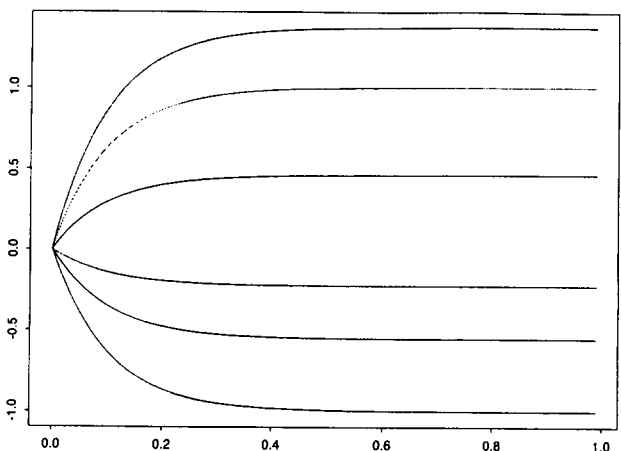


Figure 6 Hammerstein model step responses

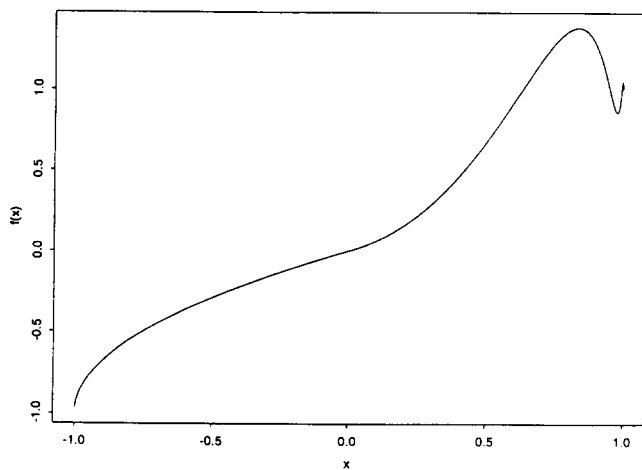


Figure 7 Static nonlinearity for block-oriented models

formed input amplitudes. Also, note that the steady-state behaviour of the Wiener and Hammerstein model is identical because both models reduce to the static nonlinearity $f(\cdot)$ at steady-state. Thus, the distinction between Wiener and Hammerstein models lies entirely in their transient behaviour.

The static nonlinearity $f(\cdot)$ defined in Equation (20) is plotted in Figure 7. Several points are worth noting. First, this function is highly asymmetric, reflecting the asymmetry of the Wiener system response. Further, note that this nonlinearity is not monotonic. As noted

earlier, it therefore follows that $f(\cdot)$ is not invertible, complicating the task of inverse-based control system design, and probably limiting achievable control system performance. In addition, note that the behaviour of $f(\cdot)$ implies a non-monotonic relationship between the input amplitude A and the steady-state response of the model. Specifically, note that the steady-state value of the step response for $A = +0.80$ is greater than the steady-state value for $A = +1.00$, while this behaviour is somewhat counter-intuitive, it does emphasize another point made earlier: it is important to consider the behaviour of any nonlinear model over the whole set of possible inputs \mathbb{U} , and not just for a few selected sequences (e.g. $\pm 10\%$ flow rate changes in this example). Both extrapolations beyond the original input sequence range and interpolations between selected inputs can behave quite unexpectedly.

One final point to note regarding the Wiener model is that the oscillatory character of positive step responses is described by the behaviour of $f(x)$ as $x \rightarrow +1$. In particular, note that $f(x)$ oscillates with increasing frequency as $x \rightarrow +1$, analogous to the behaviour of the 'topologist's sine curve' $\sin(1/x)$ as $x \rightarrow 0^{46}$. Alternatively, we may regard this function as fractal in nature, since the qualitative behaviour remains the same if we look at the function with successively finer resolution near $x = +1$. Consequently, this function would be quite difficult to identify from input/output data using either parametric or nonparametric procedures.

For a given initial steady state, the description of the CSTR behaviour given above – 'second-order for positive changes, first-order for negative changes' – actually defines a Uryson model. Specifically, define the two functions:

$$\theta_+(z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases} \quad \text{and} \quad \theta_-(z) = \begin{cases} 0 & z \geq 0 \\ z & z < 0 \end{cases} \quad (21)$$

Using these nonlinear functions, define the two Hammerstein models:

$$\begin{aligned} y_+(k) &= a_1^+ y_+(k-1) + a_2^+ y_+(k-2) + b^+ \theta_+(u(k-1)) \\ y_-(k) &= a_1^- y_-(k-1) + b^- \theta_-(u(k-1)) \end{aligned} \quad (22)$$

and combine these two 'Hammerstein channels' into the Uryson model response:

$$y(k) = y_+(k) + y_-(k) \quad (23)$$

For properly chosen values of the parameters a_1^+ , a_2^+ , a_1^- , b^+ , and b^- , we may obtain the desired qualitative behaviour from this model. Also, note that we may represent the autoregressive dynamics described in Equation (22) for the positive and negative channels of the Uryson model by equivalent infinite-order moving-average models, excited by $\theta_+(\theta(k-1))$ and $\theta_-(\theta(k-1))$, respectively. These two models may be combined into a single infinite-order nonlinear moving average model with a

very special structure⁴⁷. One consequence of this special structure is that this Uryson model exhibits an interesting 'nearly linear' scaling property. Specifically, if $\{y(k)\}$ is the response to an input sequence $\{u(k)\}$, then the response to an input sequence $\{\lambda u(k)\}$ is $\{\lambda y(k)\}$ for any $\lambda > 0$. This result follows from the fact that both of the functions $\theta_-(x)$ and $\theta_+(x)$ are homogeneous⁴⁸ with respect to positive scaling constants λ : $\theta_{\pm}(\lambda x) = \lambda \theta_{\pm}(x)$ for $\lambda \geq 0$. Thus, unlike NARMAX Models 4 and 5 discussed below, the qualitative behaviour of the step response is not a function of the input step magnitude $|A|$, only its sign.

In contrast to the Wiener model considered above, the Uryson model is invertible, although the inverse is a little unusual. Specifically, define the following quantities:

$$\begin{aligned} u_-(k-1) &= [y(k) - a^- y(k-1)] / b^- \\ u_+(k-1) &= [y(k) - a^+ y(k-1) - a_2^+ y(k-2)] / b^+ \end{aligned} \quad (24)$$

The input $u(k-1)$ required to drive $y(k-1)$ and $y(k-2)$ to $y(k)$ will be:

$$u(k-1) = \begin{cases} u_-(k-1) & \text{if } u_-(k-1) \leq 0 \\ u_+(k-1) & \text{if } u_+(k-1) \geq 0 \end{cases} \quad (25)$$

It is important to note that it is the complete separation of this Uryson model into a pair of 'non-overlapping channels' for positive and negative inputs that makes this inversion possible. In more general cases, it is not obvious under what circumstances a well-defined inverse will exist; similarly, questions have been raised regarding identifiability of Uryson models¹⁰. Finally, note that the Uryson model considered here does not have a Volterra series representation, since the nonlinearities $\theta_{\pm}(x)$ are not analytic functions.

In general, NARMAX models appear to be a more convenient choice for representing behaviour like the CSTR step response, as the following five examples illustrate:

1. $y(k) = ay(k-1) + bu(k-1) + cu(k-1)y(k-1)$
2. $y(k) = a|y(k-1)| + bu(k-1)$
3. $y(k) = ay(k-1) + bu(k-1) + c\theta_+[y(k-2)]$
4. $y(k) = dy^2(k-1) + bu(k-1)$
5. $y(k) = ay(k-1) + bu(k-1) + cu(k-1)y(k-1) + dy^2(k-1)$

Conversely, these examples also demonstrate that the general qualitative behaviour of nonlinear models with 'similar' step responses can be radically different. This observation emphasizes a point made earlier: goodness of fit alone is not a sufficient criterion for 'model adequacy'.

Model 1 is a bilinear generalization of the standard first-order linear model (see Tong⁵ for a discussion of bilinear models). For step inputs, this nonlinear model reduces to the first-order linear model with autoregres-

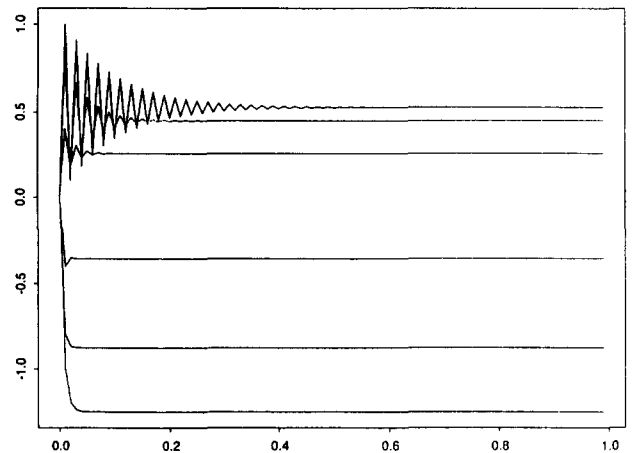


Figure 8 Model 1 step responses

sive coefficient $a + cA$. This qualitative behaviour of this step response can thus be determined from linear system theory. In particular, this model will exhibit a stable, monotonic response if $0 < a + cA < 1$, a stable, damped oscillatory response if $-1 < a + cA < 0$, and unstable step responses outside these ranges. Step responses for A in the range $-1 < A < +1$ are shown in Figure 8 for $a = -0.35$, $b = 1.0$, $c = -0.55$. While the exact details of these step responses are not identical to those of the Wiener model, the qualitative behaviour is the same: a step of amplitude $A = +1$ yields a decaying oscillation, while a step of amplitude $A = -1$ yields a monotonic response. Unlike the Wiener model, however, the steady-state response varies monotonically with A , a more intuitively pleasing behaviour than the non-monotonic variation exhibited by the Wiener model. Also unlike the Wiener model, note that Model 1 is invertible. In particular, the control input $u(k-1)$ necessary to drive the output of Model 1 from $y(k-1)$ to $y(k)$ is:

$$u(k-1) = \frac{y(k) - ay(k-1)}{b + cy(k-1)} \quad (26)$$

provided $b + cy(k-1) \neq 0$. Finally, it is important to note that the identification of this model is much easier than the Wiener model identification problem just discussed. Specifically, note that Model 1 is a parametric model, linear in the three model parameters a , b and c .

Models 2 and 3 are both NAARX models and also threshold autoregressive models⁵, similar in structure to Equation (10). In addition, these models also both exhibit the same positive homogeneous behaviour as the Uryson model discussed above. This behaviour is evident in the step responses shown in Figures 9 and 10 for these two models. Here, the parameters are $a = -0.70$, $b = 1$ for Model 2 and $a = 0.60$, $b = 1$, $c = -0.80$ for Model 3. Much more can be said about the qualitative behaviour of these models because of their simple structure. In particular, note that if the response of Model 2 to an input sequence $\{u(k)\}$ is uniformly positive, the absolute value function has no effect and the model reduces to Equation (16) with delay $d = 1$. In contrast,

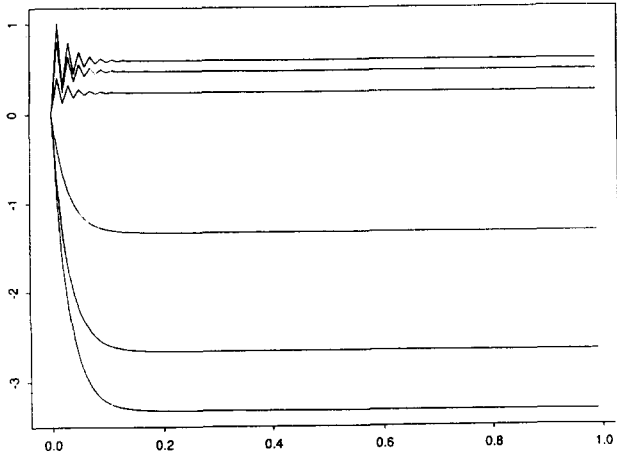


Figure 9 Model 2 step responses

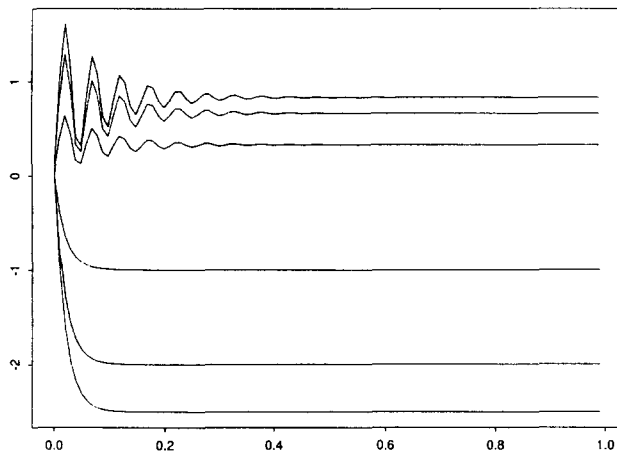


Figure 10 Model 3 step responses

if the response to an input sequence $\{u(k)\}$ is uniformly negative, the absolute value function may be replaced by a negative sign, reversing the sign of the autoregressive coefficient a . These observations describe the step responses shown in Figure 9: for $A > 0$, the model shown is effectively Equation (16) with $a = -0.70$, for $A < 0$, it is effectively the same model with $a = +0.70$.

If the response to $\{u(k)\}$ can be both positive and negative, more complicated behaviour is possible, as illustrated in Figure 11. This plot shows the response of

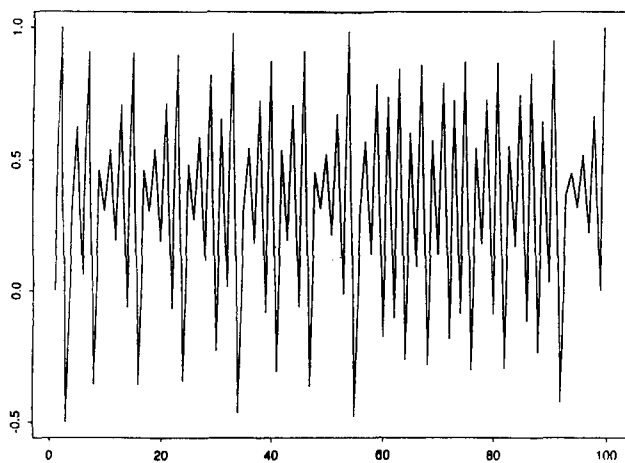


Figure 11 Chaotic step responses for Model 2

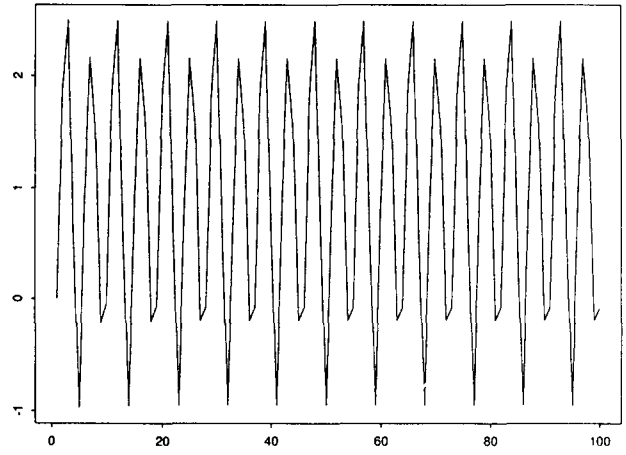


Figure 12 Sustained oscillation for Model 3

Model 2 with $a = -1.5$, $b = 1$ to a positive step input. This response is chaotic and, as noted above, scales linearly with the step amplitude A for any $A > 0$. In contrast, for $A < 0$, the step response is always negative, so the model behaves like a first-order linear model with autoregressive coefficient $a = +1.5$; this response is unstable, growing exponentially with time. Interestingly, chaotic step responses have not yet been observed in Model 3, although sustained oscillations have, as shown in Figure 12. There, the positive step response is shown for Model 3 with parameters $a = 0.35$, $b = 1$, $c = -1.2$. Finally, note that both these models exhibit unique steady-states. For Model 2, if $-1 < a < 1$, this steady state value is:

$$y_s = \begin{cases} bA/(1-a) & bA > 0 \\ bA/(1+a) & bA < 0 \end{cases} \quad (27)$$

while for Model 3, the steady state response is:

$$y_s = \begin{cases} bA/(1-a-c) & bA > 0, a+c < 1 \\ bA/(1-a) & bA < 0, a < 1 \end{cases} \quad (28)$$

Model 4 replaces the nonsmooth absolute value non-linearity in Model 2 with a quadratic term, yielding a polynomial NAARX model. The step responses for this model are shown in Figure 13 for $b = 1$, $d = -0.24$ and

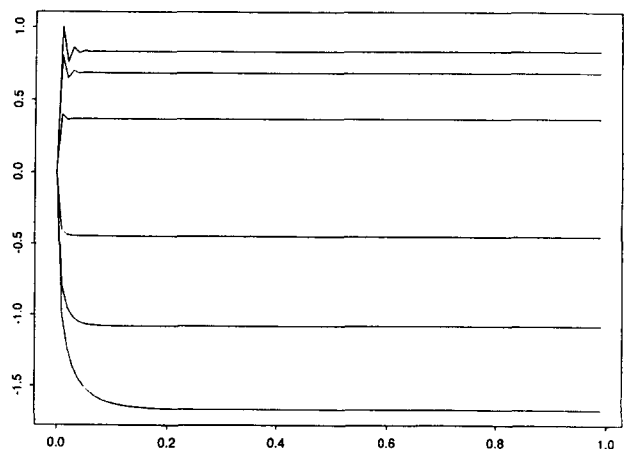


Figure 13 Model 4 step responses

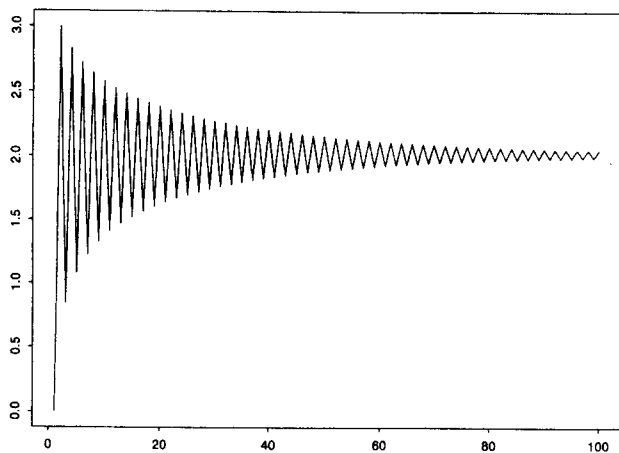


Figure 14 Model 4 step responses, $A = +3.00$

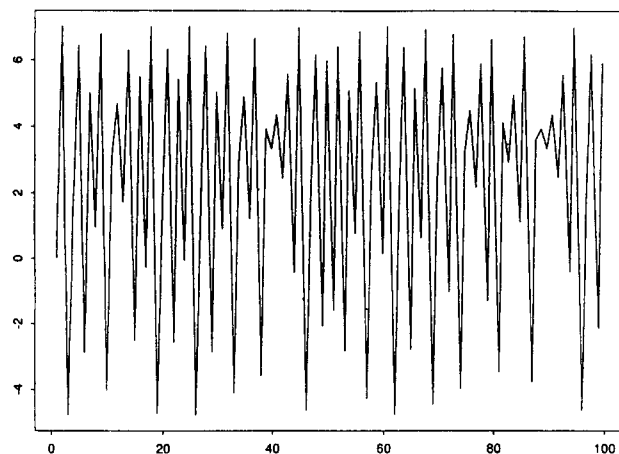


Figure 15 Model 4 step responses, $A = +7.00$

the values of A between -1 and $+1$. These responses are qualitatively similar to those shown in Figure 9 for Model 2. Profound differences exist between these models, however, despite the apparent simplicity of substituting a quadratic nonlinearity for an absolute value. First, note that Model 4 does not share the ‘nearly linear’ scaling property of Model 2: the qualitative dependence of the step response on the step amplitude A is significant as shown in Figures 14 and 15. Specifically, note that the step response for $A = +1.00$ shown in Figure 13 is only slightly oscillatory, while the step response for $A = +3.00$ shown in Figure 14 is highly oscillatory. Increasing the amplitude further to $A = +7.00$ yields the chaotic step response shown in Figure 15. Finally, note that if the step amplitude is made too large – either positive or negative – Model 4 becomes unstable. This general behaviour further illustrates yet another point made earlier: nonlinear models need to be characterized over the entire set \mathcal{U} of probable input sequences and not simply for a few ‘representative’ elements of \mathcal{U} .

Unlike Models 1, 2, and 3, Model 4 exhibits multiple steady states. In particular, substituting $y(k) = y(k-1) = y_s$ into Model 4 yields a quadratic equation for y_s , with the two roots:

$$y_s = \frac{1 \pm \sqrt{1 - 4dbA}}{2d} \quad (29)$$

Further, note that if $1 - 4dbA < 0$, this model exhibits no real steady states, suggesting oscillatory, chaotic or unstable behaviour.

Model 5 is a simple polynomial NARMAX model that includes both Models 1 and 4 as special cases. For arbitrary parameters a , b , c and d , this model will generally resemble Model 4 in its overall qualitative behaviour. For example, unless $d = 0$, this model will exhibit two steady states like Model 4. Further, the qualitative behaviour of this model will depend on the amplitude of the input sequence, like Models 1 and 4 and unlike Models 2 and 3. Typical behaviour for the positive step response would be a monotonically increasing response for small amplitude steps, damped oscillations at larger amplitudes, sustained oscillations and chaotic behaviour at still larger amplitudes, and instability at sufficiently large amplitudes.

Summary

This paper has attempted to summarize a number of the key issues in nonlinear input/output modelling, comparing and contrasting this problem with its linear counterpart where possible. Consequently, the objective of the paper has been more to raise questions than to answer them and, indeed, some of these questions appear to be very difficult to answer, even in general terms. Certain general conclusions may be drawn, however. First, the example discussed in the previous section emphasizes the point made earlier that goodness of fit alone is not a sufficient criterion for ‘model adequacy’. In particular, other considerations like the sensitivity of the identified model to errors in the data may be equally important and can be profoundly influenced by general choices of model structure, order parameters and terms (e.g. input variables) included in or excluded from the model. In addition, the results of the previous section clearly demonstrates that the overall qualitative behaviour of different models with comparable goodness of fit for a few selected inputs can be radically different: some will exhibit chaotic responses or multiple steady states while others will not, the qualitative behaviour of the responses of some models will change dramatically with the magnitude of the input sequence while others will not, etc.

As noted earlier and illustrated in the previous two sections, specification of a reasonably ‘complete’ set \mathcal{U} of input sequences $\{u(k)\}$ is important in nonlinear system identification. In particular, the qualitative nature of a nonlinear model’s behaviour can depend very strongly on its input sequence, in contrast to the behaviour of linear models. In particular, note that if $\{u_1(k)\}$ and $\{u_2(k)\}$ are two input sequences to a linear model and $\{y_1(k)\}$ and $\{y_2(k)\}$ are the corresponding output sequences, the response to any ‘intermediate’ sequence $\tilde{u}(k) = \epsilon u_1(k) + (1 - \epsilon)u_2(k)$ for $0 < \epsilon < 1$ is simply $\epsilon y_1(k) + (1 - \epsilon)y_2(k)$. Such ‘interpolatory’ behaviour may or may not hold for nonlinear models, as the NARMAX models considered in the previous section demonstrate.

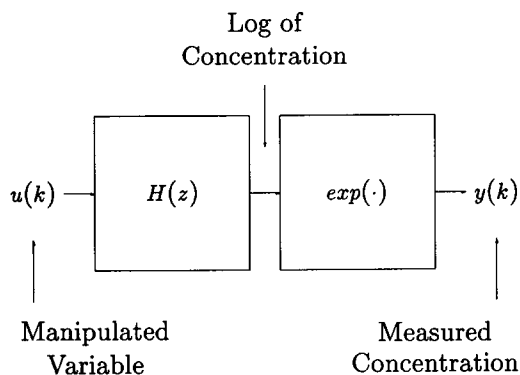


Figure 16 Implicit Wiener model for high-purity columns

This observation has two practical implications for nonlinear model development. First, because the real-world process \mathcal{P} being modelled is itself nonlinear, it is important to characterize its behaviour over the set \mathcal{U} of practically important input sequences as completely as possible. Second, once a candidate structure is selected for the empirical model \mathcal{M} of this process, it is important to characterize *its* qualitative behaviour over the set \mathcal{U} as well. It is only after both of these characterizations have been made that we can hope to estimate 'model adequacy'.

The subharmonic generation example discussed earlier emphasized the difference between 'nonlinear autoregressive' and 'nonlinear moving average' behaviour. In particular, it was shown that subharmonic generation is an 'inherently autoregressive' phenomenon that has no 'moving average representation'. The CSTR step response example discussed in the previous section is less definite, but the results suggest that while this response is 'barely representable' by moving average models (i.e. the infinite-order moving average equivalents of the Wiener and Uryson models), it is much more easily represented by autoregressive models. Conversely, this example also demonstrated that nonlinear autoregressive models can exhibit potentially undesirable behaviour (e.g. multiple steady states, chaotic step responses, amplitude-dependent instabilities etc.). If such qualitative behaviour is inherent in the process \mathcal{P} , its presence in the empirical model \mathcal{M} is reasonable, but this is not always the case. In contrast, block-oriented and other 'nonlinear moving average' models tend to exhibit less of this type of behaviour. Consequently, intermediate cases like the high-purity distillation column pose particularly intriguing questions. Specifically it was noted earlier that while the Hammerstein models investigated by Eskinat *et al.*¹¹ do not adequately describe high-purity column dynamics, the more general question remains open: can these dynamics be adequately described by *any* Hammerstein model, or is the class of Hammerstein models simply too restrictive? A popular alternative to dealing with high-purity column dynamics is to model the logarithm of product concentration instead of the concentration itself. Note that this

alternative implicitly constructs a Wiener model of the process, as shown in Figure 16. That is, since the relationship between the manipulated variable $\{u(k)\}$ and the logarithm of product concentration is assumed linear, the relationship between $\{u(k)\}$ and measured concentration $\{y(k)\}$ is a Wiener model with an exponential compensating nonlinearity, as shown. Nonparametric identification of the nonlinearity in either Hammerstein or Wiener models^{9,10,13} would permit the search for other, more effective nonlinearities in either of these approaches to modelling high-purity distillation columns.

Finally, it is important to note the role that certain classes of nonlinear models can play in the development of algorithms and methodologies, even when they are not the best choice of model structure for practical applications. In particular, it was noted in the previous section that the Volterra model was a poor choice for describing the CSTR dynamics considered there because of the extreme number of parameters required for its specification, even in the restrictive case of second-order Volterra models. Conversely, the investigation of this general class of models described by Pearson *et al.*³⁸ has yielded some very useful insights into the challenging problem of input sequence design for nonlinear system identification. Specifically, these results make it clear that PRBS inputs are singularly inappropriate for Hammerstein model identification, a conclusion that is reasonable in retrospect, important in view of the popularity of PRBS for linear modelling, but not obvious beforehand.

Acknowledgements

Fruitful and on-going discussions and collaborations with many people have contributed significantly to the development of the ideas presented here. Space does not permit an exhaustive list of acknowledgements, but I am particularly indebted to B.A. Ogunnaike, F.J. Doyle, III, W.H. Ray, M. Pottmann and M. Pawlak for their ideas and criticisms.

References

- Congalidis, J.P., Richards, J.R. and Ray, W.H. *AiChE J.* 1989, **35**, 891
- Pearson, R.K. and Ogunnaike, B.A. 'Nonlinear Process Control', ch. 4 (Eds. M.A. Henson and D.E. Seborg), Prentice-Hall, Englewood Cliffs, NJ, in press
- Ljung, L. 'System Identification: Theory for the User', Prentice-Hall, New York, 1987
- Kay, S.M. 'Modern Spectral Estimation', Prentice-Hall, New York, 1988, 112
- Tong, H. 'Non-linear Time Series', Oxford University Press, New York, 1990
- Nikias, C.L. and Petropulu, A. 'Higher-Order Spectra Analysis', Prentice-Hall, New York, 1993
- Billings, S.A. *IEE Proc.* 1980, **127**, 272
- Haber, R. and Unbehauen, H. *Automatica* 1990, **26**, 651
- Greblicki, W. and Pawlak, M. *IEEE Trans. Inform. Theory* 1989, **35**, 4091
- Greblicki W. and Pawlak, M. 'Nonparametric Identification of

- Nonlinear Block-Oriented Systems', Preprints, 9th IFAC/IFORS Symposium, Budapest, Hungary, 1991, 720
- 11 Eskinat, E., Johnson, S.H. and Luyben, W.L. *AIChE J.* 1991, **37**, 255
 - 12 Su, H.T. and McAvoy, T.J. *Ind. Eng. Chem. Res.* 1993, **32**, 1927
 - 13 Greblicki, W. *IEEE Trans. Inform. Theory* 1992, **38**, 1487
 - 14 Brillinger, D.R. *Biometrika* 1977, **64**, 509
 - 15 Billings, S.A. and Voon, W.S.F. *Int. J. Control* 1986, **44**, 803
 - 16 Chen, R. and Tsay, R.S. *J. Am. Stat. Assoc.* 1993, **88**, 955
 - 17 McCulloch, W.S. *Bull. Math. Biophys.* 1943, **5**, 115
 - 18 Rumelhart, D.E. et al. 'Parallel Distributed Processing', vol. 1, MIT Press, Cambridge, MA, 1986
 - 19 Cybenko, G. *Math Controls, Sign. Syst.* 1989, **2**, 303
 - 20 Kramer, M.A. *AIChE J.* 1991, **37**, 233
 - 21 Narendra, K.S. and Parthasarathy, K. *IEEE Trans. Neur. Netw.* 1990, **1**, 4
 - 22 Elman, J.L. *Cogn. Sci.* 1990, **14**, 179
 - 23 Scott, G.M. and Ray, W.H. *J. Proc. Cont.* 1993, **3**, 163
 - 24 Pottmann, M. and Seborg, D.E. *J. Proc. Cont.* 1992, **2**, 189
 - 25 Nayfeh, A.H. and Mook, D.T. 'Nonlinear Oscillations', John Wiley, New York, 1979
 - 26 Chien, I.L. and Ogunnaike, B.A. 'Modeling and Control of High-Purity Distillation Columns,' presented at AIChE National Meeting, Miami, November, 1993
 - 27 Hardle, W. 'Applied Nonparametric Regression', Cambridge University Press, New York, 1990
 - 28 Draper, N. and Smith, H. 'Applied Regression Analysis', 2nd ed., John Wiley, New York, 1981
 - 29 Hernandez, E. and Arkun, Y. *AIChE J.* 1993, **39**, 446
 - 30 Pottmann, M., Unbehauen, H. and Seborg, D.E. *Int. J. Control* 1993, **57**, 97
 - 31 Poljak, B.T. and Tsympkin, J.A.Z. *Automatica* 1980, **16**, 53
 - 32 Huber, P.J. 'Robust Statistics', John Wiley, New York, 1981
 - 33 Leger, C. and Altman, N. *J. Am. Statist. Assoc.* 1993, **88**, 547
 - 34 Morari, M. and Zafiriou, E. 'Robust Process Control', Prentice-Hall, Englewood Cliff, NJ, 1989
 - 35 Monaco, S. and Normand-Cryot, D. *Syst. Control Lett.* 1988, **11**, 229
 - 36 Klambauer, G. 'Mathematical Analysis', Marcel Dekker, New York, 1975
 - 37 Koh, T. and Powers, E.J. *IEEE Trans. Acoust. Speech Sig. Proc.* 1985, **33**, 1445
 - 38 Pearson, R.K., Ogunnaike, B.A. and Doyle, F.J., III, submitted for publication
 - 39 Bendat, J.S. 'Nonlinear System Analysis and Identification from Random Data', John Wiley, New York, 1990
 - 40 Martin, R.D. and Thomson, D.J. *Proc. IEEE* 1982, **70**, 1097
 - 41 Box, G.E.P. and Tiao, G.C. 'Bayesian Inference in Statistical Analysis', Addison-Wesley, Reading, MA, 1973
 - 42 Papoulis, A. 'Probability, Random Variables, and Stochastic Processes', McGraw-Hill, New York, 1965
 - 43 Cho, Y.S. and Powers, E.J. Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, May, 1991, 3117
 - 44 Tseng, C.H. and Powers, E.J. Proc. Third International Symposium on Signal Processing and Its Applications, Queensland, Australia, August, 1992, 212
 - 45 Nahas, E.P., Henson, M.A. and Seborg, D.E.. *Comput. Chem. Eng.* 1992, **16**, 1039
 - 46 Munkres, J.R. 'Topology: A First Course'. Prentice-Hall, Englewood Cliffs, NJ, 1975
 - 47 Pearson, R.K. in preparation
 - 48 Davis, H.T. 'Introduction to Nonlinear Differential and Integral Equations'. Dover, New York, 1962