J. Proc. Cont. Vol. 8, No. 2, pp. 79–88, 1998

Printed in Great Britain
0959-1524/98 $19.00 + 0.00

ELSEVIER

PII: S0959-1524(97)00029-2

Full Papers

Control relevant model reduction of Volterra series models

Wei-Ming Ling and Daniel E. Rivera*

Department of Chemical, Bio, and Materials Engineering and Control Systems Engineering Laboratory, Manufacturing Institute, Arizona State University, Tempe, Arizona 85287-6006, USA

Received 31 July 1996; revised 1 April 1997

This paper presents a two-step method for control-relevant model reduction of Volterra series models. First, using the nonlinear IMC design as a basis, an explicit expression relating the closed-loop performance to the open-loop modeling error is obtained. Secondly, an optimization problem that seeks to minimize the closed-loop error subject to the restriction of a reduced-order model is posed. By showing that model reduction of kernels with different degrees can be decoupled in the problem formulation, the optimization problem is simplified into a mathematically more convenient form which can be solved with significantly less computational effort. The effectiveness of the proposed method is illustrated on a polymerization reactor example where a second-order Volterra model with 85 parameters is reduced to a Hammerstein model with 3 parameters. Despite the lower 'open-loop' predictive ability of the control-relevant model, the closed-loop performance of the reduced-order control system closely mimics that of the full order model. © 1993 Elsevier Science Ltd. All rights reserved

Keywords: control-relevant modeling; Volterra series; model reduction

Various 'general' input/output model structures have been conceived to describe nonlinear systems. Among them, the Volterra series[1,2] and the NARMAX (Nonlinear AutoRegressive Moving Average with eXgeneous input)[3] models are among the most widely used model structures. Theoretically, these models can represent a large class of nonlinear systems with arbitrary accuracy. But in practice, some structural restrictions must be imposed on these 'general' models due to the number of model parameters involved. For example, the finite Volterra model according to

$$y(k) = \sum_{n=1}^{N}\sum_{i_1=1}^{M}\cdots\sum_{i_n=i_{n-1}}^{M} h_n(i_1,\ldots,i_n) \times u(k-i_1)\cdots u(k-i_n) \qquad (1)$$

reaches steady state in M sampling intervals to a step input. If it is used to describe a process with a settling time of 30 min and a sampling time of 2 min, then $M \approx 15$ must be selected. For a third-order Volterra model ($N=3$), about 815 parameters will be included by taking into account kernel symmetry. The incentive to seek parsimonious, reduced-order alternatives to these models is clear.

While restrictions imposed on 'general' nonlinear models may make them more tractable in practice, they inevitably introduce model errors. Consider the widely used Hammerstein model which consists of a zero-memory nonlinear element followed by a linear dynamic element. It can be represented as a Volterra series model with the off-diagonal model parameters set to zero (i.e. $h_n(i_1, i_2, \ldots, i_n) = 0$, unless $i_1 = i_2 = \cdots = i_n$) if the nonlinear function representing the zero-memory nonlinear element is analytic. Step responses of a Hammerstein model display the same nominal speed-of-response in spite of changes in the size of the step input applied. Few real-life processes display exactly this kind of dynamic behavior. Therefore, if reduced models such as the Hammerstein model are used for control design purposes, an important question to ask is the following: how can model error be 'distributed' such that a reduced-order model retains the most important information related to closed-loop performance? In this paper, the control-relevant model reduction problem of Volterra series models is investigated through the following two steps:

1. First, using the Nonlinear Internal Model Control (NIMC) design procedure as a basis, an explicit expression relating the closed-loop control error

*To whom correspondence should be addressed. Fax: (602) 965-2910; e-mail: daniel.rivera@asu.edu

(defined as the difference between the set point and the system output) and the modeling error is obtained.

2. An optimization problem that seeks to minimize the closed-loop error subject to the restriction of a reduced-order model is then posed. The optimization problem is further manipulated into a mathematically simpler form, leading to a solution that, although suboptimal, can be solved with significantly less computational effort.

We demonstrate that by incorporating closed-loop consideration into the modeling procedure, significant model reduction can be achieved with only slight degradation in closed-loop performance. The problem is philosophically similar to the linear problem, previously analyzed by Rivera and Morari[4]. The problem formulation is simple enough to be solved via general-purpose optimization routines, such as those found in the Optimization Toolbox in MATLAB. Effectiveness of the proposed method is illustrated through a polymerization reactor example[5]. In the example, a second-order Volterra model with 85 parameters is reduced to a Hammerstein model with three parameters. Despite the lower 'open-loop' predictive ability of the control-relevant restricted complexity model, the closed-loop performance of the reduced-order control system closely mimics that of the full order model.

The paper is organized as follows. The next section consists of a brief discussion of the Volterra series representation of nonlinear systems. Following this, the nonlinear IMC (NIMC) design is introduced as a bridge to access the model control relevance. Next, the control-relevant model reduction problem is formulated and a corresponding solution procedure is proposed. The model validation issue is then considered, followed by an application example. Finally, a brief summary and some conclusions are given.

Volterra series model representation of nonlinear systems

In this paper, we shall consider the model reduction problem of a finite Volterra series model of the following form:

$$y(k) = \sum_{n=1}^{N} \sum_{i_1}^{M} \cdots \sum_{i_n = i_{n-1}}^{M} h_n(i_1, \ldots, i_n) \times u(k - i_1) \cdots u(k - i_n) \tag{2}$$

Equation (2) is a typical Volterra series model obtained from an identification procedure. A relative recent work on the theoretical justification of the Volterra series model representation of nonlinear systems can be found in Boyd and Chua[6]. A large class of nonlinear time invariant system models are covered by the Volterra series model (2). For example, the widely used block-oriented models[7] and bilinear models have equivalent Volterra series model representations under certain conditions. For a detailed discussion of relationships between various model structures, we refer to Pearson[8].

In spite of its theoretical attractiveness, the number of parameters in the Volterra series model often poses a major obstacle in applications. To limit the number of model parameters, the model structure (2) is often truncated to second or third order. But even for a third order system, the number of model parameters involved may still pose a problem in practice, as noted in the introductory section; meanwhile, a second-order Volterra model may only describe system nonlinearity in a very limited operation range. In order to include the higher-order nonlinearity without introducing too many model parameters, some additional model structure restrictions are often imposed which lead to using a Hammerstein model, a Wiener model, or other structures. For example, if we truncate the Volterra model in (2) to the kth order and further restrict that $h_n(i_1, i_2, \cdots, i_n) = 0$ $\forall n = 1, 2, \cdots k$ (unless $i_1 = i_2 = \cdots = i_k$ or $i_j = i_1 \pm 1, j = 1, 2, \cdots, k$), the Volterra model is reduced to a tridiagonal model[9]. Other approaches examined in the literature to reduce the model parameters include using the orthonormal basis function expansions, such as the Laguerre series expansion[10,11]. The orthogonal least squares method[12] which has been extensively studied for NARMAX model estimation can also be applied in the Volterra series model case. Control-relevance issues associated with Volterra and Volterra-Laguerre expansions have been examined by Zheng and Zafiriou[11]; otherwise, little attention has been paid to model reduction/parameter estimation problems which incorporate closed-loop performance considerations.

Our objective is to formulate and solve a model reduction problem which explicitly minimizes a 2-norm closed-loop error criterion. To formulate this control-relevant model reduction problem, an expression relating the closed-loop tracking error to the open-loop modeling error is needed. The problem requires defining a control design procedure, and NIMC design[13] is naturally introduced for this purpose, as developed in the ensuing section.

NIMC design: relating closed-loop performance to the open-loop modeling error

Figure 1 is the NIMC structure of a Volterra model where p and $\tilde{p}$ denote the plant (full-order Volterra

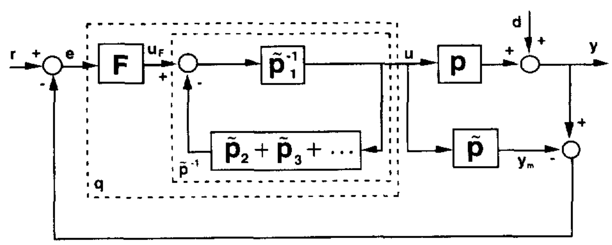

Figure 1 IMC structure of the Volterra model

model) and the plant model (reduced order model), respectively; q is the IMC controller which consists of the inverse plant model $\tilde{p}^{-1}$ augmented by a linear filter F. The linear filter F is introduced for robustness and other considerations such as physical realizability. Also, when e does not belong to the range space of operator $\tilde{p}$, it is assumed that F will project e into the appropriate space since by definition of the inverse operator, the domain of the operator $\tilde{p}^{-1}$ is the range of $\tilde{p}$.

An advantage of using the Volterra series model representation is that explicit expressions for the inverse Volterra operator $\tilde{p}^{-1}$ and consequently the NIMC controller q can be derived. Assume that the full order model p represented in the operator notation is given by

$$p = p_1 + p_2 + \cdots + p_m \tag{3}$$

where operator p_i represents the ith order Volterra kernel. The reduced model $\tilde{p}$ is given by

$$\tilde{p} = \tilde{p}_1 + \tilde{p}_2 + \cdots + \tilde{p}_n \tag{4}$$

with some implicit structure restrictions on $\tilde{p}_1, \tilde{p}_2, \cdots, \tilde{p}_n$. For example, $\tilde{p}_1$, $\tilde{p}_2$, $\cdots$, $\tilde{p}_n$ could represent Volterra kernels of an equivalent block-oriented model. Here, the order n of the reduced model is allowed to be different from that of the full order model (m in (3)).

To derive the inverse Volterra operator, let $\tilde{p}^{-1}$ be represented as the Volterra series:

$$\tilde{p}^{-1} = g_1 + g_2 + g_3 + \cdots \tag{5}$$

Then

$$\begin{aligned} I = \tilde{p} * \tilde{p}^{-1} &= (\tilde{p}_1 + \tilde{p}_2 + \cdots + \tilde{p}_n) \\ &* (g_1 + g_2 + g_3 + \cdots) \end{aligned} \tag{6}$$

where I denotes the usual identity operator and '$*$' denotes the composition of related operators. The following relationships are then obtained by expanding the composite operator in Equation (6) and equating like terms on both sides of the resulting equation[1]

$$\begin{aligned} g_1 &= \tilde{p}_1^{-1} \\ g_2 &= -\tilde{p}_1^{-1} * \tilde{p}_2 * g_1 \\ g_3 &= -\tilde{p}_1^{-1} * (\tilde{p}_2 * (g_1 + g_2) \\ &\quad - \tilde{p}_2 * g_1 - \tilde{p}_2 * g_2 + \tilde{p}_3 * g_1) \\ &\vdots \end{aligned} \tag{7}$$

Equivalently, the infinite Volterra series (7) can be represented as a feedback structure, as shown in *Figure 1*[13]. Implementation of the feedback structure is straightforward, while the Volterra series (7) is valuable for problem analysis.

Augmenting the inverse Volterra model $\tilde{p}^{-1}$ with the linear filter F gives the NIMC controller q in *Figure 1*. In the Volterra series form, q is given by

$$\begin{aligned} q_1 &= \tilde{p}_1^{-1} F \\ q_2 &= -\tilde{p}_1^{-1} * \tilde{p}_2 * q_1 \\ q_3 &= -\tilde{p}_1^{-1} * (\tilde{p}_2 * (q_1 + q_2) - \tilde{p}_2 * q_1 \\ &\quad - \tilde{p}_2 * q_2 + \tilde{p}_3 * q_1) \\ &\vdots \end{aligned} \tag{8}$$

By properly selecting F, q_1 can be assumed stable and causal and q_i (i=2, 3, $\cdots$) causal. The following three points are particularly worth noting regarding the above expression (8). First, the linear kernel inverse $\tilde{p}_1^{-1}$ is the only inverse operator required in (8). Second, equations in (8) display a recursive pattern, i.e. q_1 in the first equation only depends on $\tilde{p}_1$; once q_1 is determined, q_2 in the second equation only depends on $\tilde{p}_2$; and so on. Third, q is generally an infinite Volterra series. For $e(t) \in l_\infty$ (*Figure 1*), it can be shown that the Volterra series q converges if $\lambda_e = \sup_t |e(t)|$ is sufficiently small (cf. Ref. 2).

The NIMC structure provides improved perspectives on modeling requirements for control. We know that if p is input/output stable, $\tilde{p}^{-1}$ is stable and causal, and in the absence of plant/model mismatch, letting F=1 will lead to 'perfect control', i.e. exact set point tracking $y(t) \equiv r(t)$. In practice, several factors will prevent us from achieving the 'perfect control'. First the inverse operator $\tilde{p}^{-1}$ may not exist or be unstable/non-causal. Second, the resulting control action u may be too aggressive and/or exceed physical constraints. Third, in the presence of inevitable modeling errors, too aggressive a controller design will give rise to robustness problems.

In a model reduction problem, using a restricted complexity model $\tilde{p}$ with fewer parameters than the true plant p will inevitably introduce modeling errors. Consequently, closed-loop performance will deteriorate for a controller design based on the reduced order model $\tilde{p}$. How then can a reduced order model $\tilde{p}$ be determined such that $\tilde{p}$ retains the most important information related to the closed-loop performance? In the following section, the problem is addressed by deriving an expression relating the control error to the modeling error $(p - \tilde{p})$.

Control-relevant model reduction problem formulation

The closed-loop performance criterion considered in this study is the 2-norm of the tracking error, which is defined as the difference of the set point r of the plant output and the actual plant output y.

$$e_c(t) = r(t) - y(t) \tag{9}$$

Using the IMC structure in *Figure 1*, the following relationships can be derived:

$$\begin{aligned} e_c = r - y &= (r - d) - p * q[e] \\ &= (r - d) - p * q * [I + (p - \tilde{p}) * q]^{-1} [r - d] \\ &= [I - p * q * (I + (p - \tilde{p}) * q)^{-1}][r - d] \end{aligned} \tag{10}$$

Expanding the identity operator I as

$$I = [I + (p - \tilde{p}) * q] * [I + (p - \tilde{p}) * q]^{-1} \tag{11}$$

and substituting into (10) leads to

$$e_c = (I - \tilde{p} * q) * (I + (p - \tilde{p}) * q)^{-1}[r - d] \tag{12}$$

To represent e_c as a Volterra series, we shall denote

$$\begin{aligned} s &\equiv I + (p - \tilde{p}) * q = \sum_{i=1}^{\infty} s_i; \\ w &\equiv [I + (p - \tilde{p}) * q]^{-1} = \sum_{i=1}^{\infty} w_i \end{aligned} \tag{13}$$

where s_i and w_i correspond to the ith order Volterra kernels. Then

$$\sum_{i-1}^{\infty} s_i = I + [\sum_{i=1}^{N}(p_i - \tilde{p}_i)] * (\sum_{j=1}^{\infty} q_i) \tag{14}$$

with $N = \max(m, n)$. Equating like terms on both sides of the above equation leads to:

$$\begin{aligned} s_1 &= I + (p - \tilde{p}_1)q_1 \\ s_2 &= (p_1 - \tilde{p}_1) * q_2 + (p_2 - \tilde{p}_2) * q_1 \\ s_3 &= (p_1 - \tilde{p}_1) * q_3 + (p_2 - \tilde{p}_2) * (q_1 + q_2) \\ &\quad - (p_2 - \tilde{p}_2) * q_1 - (p_2 - \tilde{p}_2) * q_2 \\ &\quad + (p_3 - \tilde{p}_3) * q_1 \\ &\vdots \end{aligned} \tag{15}$$

Matching the inverse operator relations in Equation (7) yields kernels of the Volterra operator w:

$$\begin{aligned} w_1 &= s_1^{-1} = I + [-(p_1 - \tilde{p}_1)q_1][I + (p_1 - \tilde{p}_1)q_1]^{-1} \\ &= I + w_1' \end{aligned} \tag{16}$$

$$w_2 = -(w_1 * s_2) * s_1^{-1} \tag{17}$$

$$\vdots \tag{18}$$

Relations in (16) are obtained based on the fact that p_1, q_1 are linear operators. Denote the operator ϵ as

$$\epsilon = I - \tilde{p} * q = I - \tilde{p} * \tilde{p}^{-1} * F = I - F \tag{19}$$

which represents the desired closed-loop response. The 2-norm of the tracking error is then given by

$$\begin{aligned} \| e_c \|_2 = \| &(\epsilon + \epsilon * w_1' + \epsilon * w_2 + \cdots \\ &+ \epsilon * w_n + \cdots)[r - d] \|_2 \end{aligned} \tag{20}$$

$$\begin{aligned} \leq \| \epsilon[r - d] \|_2 + \| &(\epsilon * w_1' + \epsilon * w_2 + \cdots \\ &+ \epsilon * w_n + \cdots)[r - d] \|_2 \end{aligned} \tag{21}$$

For $(r(t) - d(t)) \in l_\infty$, Using the method of Rugh (Ref. 2, Appendix 1.1), again it can be proved that the Volterra series in (20) converges if $\lambda_{(r-d)} = \sup_t | (r - d) |$ is sufficiently small.

The first term in (21) is a nominal performance term and not directly related to the modeling error. It comes from the fact that perfect control was detuned in the NIMC design. Once the NIMC filter F is fixed, it will always exist whether or not there is modeling error. The purpose of the control-relevant model reduction is to search a reduced process model $\tilde{p}$ which minimizes the closed-loop tracking error arising from the other parts of e_c in (21). Also, a convergent Volterra series in (21) can be approximated by a truncated finite series up to order n. The resulting truncated finite Volterra series w is called an nth order inverse and will be accurate for inputs sufficiently small[1]. A control-relevant model reduction problem is then formulated as

$$\begin{aligned} \tilde{p} = \arg\min_{\tilde{p}} \| &(\epsilon * w_1' + \epsilon * w_2 + \cdots \\ &+ \epsilon * w_n)[r - d] \|_2 \end{aligned} \tag{22}$$

Even though complexity of Volterra kernels w_i, $(i = 1, \cdots, n)$ increases rapidly as n increases, the following facts can be observed: (a) w_i still displays a recursive pattern, i.e. w_1 only depends on $\tilde{p}_1$; once $\tilde{p}_1$ is determined, w_2 only depends on $\tilde{p}_2$; and so on; (b) for most practical problems, a Volterra series model up to 4th order ($m \leq 4$ in Equation (3)) covers a sufficiently large operation range. It is reasonable to assume that the order of the reduced order model satisfies $n \leq m \leq 4$.

The problem now is how to solve the model reduction problem posed in (22) without overwhelming mathematical effort. To solve (22), only the filter F (which represents the desired closed-loop response) and the *structure* of the reduced order model $\tilde{p}$ need to be specified at the start of computations. Formula (22) forms a nonlinear optimization problem, which can be solved numerically. One reasonable initial guess for the parameters to the optimization problem is to use the results of conventional open-loop prediction error estimation,

$$\begin{aligned} \tilde{p} = \arg\min_{\tilde{p}} \| &((p_1 - \tilde{p}_1) + (p_2 - \tilde{p}_2) + \cdots \\ &+ (p_n - \tilde{p}_n))[u_o] \|_2 \end{aligned} \tag{23}$$

Solution to (23) requires choosing an input signal; a white noise input u_0 is one option. From u_0 and y (the model's response to u_0), solution to (23) can be obtained using various methods (e.g. Refs 14, 15) depending on the structure of the restricted complexity model.

When the number of parameters in the reduced order model is 'small', it can be solved directly using general purpose numerical optimization routines. But if the number of reduced-order model parameters is large, local minima and long computation times may arise. As an alternative, we consider simplifying the (potentially) formidable-looking expression (22). Following the approach of Doyle *et al.*[13], we relax the requirement to minimize the norm of the composite control error in

(22). Instead, the norm of each term is minimized separately. The control-relevant model reduction problem is then formulated as sequential solutions to the following optimization problems:

$$\tilde{p}_1 = \arg\min_{\tilde{p}_1} \| \epsilon w_1' [r-d] \|_2 \tag{24}$$

$$\tilde{p}_2 = \arg\min_{\tilde{p}_2} \| \epsilon w_2 [r-d] \|_2 \tag{25}$$

$$\vdots \tag{26}$$

$$\tilde{p}_n = \arg\min_{\tilde{p}_n} \| \epsilon w_n [r-d] \|_2 \tag{27}$$

In the above problem formulation, model reduction of kernels with different orders are decoupled. The linear kernel $\tilde{p}_1$ is obtained by solving the minimization problem in (24); having obtained $\tilde{p}_1$, the minimization problem in (25) is solved to give the second order kernel $\tilde{p}_2$, and so on. The minimization problem in (24) is exactly the same as the control-relevant linear model reduction problem formulation discussed by Rivera and Morari[4], which can be solved using the iterative prefiltered prediction-error algorithm described in Rivera *et al.*[16]. If the reduced order model is a Hammerstein, Wiener or cascade model, the decoupled minimization problems for all the nonlinear kernels form a one-dimensional search. Hence, they can be solved with much less effort than the original problem (22).

It is interesting to contrast the control-relevant formulation with the traditional 'open-loop' prediction error formulation (23). In the control-relevant formulation (22), terms representing the modeling error for each kernel $(p_i - \tilde{p}_i)$ are individually weighted based on closed-loop performance requirements. In the prediction error method problem formulation (23), the only design variable that is at our disposal is the input perturbation signal. When using a restricted complexity model, the input signal design plays an important role in 'shaping' the information content of the reduced order model estimate. For linear systems, problem analysis and consequent input signal design methodologies from the control perspective are discussed by Rivera *et al.*[17]. But for nonlinear systems, designing input perturbation which takes control requirement into account is by no means a simple task. In the control-relevant problem formulation (22) or (24)–(27), selection of the IMC filter and the setpoint/disturbance direction $(r-d)$ naturally influence the model reduction problem, without having to resort to *ad hoc* procedures.

Model validation

Once a reduced-order model $\tilde{p}$ is obtained by solving (22) or (24)–(27), it is important to verify that a controller designed on the reduced order model will satisfy certain closed-loop performance requirements.

The first step of model validation is concerned with the stability of the feedback system in *Figure 1*. The problem can be stated as: given a set of desired inputs $(r-d)$ in (22), the resulting closed-loop system based on the reduced order model must be stable. From a practical standpoint the problem can always be examined through an exhaustive set of simulations. We consider, however, a less aggressive but more general validation criteria based on the local small gain theorem studied by Zheng and Zafiriou[18]. Let $\gamma_{\mathcal{P}}$ denote the gain of a Volterra operator $\mathcal{P}$ defined in some local domain. The closed-loop system in *Figure 1* is locally stable if

$$\gamma_q \gamma_{(p-\tilde{p})} < 1 \tag{28}$$

for a sufficiently small $\lambda_{(r-d)} = \| (\lambda - d) \|_\infty$. Clearly, usefulness of the criteria (28) depends on the way in which the local gains γ_q and $\gamma_{(p-\tilde{p})}$ are estimated. Notice that $(p-\tilde{p})$ is just a Volterra operator obtained from additive parallel connection of two Volterra operators p and $-\tilde{p}$. Denote the nth order kernel of $(p-\tilde{p})$ as $k_n(i_1, i_2, \ldots, i_n)$ and

$$| (p-\tilde{p})_n | = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} | k_n(i_1, i_2, \ldots, i_n) |$$

Then for $\| u \|_\infty \leq U$,

$$\| (p-\tilde{p})[u] \|_\infty \leq \sum_i | (p-\tilde{p})_i | \, U^{i-1} \, \| u \|_\infty$$

We use $\sum_i | (p-\tilde{p})_i | \, U^{i-1}$ as an estimate of $\gamma_{(p-\tilde{p})}$. For γ_q, we have $\gamma_q \leq \gamma_{\tilde{p}^{-1}} \gamma_F$ since $q = \tilde{p}^{-1} F$, and F is a linear operator. An estimation method for $\gamma_{\tilde{p}^{-1}}$ in the case that $\tilde{p}$ is a general Volterra model is proposed in Ref. 18. Considering that $\tilde{p}$ represents a restricted complexity model in this study, a less conservative estimate for γ_q can often be obtained by exploiting the special structure of $\tilde{p}$. For example, when $\tilde{p}$ is a Hammerstein model, q may be represented as a Wiener model with a linear element $\tilde{p}_1^{-1} F$ followed by a zero memory nonlinear element denoted as f_w. Systematic methods exist for computing $\gamma_{\tilde{p}_1^{-1} F}$ and γ_{f_w}[19].

An example

The control of a CSTR polymerization reactor[5,13] is used to illustrate the effectiveness of the proposed method. The free-radical polymerization of methyl methacrylate occurs in the reactor with azo-bis-isobutyronitrile as initiator and toluene as solvent. The control objective is to regulate the output number average molecular weight (NAMW) by manipulating the volumetric flowrate u of the initiator. Under certain assumptions, the dynamics of the reactor can be described by the following equations:

$$\begin{aligned} \dot{x}_1 &= 10(6 - x_1) - 2.4568 x_1 \sqrt{x_2} \\ \dot{x}_2 &= 80u - 10.1022 x_2 \\ \dot{x}_3 &= 0.002412 x_1 \sqrt{x_2} + 0.112191 x_2 - 10 x_3 \\ \dot{x}_4 &= 245.978 x_1 \sqrt{x_2} - 10 x_4 \\ y &= x_4 / x_3 \end{aligned} \tag{29}$$

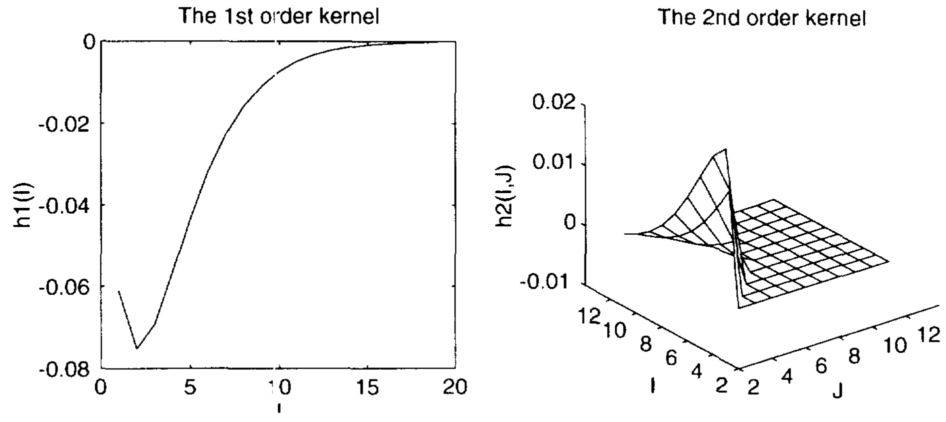

Figure 2 Second-order Volterra model kernels

Nominal operation conditions are: $x_{10} = 5.50677$, $x_{20} = 0.132906$, $x_{30} = 0.0019752$, $x_{40} = 49.3818$, $u_0 = 0.016783\,\mathrm{m^3\,hr^{-1}}$ and $y_0 = 25000.5$ NAMW.

State variables x_i, input u and output y are first normalized using $z_i = (x_i - x_{i0})/x_{i0}$, $v = (u - u_0)/u_0$ and $w = (y - y_0)/y_0$. Using the procedure discussed by Maner *et al.*[20], an approximate second-order Volterra model is then obtained via Carleman linearization.

$$w(k) = \sum_{i=1}^{M} h_1(i)v(k-i) + \sum_{i=1}^{N}\sum_{j=i}^{N} h_2(i,j)v(k-i)v(k-j) \tag{30}$$

With a sampling time of 0.04 hr, selecting $M = 19$ and $N = 11$ offers a reasonable approximate model. In this setting, the resulting second-order Volterra model contains 85 parameters. Corresponding Volterra kernels are shown in *Figure 2*.

Our objective is to reduce the Volterra model (30) to a Hammerstein model with a zero-memory nonlinear element given by

$$f(v) = v + a_1 v^2 \tag{31}$$

and a linear dynamic element according to

$$G(z) = \frac{a_2 z^{-1}}{1 + a_3 z^{-1}} \tag{32}$$

The reduced model contains only three parameters, a sharp contrast to the full model.

The control problem is assumed to track step setpoint changes shown in *Figure 3*. F is selected as a first-order filter

$$F(z) = \frac{(1-\delta)z^{-1}}{1 - \delta z^{-1}} \tag{33}$$

where $\delta = e^{-0.04/0.3}$. (The data is selected based on the fact that the process sampling time is 0.04 hr and the open loop nominal time constant is around 0.3 hr). Since the reduced order model contains only three model parameters, minimization problem (22) is directly solved to obtain model parameters. The unconstrained multivariable function minimization command **fminu** in MATLAB's Optimization Toolbox, which uses a BFGS quasi-Newton method, was applied without difficulties. We compare the results of control-relevant (CR) estimate with those from open-loop prediction error minimization (PEM). In the prediction error method, a zero-mean Gaussian white noise perturbation signal with variance 0.1 is used to generate the parameter estimation data and model parameters are estimated using the algorithm proposed by Narendra and Gallman[14]. Model parameter estimates for both the CR and PEM model are collected in *Table 1*. For comparison, model parameters estimated using the decomposed control-relevant approach (24)–(27) are also given in *Table 1*. In the decomposed approach, an ARX model generated from iterative prefiltering of data, as described by Rivera *et al.*[16], produces the linear model parameters, followed by a one-dimensional search for the nonlinear parameter.

To compare the 'open-loop' predictive abilities of different models, three cross validation data sets were

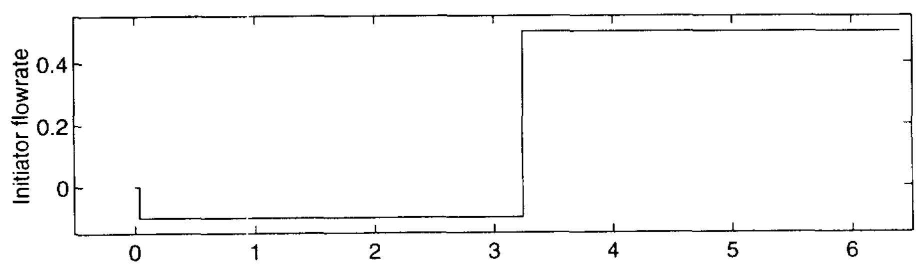

Figure 3 Setpoint changes used in the control-relevant model reduction

Table 1 Parameters of reduced models

Model	Model parameters		
	a_1	a_2	a_3
PEM model	−0.3453	−0.0918	−0.7788
CR model	−0.7999	−0.0967	−0.7610
CR model (decomposed approach)	−0.7023	−0.0947	−0.7754

generated by exciting the plant with a zero-mean Gaussian white noise (GWN) of variance 0.1 and two pseudo random ternary (PRT) inputs. Since PEM model parameters are estimated using a GWN input, here PRT inputs are introduced as a different kind of perturbation signals for comparison. Both PRT signals are generated using the same number of shift registers ($n=5$) and the same switching time ($T_{sw}=0.2$ hr), but with different signal levels. The superior 'open-loop' predictive ability of the PEM model is obvious, as evidenced by the mean square prediction error in *Tables 2* and *3* and *Figure 4*. Note that for the PRT input with signal levels between ± 0.8, the PEM model still gives the smallest prediction error, even though the second-order Volterra model (30) cannot offer good approximation in this input range.

Table 2 Open-loop performance compared to the 2nd-order Volterra model

Model	Mean square prediction error$\times 10^4$		
	GWN input	PRT input (−0.4∼0.4)	PRT input (−0.8∼0.8)
PEM	4.6501	4.9790	58.7510
CR	15.6930	12.5148	174.6066
CR (decomposed)	13.2008	10.3404	145.0544

Table 3 Open-loop performance compared to the true plant

Model	Mean square prediction error$\times 10^4$		
	GWN input	PRT input (−0.4∼0.4)	PRT input (−0.8∼0.8)
Volterra	1.4746	3.5887	39.2115
PEM	5.5110	6.2088	78.7935
CR	15.7066	14.5354	176.2024
CR (decomposed)	13.2741	11.7620	144.7601

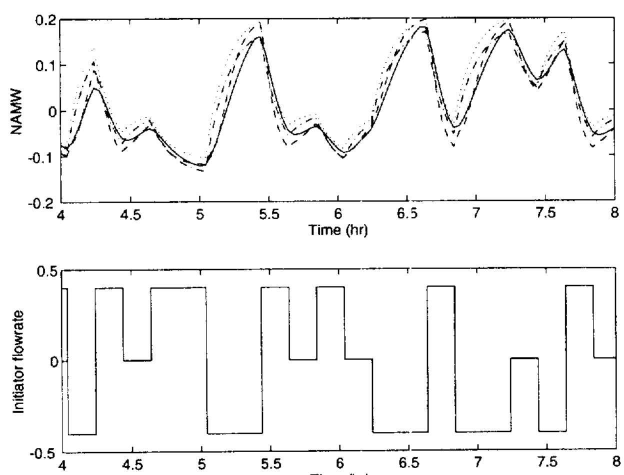

Figure 4 Open-loop responses. solid: true plant; dashed: 2nd-order Volterra model; dashdot: PEM model; dotted: CR model

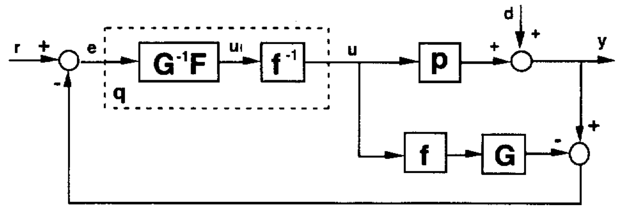

Figure 5 IMC structure of the Hammerstein model

We use both the PEM and CR models, along with the second-order Volterra model, to design nonlinear IMC controllers, tuned with the same filter given in (33). For a Hammerstein model, the IMC controller q reduces to a Wiener model (*Figure 5*). For f being a polynomial of degree two, two solutions can be found for $f(u)=u_i$. Here the control action u is taken as the solution with the smaller deviation from the steady state value:

$$u=\frac{-1+\sqrt{1+4a_1u_i}}{2a_1}$$

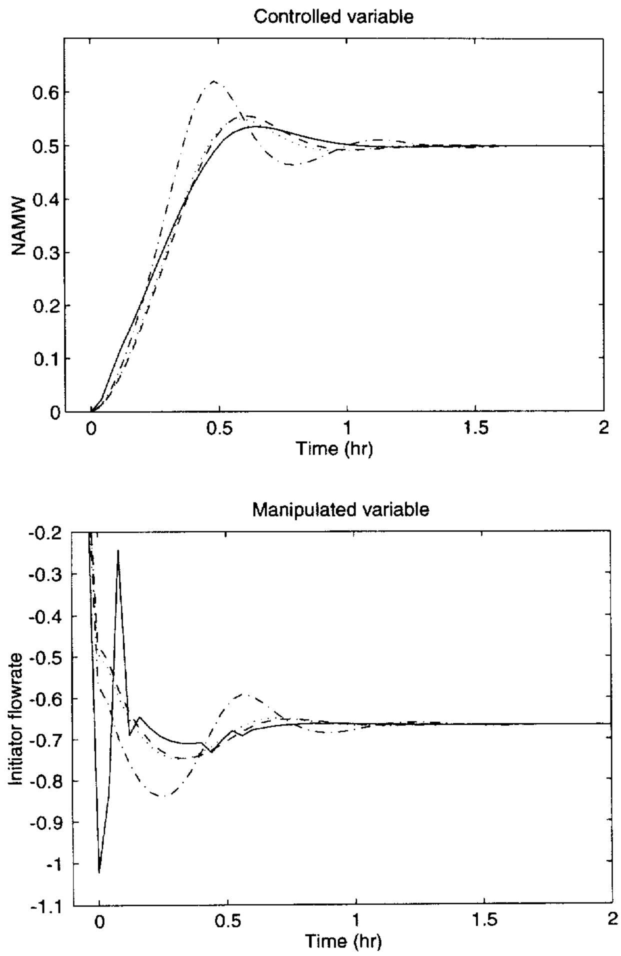

Figure 6 Closed-loop responses for a +50% step set point change. solid: 2nd-order Volterra model; dashdot: PEM model; dashed: CR model; dotted: CR model (decomposed approach)

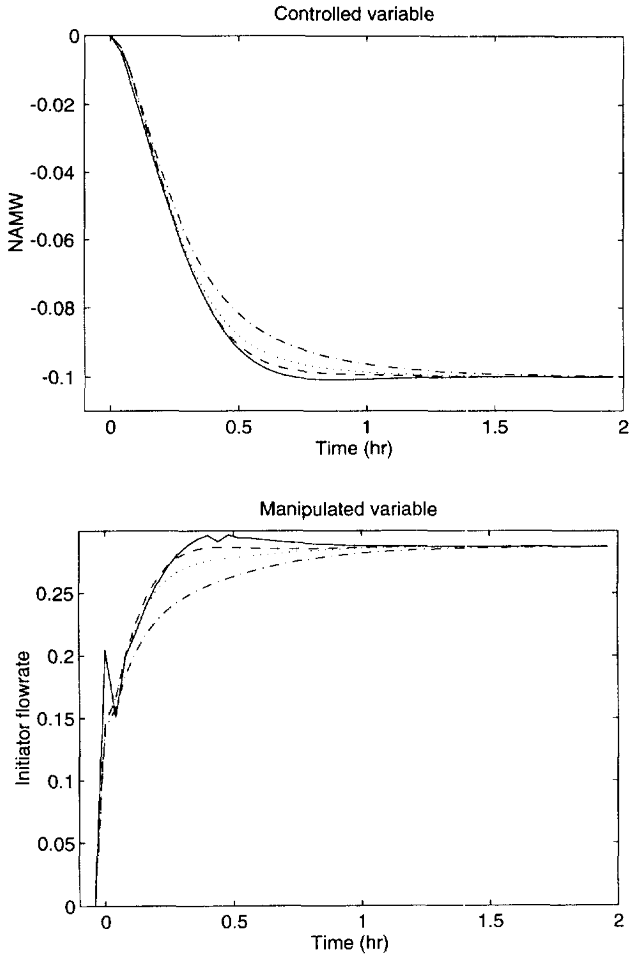

Figure 7 Closed-loop responses for a -10% step set point change. solid: 2nd-order Volterra model; dashdot: PEM model; dashed: CR model; dotted: CR model (decomposed approach)

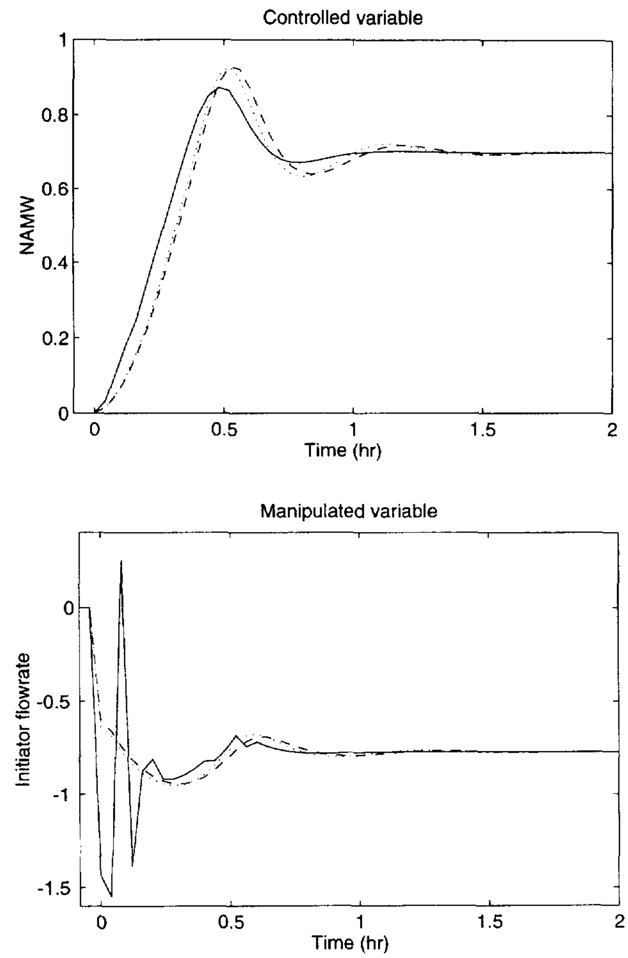

Figure 8 Closed-loop responses for a $+70\%$ step set point change. solid: 2nd-order Volterra model; dashed: CR model; dotted: CR model (decomposed approach); PEM model response (not shown) is unstable

It is clear that when $a_1 < 0$, no real solution u exists for $u_i > -1/4a_1$.

Four controllers are used to control the original system (29). Closed-loop responses to different step set point changes are given in *Figures 6–9*. *Figures 6* and *7* show the responses to set point changes of $+50\%$ and -10%, respectively, which are just input ranges used in the control-relevant model reduction procedure. Responses to a -15% step set point change are given in *Figure 9*. To a $+70\%$ set point change, the controller from the PEM model gives an unstable response, so only closed-loop responses of the full-order Volterra and the CR models are given in *Figure 8*. As seen in *Figures 6–9*, the control-relevant model clearly outperforms the PEM model in the closed-loop. These results give yet another example of what has been demonstrated many times in linear model control-relevance—a model that appears to be 'worse' in the open-loop sense turns out to be better suited for closed-loop control. Furthermore, closed-loop responses of control-relevant models closely follow those of the original Volterra model.

Model validation results using the local small gain criteria (28) are given in *Table 4*. In this table $\gamma_{(p-\tilde{p})}$ is estimated using the full second order Volterra model approximation to the true plant. Since this is a local gain, two input ranges are evaluated ($U=0.15$ and $U=0.5$). Results of *Table 4* indicate that in both cases and for the same tuning parameters, the control-relevant models possess *higher* values of the loop gain ($\gamma_q\gamma_{(p-\tilde{p})}$) compared to the open-loop PEM model. This occurs in spite of the fact that control-relevant models display superior closed-loop performance. However, this result is not surprising since similar behavior is observed in the linear control-relevant model reduction problem[4]. In the linear case, performance considerations in the control-relevant problem formulation lead to a weighted small gain condition $\min_{\tilde{p}} \| W(\omega)q(p-\tilde{p}) \|$, where the weight includes the plant sensitivity function and setpoint-disturbance directions. So while stability is guaranteed by keeping the loop gain less than one

Table 4 Estimated operator gain and uncertainty coefficient

Model	$U=0.15$			$U=0.5$		
	γ_q	$\gamma_{(p-\tilde{p})}$	$\gamma_q\gamma_{(p-\tilde{p})}$	γ_q	$\gamma_{(p-\tilde{p})}$	$\gamma_q\gamma_{(p-\tilde{p})}$
PEM	2.5412	0.1674	0.4254	2.9125	0.3185	0.9276
CR	2.8085	0.1932	0.5426	4.1189	0.4074	1.6780
CR (decomposed approach)	2.6508	0.1874	0.4968	3.6553	0.3920	1.4329

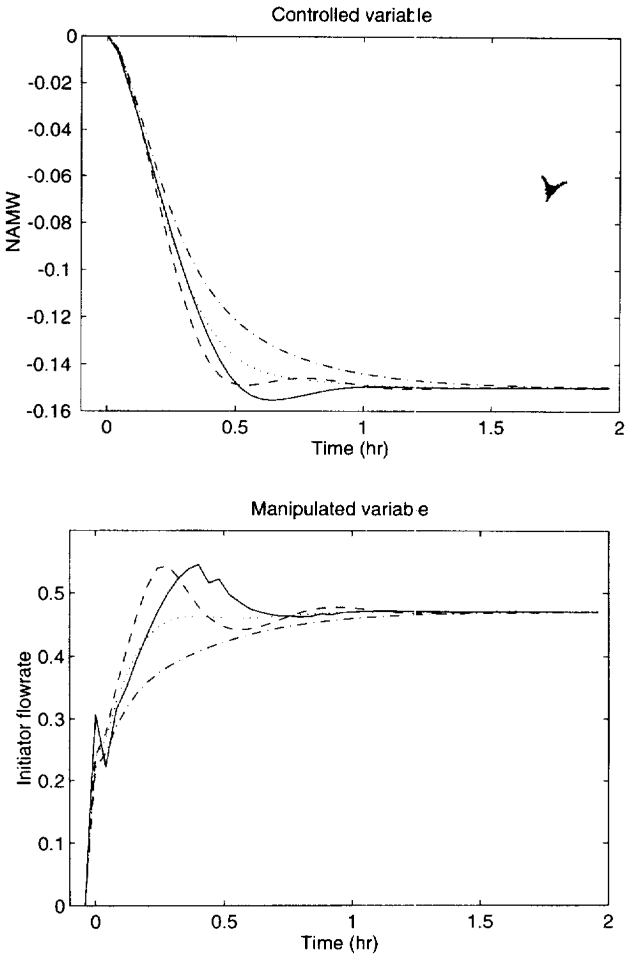

Figure 9 Closed-loop responses for a −15% step set point change. solid: 2nd-order Volterra model; dashdot: PEM model; dashed: CR model; dotted: CR model (decomposed approach)

Table 5 Step set point changes with stable responses: the upper and lower bounds

Model	Upper bound	Lower bound
2nd-order Volterra model	0.799	−0.171
PEM model	0.676	−0.312
CR model	0.809	−0.153
CR model (decomposed approach)	0.799	−0.179

over all frequencies, performance considerations will usually require that the loop gain be minimized only over a specific bandwidth. Hence, it is common to see in the linear case (and evidenced in this case study) problems where a superior control-relevant model is obtained without the need to minimize a small gain condition.

The reduced order model is further validated by testing the admissible input range of the closed-loop system. In *Table 5* the maximum step set point changes, which can be applied to the system to give stable responses and real solution for input u, are obtained based on the simulation. As seen in *Table 5*, setpoint ranges obtained for the control-relevant models meet or exceed those of the second-order Volterra model.

Summary and conclusions

In spite of the recognition that the number of model parameters in Volterra series models poses a major obstacle in practice, little attention has been paid to the control-relevant model reduction problem in the literatures. In this paper, using the nonlinear IMC design as a basis, an explicit expression relating the closed-loop control error to the open-loop modeling error is derived. A control-relevant model reduction problem is then formulated to directly minimize the closed-loop control error subject to the constraint of a reduced order model structure. The structure of the reduced order model is specified by users. It certainly can be determined in a preliminary step using the orthogonal least squares procedure, and the control-relevant method is then used to enhance the closed-loop performance. A more attractive application of the proposed method is that it can be used to generate models with restricted complexity such as the block-structured models, which are amenable to control design and analysis.

Acknowledgements

Financial support from the Dupont Educational Aid Program and the ASU–Honeywell Control System Engineering Program is gratefully acknowledged. We are thankful to Drs Ron Pearson and Tunde Ogunnaike of Dupont and Professor Frank Doyle of Purdue for their assistance in this work.

References

1. Schetzen, M., *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, New York, 1980.
2. Rugh, W. J., *Nonlinear System Theory*. The John Hopkins University Press, Baltimore, 1981.
3. Leontaritis, I. J. and Billings, S. A., Input-output parametric models for nonlinear systems. Part I—deterministic nonlinear systems. Part II—stochastic nonlinear systems. *Int. J. Control.*, 1985, **41**, 303.
4. Rivera, D. E. and Morari, M., Control-relevant model reduction problems for SISO H_2, H_∞, and μ control synthesis. *Int. J. Control*, 1987, **46**, 505.
5. Congalidis, J. P., Richards, J. R. and Ray, W. H., Feedforward and feedback control of a solution copolymerization reactor. *AIChE Journal*, 1989, **35**, 891.
6. Boyd, S. and Chua, L. O., Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions of Circuits and Systems*, 1985, **32**, 1150.
7. Haber, R. and Unbehauen, H., Structure identification of nonlinear dynamic systems—a survey on input/output approaches. *Automatica*, 1990, **26**, 651.
8. Pearson, R. K., Nonlinear input/output modeling. *J. Process Control*, 1995, **4**, 197.
9. McCullough, G. R., Identification of second order volterra series approximations to nonlinear systems. Honors thesis, Purdue University, IN, 1994.
10. Kurth, J. and Rake, H., Identification of nonlinear systems with reduced Volterra-series. In Proceedings of the 10th IFAC Symposium on System Identification, preprint, Vol.1, pp. 143–150. Copenhagen, 1994.
11. Zheng, Q. and Zafiriou, E., Nonlinear system identification for control using Volterra-Laguerre expansions. In Proceedings of the American Control Conference, 1995, Seattle, Washington, 2195.
12. Chen, S., Billings, S. A. and Luo, W., Orthogonal least squares

methods and their application to non-linear system identification. *Int. J. Control*, 1989, **50**, 1873–1896.
13. Doyle III, F. J., Ogunnaike, B. A. and Pearson, R. K., Nonlinear model predictive control using second order Volterra models. *Automatica*, 1995, **31**, 697.
14. Narendra, K. S. and Gallman, P. G., An iterative method for the identification of the nonlinear systems using the Hammerstein model. *IEEE Trans. Aut. Control*, 1966, **12**, 546.
15. Wigren, T., Recursive identification based on the nonlinear Wiener model. Ph.D. dissertation, Uppsala University, Sweden, 1990.
16. Rivera, D. E., Pollard, J. F. and Garcia, C. E., Control-relevant prefiltering: a systematic design approach and case study. *IEEE Transactions on Autom. Cntrl.*, 1992, **37**, 964.
17. Rivera, D. E., Chen, X. and Bayard, D. S., Experimental design for robust process control using Schroeder-phased input signals. In *Proceedings of the American Control Conference*, 1993, San Francisco, California, 895.
18. Zheng, Q. and Zafiriou, E., A local small gain theorem and its use for robust stability of uncertain feedback Volterra systems, presented at Annual AIChE Meeting, St. Louis, MO, 1993.
19. Desoer, C. A. and Vidyasagar, M., *Feedback System: Input-Output Properties*. Academic Press, New York, 1975.
20. Maner, B. R., Doyle III, F. J., Ogunnaike, B. A. and Pearson, R. K., Nonlinear model predictive control of a simulated multivariable polymerization reactor using second-order Volterra models. *Automatica*, 1996, **32**, 1285.