Constrained MIMO dynamic discrete-time modeling exploiting optimal experimental design

Derrick K. Rollins *, Nidhi Bhandari

Department of Chemical Engineering, 2114 Sweeney Hall, Iowa State University, Ames, IA 50011, USA

Received 11 December 2002; received in revised form 20 August 2003; accepted 2 October 2003

Abstract

This article presents a new multiple input, multiple output (MIMO) constrained discrete-time modeling (DTM) approach for dynamic block-oriented processes that does not require the nonlinear steady state characteristics to be known prior to model development. This approach uses an efficient statistical experimental design to provide design points for sequential step tests. The DTM is developed from this data in two stages. In the first stage, the ultimate response (steady state) model is determined from just the ultimate response data of the sequential step tests. In the second stage, the dynamic parameters are estimated under the constraint of the fitted ultimate response model obtained in the first stage. The constrained formulation is given for MIMO Hammerstein and Wiener block-oriented systems. Comparison of the proposed constrained DTM method is made with unconstrained DTM and constrained continuous-time modeling (CTM). Prediction accuracy of the proposed method is significantly better than unconstrained DTM and comparable to constrained CTM for the process studied.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Wiener system; Hammerstein system; ARMAX; NARMAX; Predictive modeling; Dynamic modeling; Block-oriented modeling

1. Introduction

Dynamic predictive models that address nonlinear behavior are essential for optimal operation and control of many process. In recent years, dynamic models have been developed for a wide range of processes [1–4]. This article addresses the class of dynamic modeling that has been described as block-oriented systems, which intersects the class of Volterra models (see Seinfeld and Lapidus [5] and Pearson and Oggunnaike [6]). The strength of the block-oriented approach lies in the use of semi-empirical model forms (also called “gray-box modeling”) to represent dynamic behavior. The two most popular systems are the Hammerstein and Wiener systems. As Fig. 1 illustrates, the first block in the Hammerstein system is the static gain function which is typically nonlinear in the inputs. This function enters the second block consisting of a linear dynamic transfer function. The Wiener system is similar to the Hammerstein system but reverses the order of the blocks as shown in Fig. 2. Its advantages over the Hammerstein system are that it allows each input to have separate dynamics and it is able to treat nonlinear dynamic systems through nonlinear functions of the linear dynamic outputs from the first blocks.

In terms of application, the dominant context of modeling Hammerstein processes has been discrete-time modeling (DTM) which falls into the class of NARMAX (nonlinear auto regressive moving average with exogenous inputs) models. Pearson and Oggunnaike [6] point out that this popularity is related to the discrete environment of digital control, measurement, and sampling. However, this environment can cause some critical drawbacks for discrete-time modeling (DTM) when sampling is infrequent, nonconstant or not online (see Chen and Rollins [7] for details). In addition, a discrete-time method is only able to handle inputs as piecewise constant step functions which can lead to poor prediction when sampling is not fast enough.

Continuous-time modeling (CTM) has seen limited application in block-oriented systems. CTM Volterra modeling has been the most widely used approach for
both Hammerstein and Wiener systems. The critical drawback of this approach has been the lack of closed-form solutions. Rollins et al. [8] introduced a CTM approach for the Hammerstein system that they called H-BEST for Hammerstein block-oriented exact solution technique. This approach provides an exact closed-form solution to a true Hammerstein system. Rollins et al. [8] exploited this solution in the development of a strategy that uses optimal statistical design of experiments (SDOE) to collect data rich in ultimate and dynamic information. The model for this method is developed in two stages. The ultimate response model is developed in the first stage using only the ultimate response data. In the second stage, the dynamic model is developed while constrained to give the ultimate response behavior from fitted ultimate response model.

CTM methods for Wiener systems are not common in the control literature. Recently, Bhandari and Rollins [9] introduced their continuous-time Wiener analog to H-BEST that they called W-BEST for Wiener block-oriented exact solution technique. It has the exact same properties as H-BEST but for the Wiener system. That is, it is a constrained (for ultimate response) two stage method that fully exploits SDOE.

Recently, Pearson and Pottmann [10] introduced constrained DTM for block-oriented systems including H-BEST for Hammerstein block-oriented exact solution technique. This approach provides an exact closed-form solution to a true Hammerstein system. Rollins et al. [8] exploited this solution in the development of a strategy that uses optimal statistical design of experiments (SDOE) to collect data rich in ultimate and dynamic information. The model for this method is developed in two stages. The ultimate response model is developed in the first stage using only the ultimate response data. In the second stage, the dynamic model is developed while constrained to give the ultimate response behavior from fitted ultimate response model.

Both Hammerstein and Wiener systems. The critical drawback of this approach has been the lack of closed-form solutions. Rollins et al. [8] introduced a CTM approach for the Hammerstein system that they called H-BEST for Hammerstein block-oriented exact solution technique. This approach provides an exact closed-form solution to a true Hammerstein system. Rollins et al. [8] exploited this solution in the development of a strategy that uses optimal statistical design of experiments (SDOE) to collect data rich in ultimate and dynamic information. The model for this method is developed in two stages. The ultimate response model is developed in the first stage using only the ultimate response data. In the second stage, the dynamic model is developed while constrained to give the ultimate response behavior from fitted ultimate response model.

Both Hammerstein and Wiener systems. The critical drawback of this approach has been the lack of closed-form solutions. Rollins et al. [8] introduced a CTM approach for the Hammerstein system that they called H-BEST for Hammerstein block-oriented exact solution technique. This approach provides an exact closed-form solution to a true Hammerstein system. Rollins et al. [8] exploited this solution in the development of a strategy that uses optimal statistical design of experiments (SDOE) to collect data rich in ultimate and dynamic information. The model for this method is developed in two stages. The ultimate response model is developed in the first stage using only the ultimate response data. In the second stage, the dynamic model is developed while constrained to give the ultimate response behavior from fitted ultimate response model.
Wiener and Hammerstein systems. The main limitation of this work was the requirement of known nonlinear steady state (i.e., ultimate response) characteristics. In addition, the formulation and their examples were demonstrated for single input, single output (SISO) processes only.

In this article we present constrained DTM for Hammerstein and Wiener systems that overcomes the aforementioned limitations of the approach introduced by Pearson and Pottmann. More specifically, the proposed approach does not require known nonlinear ultimate response behavior and we give the MIMO model formation for these two block-oriented systems. We developed the proposed DTM approach from the MIMO continuous-time H-BEST and W-BEST methods mentioned above. Like these CTM approaches, our proposed DTM is done in two stages and uses data from a SDOE which contains information rich in ultimate response and dynamic response data. For a discussion on the superiority of SDOE over the popular pseudorandom sequence design see Rollins et al. [11] and Bhandari and Rollins [9].

The ultimate response data from the step-test type SDOE are used to model the ultimate response behavior in the first stage modeling. In the second stage, all the dynamic parameters are estimated simultaneously using all the data, and constraining the outputs to give the ultimate response model obtained in first stage. For modeling SISO systems this step (i.e., the second stage) is exactly the same as given by Pearson and Pottmann. However, for modeling MIMO systems, it follows the CTM procedures of Rollins et al. [12] (for H-BEST) and Bhandari and Rollins [9] (for W-BEST). Thus, our two stage MIMO proposed methods are discrete-time analogs of H-BEST and W-BEST.

One should note that there are three critical differences between our proposed approach and the work of Pearson and Pottmann. The first one is that we do not require any prior knowledge of static behavior. We can rely totally on the data collected during the experiment to provide all the knowledge to determine the $f_i(u)$'s in Fig. 1 or the $f_i(v)$'s in Fig. 2. As described above, this is our first stage modeling which has no counter-part in the procedure proposed by Pearson and Pottmann. Secondly, Pearson and Pottmann did not address the issue of experimental design (actually, for single input, single output this is not really much of an issue) and our proposed procedure is embedded in the use of SDOE. Thirdly, their work is restricted to SISO but our proposed method is also for MIMO processes. Note that, with given ultimate response behavior for each output, one could extend the Pearson and Pottmann model formulation very easily to single input, multiple output by replacing one output with another. However, the extension to multiple input systems is considerably more challenging, especially for the Wiener system. First, to extend the SISO Wiener approach to multiple input, single output (MISO) one must provide a way to obtain the static nonlinear function of the $v_{ij}$'s, i.e., the $f_i(v)$'s. Relying on a priori knowledge for just one input and one output can be a challenge on its own, but the challenge grows substantially as the number of inputs increase and to a lesser degree as the number of outputs increase. Thus, a MIMO approach does not seem practical without two stage modeling, the first stage of which uses experimental data to obtain ultimate response behavior as in the proposed approach. Secondly, in our development of the constrained discrete-time MIMO model formulations we found it to be quite challenging and requiring meticulous and careful attention to detail. Our creation of the views given in Figs. 1 and 2 (especially Fig. 2 for the Wiener system) aided considerably in this development.

Since the proposed approach can rely completely on the experimental design for total model development, it is important to clearly articulate its scope. Hence, the scope is as follows:
1. The structure of the system can be adequately modeled by either Hammerstein or Wiener. All other block-oriented structures are outside the scope of this article. See [13] for details on general structural identification.

2. Complete model form identification is within the scope of this article. That is, the scope includes the determination of the forms for the static functions and the linear dynamic functions, and the estimation of all parameters for these forms.

In this article we present and evaluate the proposed method as follows. The MIMO constrained DTM structures are presented as analogs to W-BEST and H-BEST in Sections 2 and 3, respectively. Following this, in Section 4, we illustrate the methodology in an example using a true Wiener process. In Sections 5 and 6, we describe the physical process and the model building procedure we used in a comparative study to evaluate the proposed method. The results of this study are given in Section 7. We close, in Section 8, with a summary and concluding remarks.

2. The constrained MIMO Wiener model

This section illustrates the formulation for our proposed MIMO discrete-time constrained Wiener method that we will call C-WDTM. Working from the description of the Wiener system given in Fig. 2 with $q$ outputs and $p$ inputs, a general deterministic mathematical model (i.e., the expectation form) is given by Eqs. (1) and (2):

$$a_{ij,n} \frac{d^2 v_{ij}(t)}{dt^2} + a_{ij,n-1} \frac{d^{n-1} v_{ij}(t)}{dt^{n-1}} + \cdots + a_{ij,1} \frac{dv_{ij}(t)}{dt} + v_{ij}(t) = b_{ij,m} \frac{d^m u_{ij}(t)}{dt^m} + b_{ij,m-1} \frac{d^{m-1} u_{ij}(t)}{dt^{m-1}} + \cdots + b_{ij,1} \frac{du_{ij}(t)}{dt} + u_{ij}(t)$$

$$\eta_{ij}(t) = f_i(v_{ij}(t))$$

(1)

(2)

where $i$ refers to the output with $i = 1, \ldots, q$, $j$ refers to the input with $j = 1, \ldots, p$, and $v_i(t) = [v_{ij1}, v_{ij2}, \ldots, v_{ijp}]^T$. Note that, for simplicity, Eq. (1) is written without dead time and there are no restrictions placed on the static function given by Eq. (2). The complete model, under sampling, for the scope of the proposed methodology is:

$$y_{i\ell,t} = \eta_{i\ell} + e_{i\ell,t}$$

(3)

where

$$e_{i\ell,t} \sim N(0, \sigma^2_{e}) \quad \forall \ell$$

(4)

$y_{i\ell,t}$ is the $\ell$th measurement of output $i$ taken at time $t$ and $e_{i\ell,t}$ is the corresponding error term, $\ell = 1, \ldots, N$. In addition, we are making the common assumption under least squares estimation for the error term. That is, we are assuming that the measurements are stochastically independent of one another.

Next we convert Eq. (1) to an approximate discrete form (see [14] for assistance) in Eq. (5) below.

$$v_{i\ell,j} = \delta_{ij,1} v_{i\ell,j-1} + \delta_{ij,2} v_{i\ell,j-2} + \cdots + \delta_{ij,n} v_{i\ell,j-n} + \alpha_{ij,1} u_{i\ell,j-1} + \alpha_{ij,2} u_{i\ell,j-2} + \cdots + \alpha_{ij,m} u_{i\ell,j-m}$$

(5)

When $m = 0$, Eq. (5), can be written as

$$v_{i\ell,j} = \sum_{k=1}^{n} \delta_{ij,k} v_{i\ell,j-k} + \alpha_{ij,1} u_{i\ell,j-1}$$

(6)

and when $m > 0$, Eq. (5) can be written as

$$v_{i\ell,j} = \sum_{k=1}^{n} \delta_{ij,k} v_{i\ell,j-k} + \sum_{l=1}^{m} \alpha_{ij,l} u_{i\ell,j-l} + \alpha_{ij,m} u_{i\ell,j-m} + \alpha_{ij,m+1} u_{i\ell,j-m+1}$$

(7)

At steady state, $v_{i\ell,j} = v_{i\ell,j-1} = v_{i\ell,j-2} = \cdots$ and $u_{i\ell,j} = u_{i\ell,j-1} = u_{i\ell,j-2} = \cdots$. Therefore,

$$v_{i\ell,j} = \delta_{ij,1} v_{i\ell,j} + \delta_{ij,2} v_{i\ell,j} + \cdots + \delta_{ij,n} v_{i\ell,j} + \alpha_{ij,1} u_{i\ell,j} + \alpha_{ij,2} u_{i\ell,j} + \cdots + \alpha_{ij,m} u_{i\ell,j} + \alpha_{ij,m+1} u_{i\ell,j}$$

(8)

$$\Rightarrow v_{i\ell,j} = \frac{\alpha_{ij,1} + \alpha_{ij,2} + \cdots + \alpha_{ij,m} + \alpha_{ij,m+1} u_{i\ell,j}}{1 - \delta_{ij,1} - \delta_{ij,2} - \cdots - \delta_{ij,n}} \cdot u_{i\ell,j}$$

(9)

$$K = \frac{\alpha_{ij,1} + \alpha_{ij,2} + \cdots + \alpha_{ij,m} + \alpha_{ij,m+1}}{1 - \delta_{ij,1} - \delta_{ij,2} - \cdots - \delta_{ij,n}} = 1$$

(10)

since the gain in Eq. (1) is unity. Solving for $\alpha_{ij,m+1}$ from Eq. (10) gives,

$$\alpha_{ij,m+1} = 1 - \delta_{ij,1} - \delta_{ij,2} - \cdots - \delta_{ij,n} - \alpha_{ij,1} - \alpha_{ij,2} - \cdots - \alpha_{ij,m}$$

(11)

When $m = 0$, Eq. (11) reduces to

$$\alpha_{ij,1} = 1 - \sum_{k=1}^{n} \delta_{ij,k}$$

(12)

Substituting Eq. (12) into Eq. (6) gives the MIMO constrained form of $v_{i\ell,j}$ when $m = 0$.

$$v_{i\ell,j} = \sum_{k=1}^{n} \delta_{ij,k} v_{i\ell,j-k} + \left[1 - \sum_{k=1}^{n} \delta_{ij,k}\right] u_{i\ell,j-1}$$

(13)

Similarly, substituting Eq. (11) into Eq. (7) gives the MIMO constrained form of $v_{i\ell,j}$ when $m > 0$.

$$v_{i\ell,j} = \sum_{k=1}^{n} \delta_{ij,k} v_{i\ell,j-k} + \sum_{l=1}^{m} \alpha_{ij,l} u_{i\ell,j-l} + \left[1 - \sum_{k=1}^{n} \delta_{ij,k}\right] u_{i\ell,j-(m+1)}$$

(14)
The first step in developing a C-WDTM is the selection of the experimental design. For block-oriented modeling, this selection can be wisely made based strictly on knowledge or a priori assumptions of ultimate response behavior in the input space as described in Rollins et al. [11]. That is, this selection can be strictly based on the assumed forms of the \( f_i(v_i)'s \). SDOE selects the design that maximizes information content using a priori assumptions or knowledge under a specified optimality criterion. This means that it will provide the design with minimum design points and maximum ability to accurately estimate model coefficients. For example, if one correctly assumes the \( f_i(v_i)'s \) in the input space have approximate quadratic behavior, and only second order interactions are significant among input variables, then fractional factorial designs with modifications to include efficient estimation of quadratic behavior will provide optimal designs under these conditions. Two widely used designs meeting these criteria are central composite designs (CCD) and Box-Behnken designs (BBD). Both these designs, as well as many others, can be obtained from commonly used general purpose statistical software packages. Note that, under these conditions, second order least squares model forms including terms for the interactions will provide the model forms for the \( f_i(v_i)'s \). Later we will illustrate this approach with second order model forms in two examples.

As stated previously, this approach places no restrictions on static behavior, i.e., the \( f_i(v_i)'s \). If the ultimate response behavior is highly complex and nonlinear, multiple regression models will not be adequate to describe this behavior. In this case sufficient knowledge will be required for the forms of the \( f_i(v_i)'s \). This knowledge can then be exploited in optimal SDOE to obtain an efficient design to accurately estimate the parameters of the \( f_i(v_i)'s \). This approach has been used widely by chemical engineers in semi-theoretical modeling to obtain unknown physical coefficients (i.e., parameters) and examples can be found in [15–17].

After selecting the experimental design, the next step consists of running the design points as a series of sequential step tests. The challenge in developing a MIMO Wiener model is determining the \( v_{ij}'s \) and the \( f_i(v_i)'s \) when the \( v_{ij}'s \) are not observable; i.e., when only the inputs (the \( u_i's \)) and the measured outputs (the \( y_j's \)) are observable. Our proposed two stage MIMO constrained modeling approach overcomes this difficulty by first modeling the static behavior and then modeling the dynamic behavior. More specifically, in the first stage, the nonlinear functions of the \( v_{ij}'s \) are determined and in the second stage the dynamic models to produce the \( v_{ij}'s \) are found. In the first stage we exploit a fundamental result to overcome the challenge of obtaining the \( f_i's \) when the \( v_{ij}'s \) are not observable. More specifically, Eq. (1) tells us that at steady state, \( v_{ij}(\infty) = u_i(\infty) \). Let us write this as \( v_{ij}^\infty = u_i^\infty \). Thus, at steady state,

\[
\eta_i^\infty = f_i(v_{ij}^\infty) = f_i(u_i^\infty)
\]

where \( \eta_i^\infty \) is the true ultimate response of \( \eta_i(t) \) given the corresponding input vector \( u^\infty = [u_1^\infty, u_2^\infty, \ldots, u_p^\infty] \) at this steady state. Therefore, the first stage modeling consists of the idea of using the ultimate response data from the sequential step tests to obtain the models for the static behavior. Note that this would not be possible if the experimental design did not produce adequate ultimate response data. Thus, the idea to run the design points as a series of step tests is linked to this idea of obtaining the \( f_i's \) by exploiting Eq. (15).

In the second stage we rely on the hypothesis that the ultimate response functions (i.e., the \( f_i's \)) that we obtained from the inputs in the first stage are adequate to describe the static portion of the response under dynamic behavior. The second stage modeling seeks to verify this hypothesis by using all the data from the experimental design to obtain accurate \( G_{ij}'s \) (see Fig. 2) to produce the \( v_{ij}'s \) that are used with the \( f_i's \) determined in the first stage to accurately predict the \( \eta_i's \). More specifically, this modeling stage consists of an iterative process of first selecting \( m \) and \( n \), using Eq. (13) \((m = 0)\) or Eq. (14) \((m > 0)\) with Eq. (2) and the data from the experimental design to find least square estimates of the \( \delta_j's \) and the \( \omega_{ij}'s \). We have found the Solver routine in Microsoft Excel to be sufficient to carry out this dynamic fitting step. If the fit is poor, the process is repeated with a different set of \( m \) and \( n \) until an adequate fit is found or until it is concluded that one cannot be found. Note that, the visual attributes of the step tests provide assistance in the selection of \( m \) and \( n \). After successful completion of this step, the final step consists of using an arbitrary input sequence to test the predictive accuracy of the model.

3. The constrained MIMO Hammerstein model

This section illustrates the formulation for our proposed MIMO discrete-time constrained Hammerstein method that we will call C-HDTM. Working from the description of the Hammerstein system given in Fig. 1 with \( q \) outputs and \( p \) inputs, a general deterministic mathematical model is given by Eqs. (16) and (17) below:

\[
ad_{n+1} \frac{d^n \eta_i(t)}{dr^n} + a_{n+1-1} \frac{d^{n-1} \eta_i(t)}{dr^{n-1}} + \cdots + a_1 \frac{d \eta_i(t)}{dr} + \eta_i(t) \\
= b_{m+1} \frac{d^m v_i(t)}{dr^m} + b_{m+1-1} \frac{d^{m-1} v_i(t)}{dr^{m-1}} + \cdots + b_1 \frac{dv_i(t)}{dr} + v_i(t)
\]

\[
v_i(t) = f_i(u(t))
\]
where $i = 1, \ldots, q$ and $\mathbf{u}(t) = [u_1, u_2, \ldots, u_p]^T$. Note that, for simplicity, as in the case for the Wiener model, Eq. (16) is written without dead time and again, there are no restrictions placed on the static function given by Eq. (17). Next we convert Eq. (16) to an approximate discrete form as shown in Eq. (18) below.

$$
\eta_{i,t} = \delta_{i,1}\eta_{i,t-1} + \delta_{i,2}\eta_{i,t-2} + \cdots + \delta_{i,n}\eta_{i,t-n} + \omega_{i,1}v_{i,t-1} + \omega_{i,2}v_{i,t-2} + \cdots + \omega_{i,m}v_{i,t-m} + \omega_{i,m+1}v_{i,t-(m+1)}
$$  

(18)

When $m = 0$, Eq. (18), can be written as

$$
\eta_{i,t} = \sum_{k=1}^{n} \delta_{i,k}\eta_{i,t-k} + \omega_{i,1}v_{i,t-1}
$$  

(19)

and when $m > 0$, Eq. (18) can be written as

$$
\eta_{i,t} = \sum_{k=1}^{n} \delta_{i,k}\eta_{i,t-k} + \sum_{l=1}^{m} \omega_{i,l}v_{i,t-l} + \omega_{i,m+1}v_{i,t-(m+1)}
$$  

(20)

At steady state, $\eta_{i,t} = \eta_{i,t-1} = \eta_{i,t-2} = \cdots$ and $v_{i,t} = v_{i,t-1} = v_{i,t-2} = \cdots$. Therefore,

$$
\eta_{i,t} = \delta_{i,1}\eta_{i,t} + \delta_{i,2}\eta_{i,t} + \cdots + \delta_{i,n}\eta_{i,t} + \omega_{i,1}v_{i,t} + \omega_{i,2}v_{i,t} + \cdots + \omega_{i,m}v_{i,t} + \omega_{i,m+1}v_{i,t}
$$  

(21)

$$
\Rightarrow \quad \eta_{i,t} = \frac{\omega_{i,1} + \omega_{i,2} + \cdots + \omega_{i,m} + \omega_{i,m+1}}{1 - \delta_{i,1} - \delta_{i,2} - \cdots - \delta_{i,n}}v_{i,t}
$$

(22)

where $K = \frac{\omega_{i,1} + \omega_{i,2} + \cdots + \omega_{i,m} + \omega_{i,m+1}}{1 - \delta_{i,1} - \delta_{i,2} - \cdots - \delta_{i,n}} = 1$.

Solving for $\omega_{i,m+1}$ from Eq. (23) gives

$$
\omega_{i,m+1} = 1 - \delta_{i,1} - \delta_{i,2} - \cdots - \delta_{i,n} - \omega_{i,1} - \omega_{i,2} = \cdots = \omega_{i,m}
$$

(24)

When $m = 0$, Eq. (24) reduces to

$$
\omega_{i,1} = 1 - \sum_{k=1}^{n} \delta_{i,k}
$$

(25)

Substituting Eq. (25) into Eq. (19) gives the MIMO constrained form of $\eta_{i,t}$ when $m = 0$.

$$
\eta_{i,t} = \sum_{k=1}^{n} \delta_{i,k}\eta_{i,t-k} + \left[1 - \sum_{k=1}^{n} \delta_{i,k}\right]v_{i,t-1}
$$

(26)

Similarly, substituting Eq. (24) into Eq. (20) gives the MIMO constrained form of $\eta_{i,t}$ when $m > 0$.

$$
\eta_{i,t} = \sum_{k=1}^{n} \delta_{i,k}\eta_{i,t-k} + \sum_{l=1}^{m} \omega_{i,l}v_{i,t-l} + \left[1 - \sum_{k=1}^{n} \delta_{i,k}\right]v_{i,t-(m+1)}
$$

(27)

The procedure to obtain a C-HDTM model is exactly the same as it is to obtain a C-WDTM with the only difference being the equations used. The ultimate response functions are determined exactly the same way, except in this case Eq. (14) tells us that at steady state $\eta_{i}^{\infty} = v_{i}^{\infty}$. Therefore,

$$
\eta_{i}^{\infty} = v_{i}^{\infty} = f_i(\mathbf{u}^{\infty})
$$  

(28)

Note the similarity and difference between Eqs. (28) and (15). The similarity (i.e., $\eta_{i}^{\infty} = f_i(\mathbf{u}^{\infty})$) tells us that we determine and use the same ultimate response functional relationship without consideration of the block structure. In contrast, the dissimilarity (i.e., $v_{i}^{\infty} = f_i(\mathbf{u}^{\infty})$ versus $f_i(\mathbf{v}^{\infty}) = f_i(\mathbf{u}^{\infty})$) tells us that we relate them differently to the hypothetical intermediate variables (i.e., the $v$’s) depending on the block structure. Here, for C-HDTM we obtain $f_i(\mathbf{u}(t))$, i.e., Eq. (17). Furthermore, second stage modeling (i.e., determining the linear dynamic functions) here consists of an iterative process of first selecting $m$ and $n$, using Eq. (26) ($m = 0$) or Eq. (27) ($m > 0$) with Eq. (17) and the data from the experimental design to find least square estimates of the $\delta$’s and the $\omega$’s. Next we illustrate the procedures to obtain both C-WDTM and C-HDTM models in a mathematical example to provide better understanding of the proposed modeling methods.

### 4. Mathematical process example

The purpose of this section is to illustrate, in a step by step fashion, the procedures to obtain C-WDTM and C-HDTM models. Although this example will demonstrate excellent fitting, it is not the purpose of this example to draw any conclusions on the accuracy of the proposed methods. The “true process” for this illustration is described by the two input, single output Wiener system given by Eqs. (29) and (30) below.

$$
\eta_{i}(t) = f_i(v_i(t)) = a_i v_i(t) + 10(1 - e^{-a_i t})
$$  

(29)

$$
\tau_{j1}\tau_{j2}\tau_{j3}\tau_{j4}\frac{d^4 v_{ij}(t)}{dt^4} + (\tau_{j1}\tau_{j2}\tau_{j3} + \tau_{j1}\tau_{j2}\tau_{j4} + \tau_{j1}\tau_{j3}\tau_{j4})
$$

$$
+ \tau_{j2}\tau_{j3}\tau_{j4}\frac{d^3 v_{ij}(t)}{dt^3} + (\tau_{j1}\tau_{j2} + \tau_{j1}\tau_{j3} + \tau_{j1}\tau_{j4} + \tau_{j2}\tau_{j3})
$$

$$
+ \tau_{j2}\tau_{j3}\tau_{j4}\frac{d^2 v_{ij}(t)}{dt^2} + (\tau_{j1} + \tau_{j2} + \tau_{j3} + \tau_{j4}) \frac{dv_{ij}(t)}{dt}
$$

$$
+ v_{ij}(t) = u_{ij}(t)
$$  

(30)

where $i = 1, j = 1, 2, \eta_{i}(t)$ is the true value of the output $i$ at time $t$, all initial conditions and derivatives are 0, $\tau_{j1} = 10$, $\tau_{j2} = 6$, $\tau_{j3} = 3$, $\tau_{j4} = 1$, $\tau_{j5} = 12$, $\tau_{j6} = 7$, $\tau_{j7} = 2$, $\tau_{j8} = 1$, $a_1 = 5$, $a_2 = 2$, all inputs and outputs are deviation variables, and the time unit is minutes. Since there is only one output ($i = 1$), for simplicity the subscript $i$ will be dropped. The measured output, $y(t)$, follows $y(t) = \eta(t) + \epsilon$, where $\epsilon$ is independent and normally distributed with mean 0 and variance, $\sigma^2 = 4$. The objective is to determine an accurate predictive model.
for this process limiting the choice to either a discrete-
time Hammerstein or Wiener model. Note that the true
process (Eqs. (29) and (30)) is assumed to be completely
unknown to the modeler except that it takes about 80
min for the process get reasonably close to steady state
after an input change. We obtained the solution of the
true response for this process (i.e., the values for \( \eta(t) \))
using numerical integration.

As discussed previously, the first step in this approach
is to select an experimental design. The given informa-
tion is: sampling rate = one sample every five (5) min-
utes, the operating range is \( u_1: \pm 10 - 10 \) and \( u_2: \pm 0.5 - 0.5 \).
The modeler assumes that a second order (including the
two-way interaction) ultimate response model is ade-
quate to approximate the nonlinear static behavior.
Based on this assumption and the number of inputs, the
modeler selects a full factorial 32 (three levels, two in-
puts) design and decides to replicate the center point
twice for a total of 10 design points (i.e., trials or runs).
The modeler randomizes these trials and runs the
training input sequence shown in Fig. 3. The output
response corresponding to the sequence in Fig. 3 is
shown in Fig. 4. Note that, although the true response,
\( \eta(t) \), is shown, the modeler only observes the measured
response, \( y(t) \), sampled every 5 min. In addition, ignore
“Fitted C-WDTM” for now. This is the fitted Wiener
model that the modeler will obtain later (it is shown here
to conserve space later). Thus, at this point, the modeler
only has the “data” in Fig. 4.

The next step is to find the best dynamic model. The
data set is then over-damped and selects appropriate candidates (i.e., \( n = 1, 2, 3 \) or 4; and \( m = 0 \)) for the dynamic forms. The \( n\)th order C-
HDTM model in (i.e., combining Eq. (17) with Eq. (26)) applicable to this situation is:

\[
\hat{\eta}_t = \sum_{k=1}^{n} \delta_k y_{t-k} + \left[ 1 - \sum_{k=1}^{n} \delta_k \right] v_{t-1} - \sum_{k=1}^{n} \delta_k y_{t-k} + \left[ 1 - \sum_{k=1}^{n} \delta_k \right] (\hat{\beta}_1 u_{1,t-1} + \hat{\beta}_2 u_{2,t-1} + \hat{\beta}_3 (u_{2,t-1})^2)
\]

(32)

Recall that the “\( i \)” in Eq. (32) has been dropped for
simplicity. The unknowns are the \( \delta_k \)’s. The Modeler uses
the Solver routine in Excel with Eq. (32) and all the data in
Fig. 4 to find the least squares estimates of the \( \delta_k \)’s for
various values of \( n \). The order \( n = 4 \) gives the best fit and
this fitted C-HDTM model is shown in Fig. 5. Note that,
the fitted values are shown as discrete points because
discrete models only determine discrete values.

The \( n\)th order C-WDTM (i.e., Eqs. (2) and (13)), in
estimation form, applicable to this situation is:

\[
\hat{\eta}_t = \hat{\beta}_1 \hat{v}_{1,t} + \hat{\beta}_2 \hat{v}_{3,t} + \hat{\beta}_3 (\hat{v}_{2,t})^2 \text{ where }
\]

\[
\hat{v}_{j,t} = \sum_{k=1}^{n} \delta_{jk} \hat{v}_{j,t-k} + \left[ 1 - \sum_{k=1}^{n} \delta_{jk} \right] u_{j,t-1}, \quad j = 1, 2.
\]

(33)

Again, the Modeler uses the Solver routine in Excel but
this time with Eq. (33) and all the data in Fig. 4 to find
the least squares estimates of the \( \delta_{jk} \)’s for various values
of \( n \). The order \( n = 4 \) gives the best fit and this fitted C-
WDTM model is shown in Fig. 4. Comparing the C-
HDTM model in Fig. 5 with the C-WDTM model in

Fig. 3. The training input sequence for \( u_1(t) \) and \( u_2(t) \) used in this study from the \( 3^2 \) experimental design.
Fig. 4, the modeler finds the C-WDTM model to fit better by observation.

Next the modeler tests the C-WDTM from an arbitrary input test sequence (not shown) and its performance can be seen in Fig. 6. As shown the fit is excellent and the modeler accepted this model as an accurate approximation of the true process.

In this example, measurement variability (i.e., \( \sigma^2 \) was >0) or noise was added to simulate a process under the full assumptions of the proposed method. However, under least squares, as in the proposed method, it is important to note that the size of \( \sigma^2 \) does not prohibit the accurate estimation of model parameters. That is, we could have added any level of noise and obtained accurate estimates by either increasing the sampling rate or increasing the number of design points (i.e., step tests). Theoretically, this can be seen from an analysis of the mathematical form of the estimation variance [18], Eq. (34), given below.

\[
\text{Variance}(\hat{\theta}) = \frac{\sigma^2}{D^T D}
\]  

(34)

where \( \hat{\theta} \) is the least squares estimate of \( \theta \), \( \sigma^2 \) is the variance of the output and error term (see Eqs. (3) and (4)) and \( D \) is a derivative vector with \( N \) elements, such that, the \( i \)th element represents the partial derivative of the output with respect to \( \theta \) and evaluated at the \( i \)th sample. Therefore, we see that as \( N \to \infty \), \( D^T D \) (a sum of squared values) \( \to \infty \), and hence, \( \text{Variance}(\hat{\theta}) \to 0 \). In practical terms this means that any level of \( \sigma^2 \) can be nullified by increasing the number of elements in \( D \), that is, \( N \). Thus, to illustrate the ability of the proposed method to obtain accurate estimates in a hypothetical mathematical simulation, we do not feel it is necessary to add noise when sampling is a matter of choice as in the case to follow. If one is interested in how the accuracy of the parameters varies for a particular simulated process, Eq. (34) can provide this information. In addition, Eq. (34) could prove useful in controlling estimation error in modeling real processes. More specifically, one could explore the sensitivity of the parameter estimates from assumed model forms, sampling, values of \( \sigma \), parameter values, and experimental designs before running the plant test (i.e., the experiment) and select a sampling rate and design that provides a certain degree of robustness. A study of this sort is beyond the objectives of this article. Next we apply the proposed methodology to a seven input, five output physical process with significant nonlinear and interactive static behavior and compare it to other methodologies.

5. Simulated CSTR process

This section introduces the mathematical model of the simulated continuous stirred tank reactor (CSTR) used in the comparative study, which was also used by Bhandari and Rollins [9]. This CSTR is illustrated in Fig. 7. Reactants A and B enter the CSTR as two different flow streams and form product C. The second order exothermic reaction taking place in the CSTR gives it strong nonlinear and interactive behavior (see Fig. 8). The process model consists of the overall mass balance, component (A and B) mole balances, the energy balance on the tank contents, and the energy balance on the contents of the jacket (Eqs. (32)-(38)). For details of this process see [9]. Our main objective in applying the proposed method to this physical MIMO process is to determine the extent of accuracy of our discrete-time block-oriented approach to model the physical behavior of the CSTR. That is, we wish to determine how well the expectation functions obtained by C-WDTM can describe the physical nature of the CSTR outputs to changes in inputs. Note that, this
objective is best achieved by setting the measurement variance to 0 (i.e., $r = 0$) since the focus is on the accurate determination of the expectation function. Also, as described earlier if we are able to estimate this behavior accurately, then under least squares, we will be able to accomplish this objective even when $r$ is not 0 as long as sampling is large enough to offset its effect on parameter accuracy. The next section presents the details of the C-WDTM model building for this process.

6. The training phase

In a study prior to the continuous-time modeling work of Bhandari and Rollins [9], we determined that the CSTR process of the previous section more closely follows a Wiener process than a Hammerstein process. Although our Hammerstein model fit well in the ultimate response, it did not capture the dynamic overshoot and inverse response behavior of this process. Therefore, we modeled this process in this work using C-WDTM. Hence, this section illustrates the procedure mentioned previously to obtain C-WDTM models for the CSTR. The first step is the selection of the design based on the a priori assumptions about the effects of input variables on the output variables. The input variables considered in this study are the feed flowrate of A ($q_{Af}$), the feed temperature of A ($T_{Af}$), the feed concentration of A ($C_{Af}$), the feed flowrate of B ($q_{Bf}$), the feed temperature of B ($T_{Bf}$), the feed concentration of B ($C_{Bf}$) and the coolant flowrate ($q_c$) to the jacket. The output variables considered in this study are the concentrations of species A, B, and C in the reactor (i.e., $C_A$, $C_B$, and $C_C$, respectively), the temperature in the reactor ($T$) and the coolant temperature ($T_c$) in the jacket. Thus, in all we have seven (7) inputs and five (5) outputs for this study. For this seven input system, our goal is to choose a design with minimum design points or runs that enables the testing and estimation of all the main effects and second order effects (the two factor interactions and the quadratic effects). This consists of a model with 36 terms. An experimental design meeting these criteria is a three level Box-Behnken design (BBD) with 62 trials or runs including six replicated center points. Since this is a simulation study without noise, the replicated center points are not needed and thus, the selected design has only 56 runs. The three levels for each input variable are designated (i.e., coded) from low to high as -1, 0, and 1. The lower and upper limits on the input variable are selected so as to cover the complete input space. The values for each level for the seven inputs are given in [9]. We would like to make the following comments regarding the ability of this approach to effectively model such a large number of interactive effects (i.e., 21) which seems to be unique in block-oriented dynamic modeling. In this approach, the experiment design, with its property of orthogonality (no pairwise correlations), keeps all the effects, including the interactions, uncorrelated to allow for optimal efficiency and assessment of the significance of these terms. In addition, the number of trials (in this case 56) is quite small relative to the number of terms (36) in the model, which again speaks to the efficiency of SDOE and its use in this approach. The only article that we have found that addresses interactive effects in dynamic block-oriented modeling is the one by Eskinat et al. [3]. However, they did not address optimal experimental design and the efficient estimation of these terms and did not attempt to include them in any application because of their limitations to handle models with a large number of terms.

The second step consists of running the experiment with the input variables changed as step changes from the values for the current run to the values for the next

---

Fig. 7. Schematic of the CSTR.

Fig. 8. The interaction plot for $C_A$ and $T_A$ for the ultimate response of $C_A$. Significant nonlinear effect is seen from the curve-linear behavior and significant interaction is observed by the nonparallelism of the curves.

---

run after the output responses have approximately reached steady state. The time interval between each of those step changes is 5 min in this study. The input or training sequence based on the SDOE design with 56 runs lasted a total of 280 min. The outputs were sampled twice every minute over the course of the SDOE. The SDOE training sequence is shown in Fig. 9.

The third step is the identification of the static gain or the ultimate response function using linear regression techniques. Only the steady-state data for each of the 56 runs are used and the general form of the static gain or ultimate response function obtained for the ith output is given in Eq. (35) below.

\[
f_i(\mathbf{v}^\infty) = f_i(\mathbf{u}) = \begin{align*}
\hat{\beta}_{i,0} + \hat{\beta}_{i,1}q_{AF} + \hat{\beta}_{i,2} C_{AF} + \hat{\beta}_{i,3} T_{AF} + \hat{\beta}_{i,4} T_{BF} \\
+ \hat{\beta}_{i,5} q_{BF} + \hat{\beta}_{i,6} C_{BF} + \hat{\beta}_{i,7} q_c + \hat{\beta}_{i,8} q_{AF}^2 \\
+ \hat{\beta}_{i,9} C_{AF} + \cdots + \hat{\beta}_{i,13} C_{BF} + \hat{\beta}_{i,14} q_c^2 \\
+ \hat{\beta}_{i,15} q_{AF} C_{AF} + \hat{\beta}_{i,16} q_{AF} T_{AF} + \cdots \\
+ \hat{\beta}_{i,34} q_{BF} q_c + \hat{\beta}_{i,35} C_{BF} q_c
\end{align*}
\]

(35)

where \( \mathbf{u}^T = [q_{AF}, C_{AF}, T_{AF}, T_{BF}, q_{BF}, C_{BF}, q_c], \ i = 1, \ldots, 5, \) and the parameter estimates (\( \hat{\beta}_{i,1} \) to \( \hat{\beta}_{i,35} \)) for the ith output are obtained using linear regression. These estimates are given in [9] for all the outputs.

The fourth and final step consists of the estimation of the dynamic parameters, the \( \delta \)'s and the \( \omega \)'s. The form of the chosen dynamic model is second order with lead (i.e., \( n = 2 \) and \( m = 1 \)), which in discrete form is given by Eq. (36) below.

\[
\delta_{ij} = \hat{\delta}_{ij}^0 \hat{v}_{ij-1} + \hat{\delta}_{ij}^1 \hat{v}_{ij-2} + \hat{\omega}_{ij}^0 u_{ij-1} + \hat{\omega}_{ij}^1 u_{ij-2} \\
= \hat{\delta}_{ij}^0 \hat{v}_{ij-1} + \hat{\delta}_{ij}^1 \hat{v}_{ij-2} + \hat{\omega}_{ij}^0 u_{ij-1} + (1 - \hat{\delta}_{ij}^1) \hat{\omega}_{ij}^1 u_{ij-2}
\]

(36)

where \( i = 1, \ldots, 5 \) and \( j = 1, \ldots, 7 \). With fixed estimates obtained in stage 1, the inputs in Eq. (35) were replaced with their corresponding \( v_{ij} \)'s. This form of the equation was combined with Eq. (36) to give Eq. (37) below where the estimates for the \( \delta \)'s and the \( \omega \)'s are found using constrained least squares modeling.

\[
\tilde{\eta}_{ij} = f_i(\mathbf{v}_{ij}) = \begin{align*}
\hat{\beta}_{i,0} + \hat{\beta}_{i,1} \hat{v}_{i1} + \cdots + \hat{\beta}_{i,7} \hat{v}_{i7} + \hat{\beta}_{i,8} (\hat{v}_{i1})^2 + \cdots \\
+ \hat{\beta}_{i,14} (\hat{v}_{i2})^2 + \hat{\beta}_{i,15} (\hat{v}_{i1}) (\hat{v}_{i2}) + \hat{\beta}_{i,16} (\hat{v}_{i1}) \\
\times (\hat{v}_{i3}) + \cdots + \hat{\beta}_{i,35} (\hat{v}_{i6}) (\hat{v}_{i7})
\end{align*}
\]

(37)

where \( i = 1, \ldots, 5 \), and the values of the estimated parameters \( \hat{\beta}_{i,1} \) to \( \hat{\beta}_{i,35} \) are given in [9]. The fit of C-WDTM is shown only for the concentration of species A (\( C_A \)) in Fig. 10. As shown, the agreement between the true and the fitted values, as seen by the high \( R^2 \) values presented in Table 1. Plots for the true and fitted responses for the other outputs are not shown for space consideration.

By combining Eqs. (36) and (37), all the parameters (the \( \beta \)'s in Eq. (37), and the \( \delta \)'s, and \( \omega \)'s in Eq. (36)) were estimated simultaneously for the unconstrained model (WDTM in Table 4). Since there are more parameters in

Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>( R^2 ) values for the outputs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( C_A )</td>
</tr>
<tr>
<td>C-WDTM</td>
<td>99.4</td>
</tr>
<tr>
<td>WDTM</td>
<td>99.6</td>
</tr>
</tbody>
</table>
the unconstrained model, it is able to explain greater variation in the data as compared to the constrained model. This is seen in the higher $R^2$ values for the unconstrained model also given in Table 1.

7. The testing phase

In this section the accuracy of the predictions for C-WDTM and WDTM are compared for the five outputs considered in this study. We also compare their prediction performance with the continuous-time W-BEST, developed in Bhandari and Rollins [9]. To generate the test sequence, all the inputs are changed arbitrarily (i.e., randomly) as a series of step changes shown in Fig. 11. The predictions for $C_A$, $C_B$, and $C_C$ are given in

Fig. 11. The input test sequences used in the comparison study.

Fig. 12. The true, C-WDTM and WDTM responses for $C_A$ to the input sequences in Fig. 11.

Fig. 13. The true, C-WDTM and WDTM responses for $C_B$ to the input sequences in Fig. 11.

Fig. 14. The true, C-WDTM and WDTM responses for $C_C$ to the input sequences in Fig. 11.
Figs. 12–14, respectively, for C-WDTM and WDTM. The predictions for the tank and coolant temperatures are given in Figs. 15 and 16, respectively, for C-WDTM and WDTM.

As Figs. 12–16 show, the predictions from C-WDTM more closely follow the process for all the outputs. The performance of the WDTM is clearly worse than C-WDTM. To quantitatively assess the extent of agreement between the true responses and the predicted responses, we define a term called the sum of squared prediction error (SSPE) given by Eq. (38) below:

$$\text{SSPE} = \sum_{k=1}^{M} (\eta_k - \hat{\eta}_k)^2$$  \hspace{1cm} (38)

where $M$ is the total number of equally spaced sampling points used over the testing interval, $\eta_k$ is the true response and $\hat{\eta}_k$ is the predicted response. For this study $M = 600$. The smaller the SSPE, the more accurate the model. The SSPE values for DTMs and W-BEST (from [9]) are summarized in Table 2.

The SSPE values for C-WDTM and W-BEST are very similar though W-BEST is slightly smaller (i.e., better) in most cases. The SSPE values for WDTM are on an average about 40% higher than C-WDTM. Hence, the use of constraints along with the two stage estimation procedure appears to greatly enhance modeling accuracy.

### 8. Concluding remarks

This article presented a two stage, constrained, MIMO discrete-time block-oriented modeling method that uses statistical design of experiments (SDOE) to generate the data. The scope of this approach currently includes Hammerstein and Wiener modeling and this
articles gives the MIMO DTM formulations for both. This approach does not require any known or assumed structural information before data collection and allows one to fully develop MIMO Hammerstein and Wiener DTM’s from the experimental data alone. Another attractive attribute of this approach is its ability to efficiently address the modeling of interaction terms for systems with several input variables. A critical strength of this approach is the use of SDOE which provides efficient information to estimate ultimate response and dynamic response behavior. For readers not familiar with or needing assistance with SDOE we recommend the book by Montgomery [19].

We believe that the second order à priori assumption will be practical for many applications in process control because the control environment tends to limit the operability range. Hence, even when process behavior is highly nonlinear, by limiting process variables to small ranges, second order behavior can be a reasonable approximation. The popular SDOE methods such as Box Behnken and Central Composite designs are D-optimal for second order behavior. However, as we discussed in Section 2, when process behavior is more complex than second order, knowledge of the behavior can be exploited to obtain optimal designs [20].

As indicated, the proposed modeling method falls under nonlinear least squares estimation which assumes white noise for the error term. A future step in this research is the extension of this method to address unmeasured disturbances that can cause the error term to exhibit serial correlation. This estimation problem is quite challenging because the estimation of the model parameters can be severely influenced in the presence of stochastic serial correlation. In addition, to obtain accurate predictions, it will likely require accurate estimation of the serial correlation parameters simultaneously with the parameters in the expectation function. As this research progresses, the ability to treat more practical situations will increase.

The proposed constrained DTM was shown to give excellent and superior results to unconstrained DTM and comparable performance to the CTM W-BEST for the process studied. Thus, we recommend this approach for block-oriented modeling when the assumption of white noise is valid. In addition, we are currently developing modeling methods for other types of block-oriented systems.

Acknowledgements

The authors wish to acknowledge partial support of this work by the Minnesota Mining and Manufacturing Company (3M), Saint Paul, Minnesota.

References