Computers chem. Engng, Vol. 18, No. 1, pp. 55–69, 1994
Printed in Great Britain.

0098-1354/94 $6.00 + 0.00

IDENTIFICATION OF CHEMICAL PROCESSES USING CANONICAL VARIATE ANALYSIS

C. D. SCHAPER,[1] W. E. LARIMORE,[2] D. E. SEBORG[1] and D. A. MELLICHAMP[1]†
[1]Department of Chemical and Nuclear Engineering, University of California, Santa Barbara, CA 93106, U.S.A.
[2]Adaptics, Inc., 40 Fairchild Drive, Reading, MA 01867, U.S.A.

(Received 20 January 1992; final revision received 6 May 1993; received for publication 3 June 1993)

Abstract—A method of identification of linear input–output models using canonical variate analysis (CVA) is developed for application to chemical processes. This approach yields both a process model and a nonparametric description of model uncertainty, utilizing CVA for selection of a state coordinate system that optimally relates past inputs and outputs to future outputs. Regression procedures are then used for estimation of the state-space model parameters, and the Akaike Information Criterion (AIC) is used to determine the model order. The primary computations involve singular value decompositions which are numerically stable and accurate.

The effectiveness of the CVA approach is first evaluated with simulated chemical processes that exhibit most of the practical problems encountered by existing system identification methods: nonlinear dynamics, unknown model orders and time delays, nonminimum phase dynamics, partial stiffness (requiring two time-scale approaches when other identification methods are used), low input excitation in some frequency bands, and measurement and process noise. The CVA methodology is then applied to the identification of models for a pilot-scale distillation column.

1. INTRODUCTION

In general, dynamic process models can be developed based on first principles (e.g. mass and energy balances) and/or experimental data. However, the development and evaluation of physically-based models can be very costly, time consuming and require a major commitment of often scarce engineering manpower. Furthermore, the resulting dynamic model of the process may not be in a form suitable for control system design. For example a physically-based model of a large industrial multicomponent distillation column typically consists of a set of 50–300 nonlinear ordinary differential equations. Although such a complex process model can be used in the final evaluation of an advanced control strategy, it does not facilitate the control system design nor can it be conveniently incorporated into a model-based, on-line control strategy. For these purposes it is desirable to have a less complex process model but one which retains the important characteristics of the process. In principle, a simpler, low-order model can be obtained from a complex model by applying model reduction techniques (Benallou *et al.*, 1986; Bonvin and Mellichamp, 1982). However, reliable model reduction methods for general classes of nonlinear process models are not available.

Process identification provides an alternative approach for developing dynamic process models through a sequence of steps:

(1) Selection of input and output variables for the process model (i.e. selection of the manipulated and controlled variables).
(2) Selection of the input excitation (e.g. step inputs, pseudo-random binary sequence).
(3) Collection of data.
(4) Selection of model structure.
(5) Calculation of model parameters.
(6) Model verification.

It should be emphasized that, in general, these steps are part of an iterative procedure. Also, the order in which these steps are performed may vary from one problem to another. For example, the model structure may be specified as part of the first step, prior to data collection rather than after.

Although the basic concept of process identification is quite straightforward, it is often the most time consuming step in the design and implementation of advanced control strategies as noted by Prett *et al.* (1987). They cite a number of important practical problems that must be addressed when

†To whom all correspondence should be addressed.

process identification is performed in an industrial plant:

(1) The output variables are affected by unmeasured disturbances.
(2) The input excitation may not be adequate to identify the process to a sufficient degree of accuracy.
(3) The approximate process models that are typically obtained via process identification are only valid for a limited range of process conditions. (For example, a linear transfer function model provides only an approximate representation for a nonlinear, time varying process.)

The appropriate choice of an input excitation for accurate identification of the process may require some prior knowledge of the plant dynamics.

A number of system identification methods have been developed in the last decade. In recent studies (Larimore and Mehra, 1984; Peloubet *et al.*, 1990), the literature of adaptive control and system identification was surveyed, and the various methods were evaluated for their potential application to the adaptive identification of wing flutter dynamics. Aircraft wing flutter is a resonant aeroelastic vibration which can result in exponentially growing oscillations and wing failure in less than 1 s. The major emphasis of the above studies was on the suitability of the various identification procedures for automatic implementation without the need for an analyst, and only linear system identification methods were addressed. All of the available methods were found to have one or more major difficulties which preclude their general application to the problem of identification in real-time using on-line computations. The least-squares, recursive least-squares (RLS) and autoregressive methods give biased estimates, while the statistically accurate methods of autoregressive, moving average (ARMA) modeling such as extended least-squares (ELS), the Box–Jenkins approach (Box and Jenkins, 1976) and self-tuning regulator (Ljung and Soderstrom, 1983) are not computationally or statistically well-conditioned. Gevers and Wertz (1982) have shown that the ARMA model structure is not globally identifiable. Thus, these methods can fail when applied to high-order mutlivariable systems. In addition, the problems of accurate initialization and slow convergence rule these out as general identification methods. Maximum likelihood estimation (MLE), which is the most precise approach, requires an expert user to deal with convergence and other problems that frequently arise.

A fundamental problem with many of the above system identification methods is that specification of model order limits the dynamics that are obtained in the fitted model. To include additional dynamics requires fitting another higher-order model with a substantial amount of additional computation and the possibility of identifiability or convergence problems. Further, this procedure may have to be repeated several times. Table 1 provides a summary of some of the major issues of on-line system identification discussed above for the various system identification procedures. Progress beyond the present state-of-the-art in system identification requires a new approach to the problem.

In order for process identification to be performed on a more routine basis, a unified approach is needed that:

(1) Is applicable to a broad class of process models (e.g. linear state-space or ARMA models of arbitrary order).
(2) Is capable of identifying both linear and nonlinear process models.
(3) Automatically determines the optimum model structure and model orders but avoids overfitting the available data (Larimore and Mehra, 1985).
(4) Requires a bare minimum of information regarding the mathematical and statistical details of the process identification method.
(5) Can provide a measure of the validity of the estimated model and an indication of whether the input excitation was adequate.
(6) Is robust with respect to mild departures in the modeling assumptions such as linearity, time invariance and Gaussian distribution and stationarity of the noise.

An approach to model identification that incorporates these important features—canonical variate analysis (CVA)—is described in this paper. In this approach the statistically significant states, i.e. all accumulation relations that are statistically significant, are first determined from the observations; thus a fundamental mathematical basis for including them is provided. Using this result, the detailed mathematical description of the system dynamics is easily obtained by regression methods. This approach

Table 1. Comparison of system identification procedures

	RLS	ELS	MLE	CVA
Unbiased parameter estimates	No	Yes	Yes	Yes
Statistical accuracy	No	Yes	Yes	Yes
Nonexpert user	Yes	Yes	No	Yes
Computationally reliable	Yes	No	No	Yes
Model order selection	Yes	No	Yes	Yes

RLS = recursive least squares.
ELS = extended least squares.
MLE = maximum likelihood estimation.
CVA = canonical variate analysis.

provides a statistical basis, as opposed to an *ad hoc* basis, for performing model identification.

The objectives of this paper are to describe the CVA process identification method, to show that it meets all the criteria in Table 1, and to illustrate the technique's suitability for a wide range of process identification problems.

2. CANONICAL VARIATE ANALYSIS

The analysis of canonical correlations and variates is a method of mathematical statistics developed by Hotelling (1936) [also see Anderson (1984)]. Concepts of canonical variables for representing random processes were explored by Gelfand and Yaglom (1959), Yaglom (1970) and Kailath (1974). The initial application of the canonical correlation analysis method to stochastic realization theory and system identification was done in the pioneering work of Akaike (1974, 1976). This initial work had a number of limitations such as no system inputs, no additive measurement noise, substantial computational burden involving numerous singular value decompositions (SVDs), a heuristic set of decisions for choosing a basis for representation of the system, and a number of approximations including computation of the Akaike Information Criterion (AIC) for decision on model order.

Some important generalizations and improvements to Akaike's canonical correlation method have been made by Larimore (1983b). [A tutorial development and discussion of available computer software is given by Larimore (1990)]. These include generalization to systems with additive measurement noise and with inputs including feedback controls. A major departure of this approach from previous work is the use of a single canonical variate analysis which gives the optimal choice of k linear combinations of the past inputs and outputs for prediction of the future outputs for any choice of k. The problem of finding an optimal Hankel norm reduced-order model (Kung and Lin, 1981) is related to the canonical variate approach (Camuto and Menga, 1982; Larimore, 1983b). The balanced realization method is a particular case of the canonical variate analysis (Desai and Pal, 1984). More recently the general approach central to CVA involving the use of the singular value decomposition on the past and future of a process has been called subspace systems identification, and recent research is reviewed in Heij and Roorda (1991), Van Overschee and De Moor (1991), Van der Klauw *et al.* (1991), and Viberg *et al.* (1991) and their cited references.

A related approach used in the chemical engineering and chemometrics fields is the partial least squares (PLS) approach (Hoskuldsson, 1988; MacGregor *et al.*, 1991; Wold *et al.*, 1984). The philosophy underlying PLS is very closely related to CVA; both are equivalent to a generalized singular value decomposition on the covariance matrix. The CVA method uses a weighting so as to maximize correlation, whereas PLS maximizes covariance. Maximizing correlation is statistically optimal for determining the rank of a covariance matrix using likelihood ratio tests (Anderson, 1984, pp. 497–498). The PLS procedure is sequential in selecting the important components, working with the residuals from the previous step. By contrast CVA simultaneously obtains all components in one generalized SVD and is consequently much more efficient computationally. While most of the literature on PLS concerns the static regression problem, some initial efforts have been made in the modeling of dynamical systems (MacGregor *et al.*, 1991).

To discuss the related research more concisely, the problems of identification, reduced-order modeling and filtering, and predictive control can be described concisely as follows (Larimore, 1983b, 1990). For a more detailed discussion, beyond the scope of this paper, consult the cited references. Consider the identification of a dynamic process model given data comprised of inputs and outputs of a multivariable process. For any sample time t, we denote a vector p_t consisting of past outputs y_t and inputs u_t and denote a vector f_t of outputs at time t or later:

$$p_t^{\mathrm{T}} = (y_{t-1}, y_{t-2}, \ldots, u_{t-1}, u_{t-2}, \ldots)^{\mathrm{T}}, \quad \text{(1a)}$$

$$f_t^{\mathrm{T}} = (y_t, y_{t+1}, \ldots)^{\mathrm{T}}. \quad \text{(1b)}$$

The covariance matrices among vectors f and p are denoted as Σ_{ff}, Σ_{pp} and Σ_{fp} which are, in practice, sample covariance matrices.

The identification problem is formulated as one of determining a number k of linear combinations of the past p_t which allow prediction of the future f_t. The set of k linear combinations of the past p_t are denoted as a $k \times 1$ vector m_t and are considered as k-order memory of the past. The optimal linear prediction $\hat{f}_t$ of the future f_t, which is a function of a reduced-order memory m_t is measured in terms of the average prediction error:

$$\mathbb{E}\{\|f_t - \hat{f}_t\|_{\Lambda}^{2\dagger}\} = \mathbb{E}\{(f_t - \hat{f}_t)^{\mathrm{T}} \Lambda^{\dagger} (f_t - \hat{f}_t)\} \quad (2)$$

where $\mathbb{E}$ is the expectation operator, $\dagger$ denotes the pseudoinverse and Λ is an arbitrary positive semidefinite symmetric matrix so that $\Lambda^{\dagger}$ is an arbitrary quadratic weighting that can be singular. The optimal prediction is obtained by determining an optimal k-order memory:

$$m_t = J_k p_t \quad (3)$$

choosing the k rows of J_k such that the optimal linear predictor $\hat{f}_t(m_t) = \Sigma_{fm}\Sigma_{mm}^{-1}m_t$ based on m_t minimizes the prediction error (2).

The solution to this problem in the completely general case where the matrices Σ_{ff}, Σ_{pp} and Λ may be singular is given by the generalized singular value decomposition as derived by Larimore (1983b): consider the problem of choosing k linear combinations $m_t = J_k p_t$ of p_t for predicting f_t, such that (2) is minimized where Σ_{pp} and Λ are possibly singular positive semidefinite symmetric matrices with ranks m and n, respectively. Then the existence and uniqueness of solutions are completely characterized by the (Σ_{pp}, Λ)-generalized singular value decomposition which guarantees the existence of matrices J, L and generalized singular values $\gamma_1, \ldots, \gamma_r$ such that:

$$J\Sigma_{pp}J^{\mathrm{T}} = I_m,$$
$$L\Lambda L^{\mathrm{T}} = I_n,$$
$$J\Sigma_{pf}L^{\mathrm{T}} = \mathrm{diag}(\gamma_1 \geqslant \ldots \geqslant \gamma_r, 0, \ldots, 0). \quad (4)$$

The solution is given by choosing rows of J_k as the first k rows of J if the kth singular value satisfies $\gamma_k \geqslant \gamma_{k+1}$. If there are r repeated singular values equal to γ_k then there is an arbitrary selection from among the corresponding singular vectors, i.e. rows of J. The minimum value is:

$$\min \mathbb{E}\{\|f_t - \hat{f}_t\|^2_{\Lambda^\dagger}\} = trL\Lambda_{ff}L^{\mathrm{T}} - \gamma_1^2 - \ldots - \gamma_r^2, \quad (5)$$

where the minimization is made subject to the condition $\mathrm{rank}(J_k\Sigma_{pp}J_k^{\mathrm{T}}) = k$.

This result not only gives a complete characterization of the solutions in selecting optimal predictors m_k for prediction of the future f_t from the past p_t, but the reduction in prediction error for all possible selections of order k is given simply in terms of the generalized singular values. Further, its importance is clear since it eliminates a considerable amount of computation to determine the selection of the order that is appropriate in a given situation.

For system identification, the use of the weighting matrix $\Lambda = \Sigma_{ff}$ results in a near maximum likelihood system identification procedure (Larimore *et al.*, 1984). In model predictive control, the weighting is chosen appropriate to the control problem (Maurath *et al.*, 1988). In the partial least squares approach, the first PLS component is obtained by setting $\Sigma_{pp} = I$ and $\Lambda = I$ in solving the generalized SVD [equation (4)]. For each successive PLS component, covariances are computed from the residuals and the procedure repeated.

The CVA method in conjunction with information-based multiple decision procedures leads to various discrete-time, state-space models and to the selection of the model state order that best fits the data. Consider the general case of the reduced-order filtering and modeling problem: given the past of the related random processes, u_t and y_t, we wish to model and predict the future of y_t by a k-order state x_t and state-space structure of the form:

$$x_{t+1} = \Phi x_t + \Gamma u_t + w_t, \quad (6)$$
$$y_t = Cx_t + Du_t + Ew_t + v_t, \quad (7)$$

where x_t is the state and w_t and v_t are white noise processes that are independent with covariance matrices, Q and R, respectively. Note that correlation between the state noise w_t and the measurement noise $Ew_t + v_t$ enters from the presence of the Ew_t term. This term is necessary because CVA determines a minimal order state that requires the Ew_t term which in particular cases can indeed yield $E = 0$. (A special case of the reduced-order filtering problem is the transfer function approximation problem where u_t and y_t are the input and output variables and an approximate state-space model is desired.)

Given the computational problem with finite data, the past and future of the process are taken to be finite lags of length d with:

$$p_t^{\mathrm{T}} = (y_{t-1}, \ldots, y_{t-d}, u_{t-1}, \ldots, u_{t-d})^{\mathrm{T}},$$
$$f_t^{\mathrm{T}} = (y_t, \ldots, y_{t-d})^{\mathrm{T}}. \quad (8)$$

Akaike (1976) proposed choosing the number d of lags by least squares autoregressive modeling using recursive least squares algorithms to minimize the Akaike information criterion (AIC) discussed below. This method insures that a sufficient number of lags is used to capture all of the statistically significant behavior in the data; it is easily generalized to include autogressive models with moving averages in the inputs. In the model identification problem, the generalized SVD described earlier determines a transformation J of the past that puts the state in a canonical form so that the memory $m_t = J_k p_t$ contains the states ordered in terms of their importance in modeling the process. The optimal memory for a given order k then corresponds to selection of the first k states.

To decide on the model state order or model structure, recent developments based upon information measures will be used. Such methods were originally developed by Akaike (1974) and involve the use of the AIC to determine the appropriate order of a statistical model. The AIC for each order k is defined by:

$$\mathrm{AIC}(k) = -2\log p(Y^N, U^N; \hat{\theta}_k) + 2fM_k, \quad (9)$$

where $p(\cdot)$ is the likelihood function, based on the observations (Y^N, U^N) at N sampling instants, with the maximum likelihood parameter estimates $\hat{\theta}_k$

obtained using a k-order model with M_k parameters. The small sample correction factor f has been recently shown to produce near optimal order selection even in very small samples (Hurvich *et al.*, 1990). The model order k is chosen to correspond with the minimum value of AIC(k). A predictive inference justification of the use of an information-based criterion such as AIC is given in Larimore (1983a) based upon the fundamental statistical principles of sufficiency and repeated sampling. The number of parameters M_k in the state-space model of (6) and (7) is determined by the general state-space canonical form as in Candy *et al.* (1979) and is far less than the number of elements in the various state-space matrices. For a system with m inputs, n outputs and k states, the number of parameters is:

$$M_k = k(2n + m) + nm + n(n + 1)/2. \quad (10)$$

The small sample correction factor is:

$$f = \frac{N}{N - \left(\frac{M_k}{n} + \frac{n+1}{2}\right)}. \quad (11)$$

The AIC provides a simple yardstick for comparing models of different state orders [see Larimore (1993) for details]. Suppose that the true state order is $\bar{k}$ and consider the large sample case. Then the quantity:

$$\chi^2(M_k - M_{\bar{k}}) = \mathrm{AIC}(\bar{k}) - \mathrm{AIC}(k) + 2(M_k - M_{\bar{k}}), \quad (12)$$

is a chi-squared random variable and the expected value of AIC(k) for k beyond the true order $\bar{k}$ is:

$$\mathrm{AIC}(k) = \mathrm{AIC}(\bar{k}) + (M_k - M_{\bar{k}}), \quad (13)$$

so that the AIC(k) as a function of k is expected to increase equal to the number of additional parameters $M_k - M_{\bar{k}}$. The variability in this last equation has a standard deviation of approximately $\sqrt{2(M_k - M_{\bar{k}})}$. This expected value and the standard deviation of AIC(k) provide a simple means of comparing models to determine if they are essentially equivalent or very different in terms of statistical significance; a detailed procedure is discussed in Larimore (1993).

In this paper, the issue of model order selection and order reduction is taken to involve only the statistical significance of the various hypothesized models. After determining the statistically significant system dynamics using CVA, one may wish to apply further model reduction methods based upon control objectives or other issues, taking as a starting point the model identified using CVA.

Once the optimal k-order memory m_t is determined, state-space equations of the form (6) and (7) for approximating the process evolution are easily computed by a multiple regression procedure (Larimore, 1983b). Since the CVA system identification procedure involves the state-space model form, it has the major advantage that the model is globally identifiable; thus the method is statistically well-conditioned in contrast to ARMA modeling methods (Gevers and Wertz, 1982). Furthermore, since the computations are primarily based on SVD, they are numerically stable and accurate with an upper bound on the required number. From the CVA method theory (Larimore *et al.*, 1984), it can be shown that difficulties, such as biased estimates, are not caused by the presence of a correlated feedback signal.

3. PROCESS IDENTIFICATION USING CVA

The CVA method of system identification was evaluated in this research by simulation of a continuous stirred tank reactor (CSTR) and a tubular autothermal reactor with internal countercurrent heat exchange (ARICHE). In addition, CVA was evaluated with experimental data from a pilot-scale distillation column.

3.1. Continuous stirred-tank reactor

3.1.1. Process description. The CSTR is a well-known example in the process control literature (Seborg *et al.*, 1989). The process typically has two inputs (feed flowrate and steam temperature) and two outputs (reactor concentration and temperature). By assuming: (i) perfect mixing; (ii) negligible ambient heat loss; (iii) first-order irreversible reaction; and (iv) identical physical properties of the inlet and outlet streams, the material and energy balances can be described by:

$$\frac{dC_A}{dt} = \frac{Q_f}{V}(C_{AF} - C_A) - k_0 C_A e^{-E/RT}, \quad (14)$$

$$\frac{dT}{dt} = \frac{Q_f}{V}(T_F - T) + \frac{k_0 C_A}{C_{P\rho}}(-\Delta H) e^{-E/RT} - \frac{UA}{VC_{P\rho}}(T - T_c), \quad (15)$$

where

k_0 = frequency factor (1/h),
E = activation energy (B.t.u./lb mol),
R = gas constant (B.t.u./lb mol °R),
T = temperature of reactor content and exit stream (°R)
C_A = concentration of reactor content and exit stream (lb mol/ft^3),
T_F = temperature of the inlet stream (°R)

Table 2. Specification of the CSTR system

Parameter	Value
k_0	2.0×10^8/h
E/R	1.0×10^4 °R
T_F	530 °R
T_c	530 °R
C_{AF}	0.27 lb mol/ft^3
$-\Delta H$	1.5×10^4 B.t.u./lb mol
UA	2000 B.t.u./h °R
Q_f	100 ft^3/h
V	50 ft^3
ρC_p	50 B.t.u./ft^3 °R

T_c = temperature of cooling, (or heating) fluid (°R),
C_{AF} = concentration of inlet stream (lb mol/ft^3),
ΔH = molar heat of reaction (B.t.u./lb mol),
U = overall heat transfer coefficient (B.t.u./ft^2 h°R);
A = overall heat transfer area (ft^2)
Q_f = volumetric flowrate (ft^3/h),
C_p = heat capacity of reactor content (B.t.u./lb °R),
V = volume of reactor (ft^3),
ρ = density of reactor content (lb/ft^3).

The nominal operating values and constants for this example are given in Table 2. Equations (14) and (15) can be linearized about an operating point and arranged into state-space form (see Seborg *et al.*, 1989).

3.1.2. Experiment design. 300 samples were generated from the CSTR simulation using a sampling period of 3 min. The settling time for both outputs is about 100 min. Pseudo-random binary sequences were applied to each input simultaneously with the amplitudes chosen as 10°R for the coolant temperature and 10 ft^3/h for the flowrate. White measurement noise was added to the output in concentration and in temperature. A 10:1 signal to noise ratio (standard deviation of the uncorrupted output to the standard deviation of the noise) was used.

3.1.3. Results and model validation. A multi-input, multi-output (MIMO) state-state model was identified. The AICs are presented as a function of model order in Table 3. From equation (10), for a 2×2 system there are 6 additional parameters for each additional state order and the standard deviation of AIC(k) in equation (12) is $\sqrt{12}$. Thus order 2 is a highly significant improvement over order 1. For orders higher than 2, the behavior of the AIC is consistent with the hypothesis that the true order is 2. The minimum AIC corresponds to a second-order model with poles of $0.85 \pm 0.073j$ compared to the poles of $0.84 \pm 0.082j$ for a linearized description of the CSTR. The Φ, Γ and C matrices of (6) and (7) are given by:

Table 3. Comparison of AIC with model order for the CSTR

Model order	AIC
0	−4964
1	−5075
2	−5141
3	−5137
4	−5133
5	−5127

$$\Phi = \begin{bmatrix} 0.80 & -0.14 \\ 0.049 & 0.82 \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} 0.038 & -0.032 \\ 0.047 & 0.034 \end{bmatrix},$$

$$C = \begin{bmatrix} 0.0023 & 0.0019 \\ -1.1 & 0.14 \end{bmatrix}$$

The outputs for this model are C_A and T. The inputs are Q_f and T_c.

Model validation analysis included simulation of the system outputs from a particular input sequence and comparison of the transfer function spectrum of the identified state-space model with one computed by a fast Fourier transform (FFT) of the input and output data. A rigorous test of the model is to use a different input sequence than the one used for the identification. This procedure is known as cross-validation. For this example, the step responses were generated from the true process model, the linearized process model and the identified process model. Step changes of 30°R and 30 ft^3/h were separately applied to the coolant temperature and flowrate for model validation purposes. The magnitudes of these step tests were chosen large enough to emphasize the system's nonlinearities.

The results are given in Figs 1–4. Note that in all cases CVA did a better job of predicting the steady-state gain of the true plant than did the linearized model. This result is obtained because the linearized model applies only to a particular operating point while CVA potentially can obtain a more representative linear approximation of the process dynamics over the range of operation corresponding to the actual data.

The frequency response of the CVA model was compared with a spectral estimate of the data (FFT model) to demonstrate the ability of CVA to model resonance peaks, high-frequency roll-offs, phase shifts and low-frequency (gain) estimation. Note, however, that neither the spectral estimate nor CVA is exact; thus the comparisons were only semi-quantitative and the results are not shown here (see Schaper, 1990). The frequency responses of the CVA model and the FFT model exhibit similar behavior at high frequencies. Inaccuracy of the low-frequency spectral estimates could have been improved through adjustment of the Hamming window. The accuracy of the CVA model at low frequencies can be seen in the step responses of Figs 1–4.

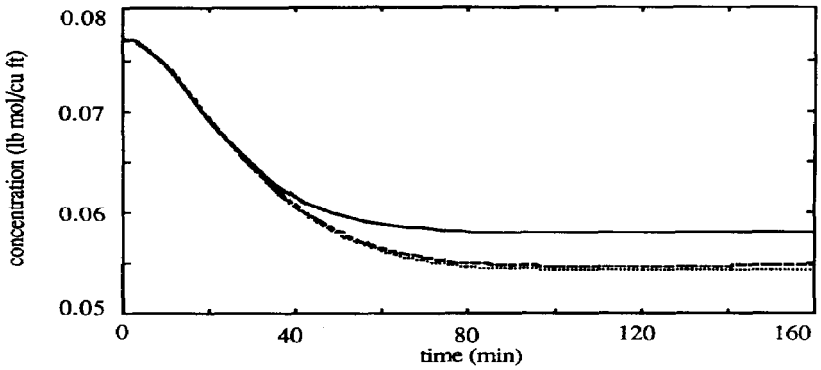

Fig. 1. The response of the reactor concentration for a step change of 30 °R in the coolant temperature: (——) nonlinear model; (– – –) CVA model; and (· · ·) linearized model.

In a second set of tests, a time delay was incorporated in the physical model by delaying the concentration measurement prior to model fitting. CVA was shown to accommodate the time delay easily by augmenting an additional state variable to the discrete state-space model. Validation results equivalent to those above were obtained (Schaper, 1990).

3.2. Autothermal reactor

3.2.1. Process description. The tubular autothermal reactor exhibits virtually every process nonideality encountered in practice (Bonvin *et al.*, 1983a,b). Included are: (1) both stable and unstable steady-state operation; (2) both inverse response (nonminimum phase) and oscillatory dynamic behavior; (3) very nonlinear behavior resulting from exponential dependence of reaction on temperature; (4) distributed process characteristics (i.e. an infinite-dimensional model); (5) highly coupled state variable structure due to internal energy flow paths; (6) inaccessible (unmeasurable) concentration state variables; (7) relatively few (i.e., 6–8) potential output variables (catalyst bed temperature measurements); (8) practical controllability/observability problems—in particular, a stong dependence of these properties on the operating state of the reactor if measurement points are chosen and remain fixed; (9) need for a two-time-scale identification technique to deal with practical stiffness in I/O relations; (10) a relatively high degree of inherent process noise related to catalyst bed channeling and mixing.

The reactor was studied by computer simulation using a process model that describes energy/temperature effects in the catalyst bed, feed section walls and insulation and reaction/concentration effects in the catalyst bed. A nonlinear computational version of the reactor model was obtained through the following procedure: (1) four partial differential equations describing energy/temperature effects in the catalyst bed, feed section walls, and insulation, and one partial differential equation describing reaction/concentration effects in the catalyst bed were developed; (2) a quasi-steady-state assumption reduced two PDEs to ODE form; (3) a 12-point orthogonal collocation procedure reduced the PDE/ODE mixed system to a set of 36

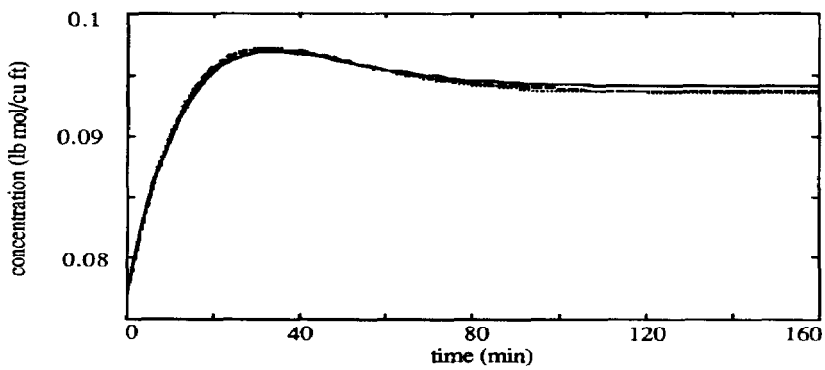

Fig. 2. The response of the reactor concentration for a step change of 30 ft^3/h in feed flowrate: (——) nonlinear model; (– – –) CVA model; and (· · ·) linearized model.

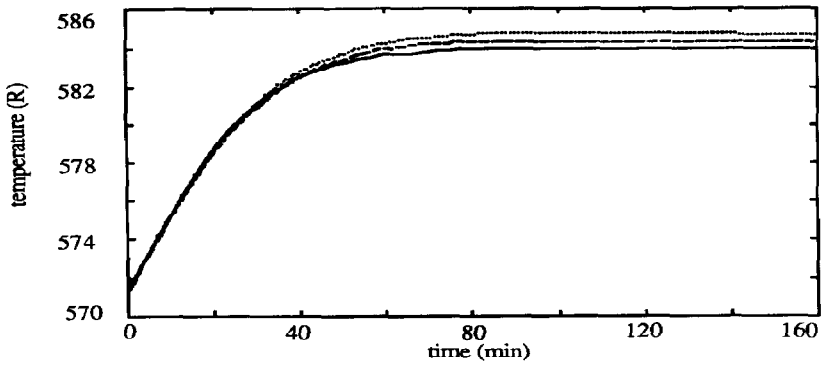

Fig. 3. The response of the reactor temperature for a step change of 30 °R in the coolant temperature: (——) nonlinear model; (– – –) CVA model; and (· · ·) linearized model.

nonlinear ODEs; (4) a single ODE relation was added to account for dynamics of the feed heating section; and (5) a random perturbation in one of the feed concentration variables was incorporated to introduce process noise. The resulting model contains 37 state variables. It has been tested extensively and shown to exhibit, both qualitatively and quantitatively, all of the features enumerated earlier for the actual process (McDermott, 1984a).

3.2.2. Experiment design. A (2 × 2) system was studied with the manipulated inputs chosen to be the heat fluxes to the top heater (q_{top}) and the bottom heater (q_{bot}). The measured outputs were chosen as the temperatures at the top of the catalyst bed (T_{115}) and the bottom of the catalyst bed (T_{991}), corresponding to axial locations 0.115 and 0.991. A set of binary sequences was applied to each input with amplitudes of 0.5 W for the top heater and 10 W for the bottom heater. Three-hundred and eighty samples were taken at a sampling period of 1 min. (The settling time is about 450 min for both outputs; thus, although many samples are used for identification, this example tests the ability of CVA to identify a model with small sample sizes relative to the settling time.) Another set of pseudo-random binary sequences was applied to the process model for later use in cross-validating the identified model.

The power spectral density of the identification input sequence of the top heater is shown in Fig. 5. This plot is to be compared with Fig. 6 in which the power spectral density is displayed for the validation input sequence of the top heater. The identification input sequence is spread over the entire frequency range while the validation input sequence is actually periodic, resulting in the periodic line spectra of Fig. 6. Thus, the ability of CVA to identify a model with good frequency response behavior in regions which have not been heavily excited can be evaluated. Of course, a good model can only be identified in the case where the true process spectrum has no sharp peaks or rapid changes corresponding to regions of no frequency excitation. The power spectral densities of the identification and validation input sequences for the bottom heater show similar characteristics.

3.2.3. Results and model validation. A multi-input, multi-output state-space model was identified given the output generated by the input corresponding to

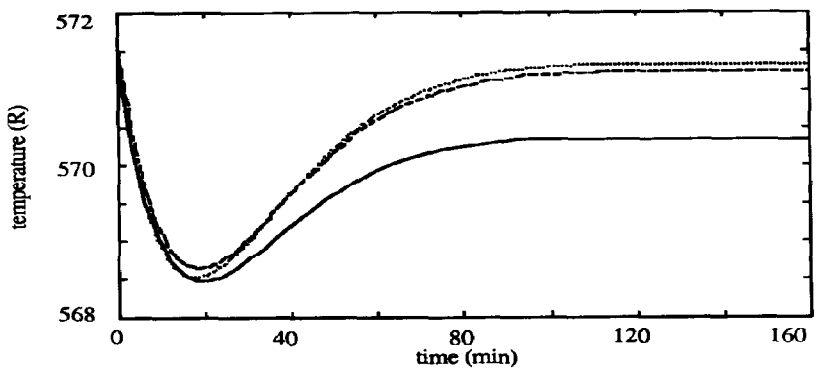

Fig. 4. The response of the reactor temperature for a step change of 30 ft^3/h in the feed flowrate: (——) nonlinear model; (– – –) CVA model; and (· · ·) linearized model.

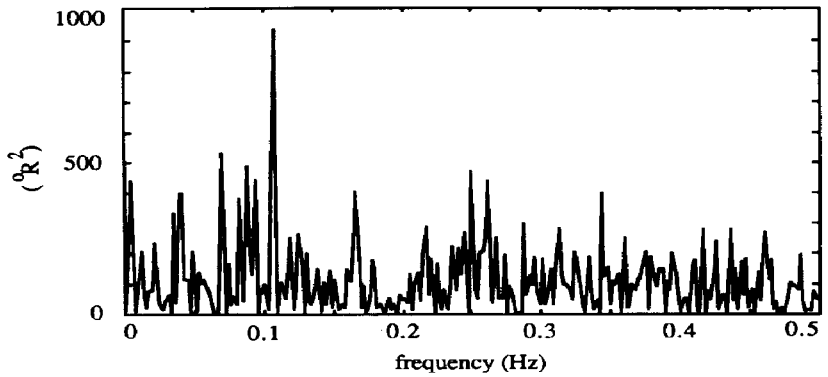

Fig. 5. The power spectral density of identification input sequence of the top heater.

the power spectral density of Fig. 5. The AICs are given in Table 4 in which it is seen that a sixth-order model is optimal. The AICs beyond order 6 are consistent with the hypothesis that the true system is order 6. For comparing orders 5 and 6, the chi-squared statistic of equation (13) is 19 with six degrees of freedom which is significant at the 0.01 level; thus, it is very unlikely that the true order is 5. The Φ, Γ and C matrices of (6) and (7) are given by: The ordering of the outputs for this model is T_{115} and T_{991}. The inputs are q_{top} and q_{bot}. The poles for this model are -0.32, $-0.10 \pm 0.53j$, 0.027, 0.73 and 0.99. Note that the poles are spread over a wide region in the z-plane. This slow mode/fast mode

$$\Phi = \begin{bmatrix} 0.87 & 0.19 & 0.051 & 0.061 & 0.044 & 0.15 \\ 0.13 & 0.76 & 0.19 & 0.028 & -0.22 & -0.12 \\ -0.027 & -0.15 & -0.097 & 0.55 & -0.58 & 0.41 \\ 0.26 & -0.30 & -0.071 & -0.69 & -0.56 & -0.074 \\ 0.33 & -0.40 & 0.014 & 0.45 & -0.045 & -0.59 \\ 0.25 & -0.22 & -0.092 & -0.069 & 0.50 & 0.44 \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} -0.0016 & -0.72 \\ -0.0066 & 1.1 \\ -0.032 & -0.23 \\ -0.0059 & 0.43 \\ 0.012 & 0.77 \\ -0.026 & 1.0 \end{bmatrix},$$

$$C = \begin{bmatrix} -3.9 & 0.92 & 0.018 & -0.0086 & -0.0043 & -0.0046 \\ -1.3 & -0.88 & 0.061 & 0.0086 & 0.0046 & 0.0004 \end{bmatrix}.$$

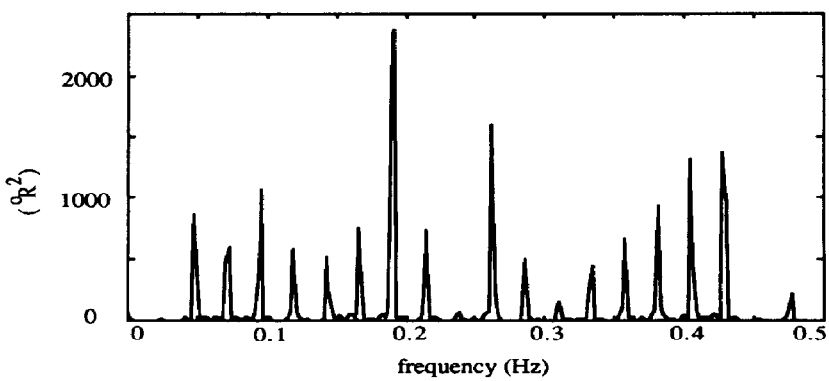

Fig. 6. The power spectral density of the validation input sequence of the top heater.

Table 4. Comparison of AIC with model order for the ARICHE

Model order	AIC
0	−1932
1	−2543
2	−2629
3	−2676
4	−2701
5	−2719
6	−2726
7	−2723
8	−2719
9	−2709
10	−2704

feature is consistent with earlier studies (McDermott *et al.*, 1986a,b) and demonstrates a stiffness characteristic of the differential equations of the process which make it difficult to identify. Note, also that a time delay of roughly three sampling periods is present from the top heater to T_{991} evident from the large magnitudes of the zeros of the input polynomials. The process is also nonminimum phase as evidenced by the zero which is outside the unit circle for the T_{991}–top heater polynomial (Schaper, 1990).

Cross-validation was performed using the input sequence with power spectral density (Fig. 6) different from the sequence used for identification (Fig. 5). Both the CVA model and the actual process model are subjected to this input sequence and the same initial condition. As in the previous case study, the outputs of the CVA-identified model closely matched the outputs of the nonlinear computer simulation model as shown in Figs 7 and 8. The gain estimation for the plant could be improved (as is evident, for instance, from the peak-to-peak tracking capability) by placing more emphasis on the low frequencies when performing the identification experiment.

The frequency response also was compared with spectral estimates obtained from the data. As in the previous case study, the spectral estimates were smoothed using a Hamming window. Both the parametric and nonparametric identification procedures produced similar results which are not shown here (see Schaper, 1990). The close agreement is important because it demonstrates that CVA can identify an accurate model in the presence of the many nonlinear features previously described. In particular, the phase lag of the transfer function from q_{top} to T_{991} is about 900°, giving an estimate of 2.5 sampling periods for the time delay which is believed to be roughly 3. This result assumes that the rest of the system is all-pass in nature.

A model was then identified using the periodic input sequence (Fig. 6). The optimal model order was 44. This high-order model is in reasonable agreement with the state order (37) used for simulation, confirming the utility of the AIC for choosing the model order. The method is also dependent on the type of input sequence used for experiment design as is usual for most identification strategies (Ljung, 1987). In this case, the periodic input sequence provides much better excitation of the process than did the PRBS sequence, leading to a more accurate description of the plant. It also suggests the need for model-order reduction techniques to be combined with the CVA method of identification if the model is to be used for controller design.

A comparison can be made between the low-order CVA-based model and other models obtained from alternative identification procedures. The comparison is qualitative in nature and considers results obtained earlier in applying other identification algorithms. In a study by Young (1988), the recursive least squares (RLS) algorithm produced a model describing T_{115} whose estimated steady-state gains were highly dependent upon the model order selected. It was demonstrated, for example, that an increase in model order from 2 to 3 produced poorer low frequency estimation results (as exemplified by comparing the estimated gain with a known gain) but better high-frequency estimation. Consequently, the selection of

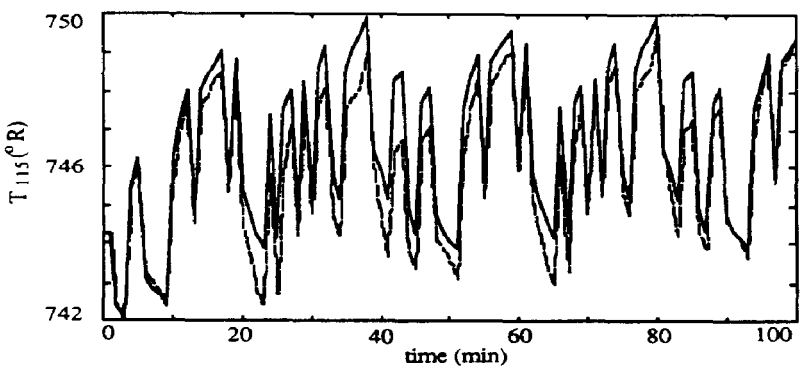

Fig. 7. Cross-validation for T_{115}: (——) reactor model; and (– – –) CVA model.

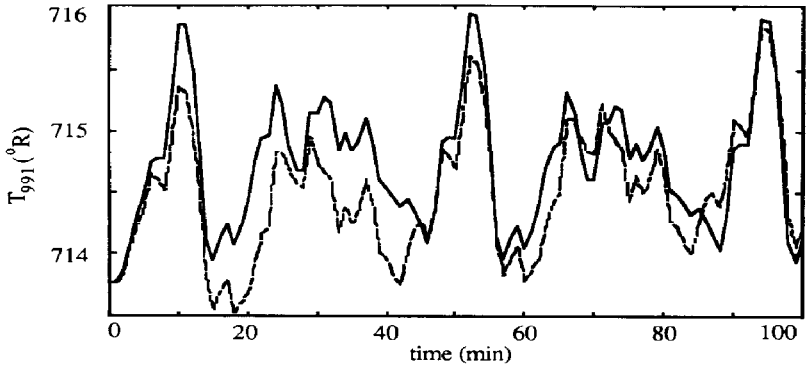

Fig. 8. Cross-validation for T_{991}: (——) reactor model; and (– – –) CVA model.

model order for predictive control strategies was somewhat arbitrary. However, in the present CVA study, the selection of model order is made clear by use of the AIC. The model order selected is high enough to capture the dynamics yet is low enough to avoid overfitting. The tradeoff between low- and high-frequency estimation is handled automatically through CVA.

Young (1988) improved the RLS estimation of T_{115} by augmenting a moving average polynomial to an ARX model structure. However, by so doing, the parameter estimation algorithm was no longer analytical; an extended least squares algorithm had to be employed which incorporated an iterative search procedure. The convergence of this search to a globally optimal point is not guaranteed. The CVA strategy avoids this iterative search while estimating the parameters for an equivalent model structure. This estimation is performed through a series of numerically accurate singular value decompositions. The noniterative scheme employed by CVA is a significant feature and is a definite advantage over the extended least squares algorithm. The drawback of this feature is that it prevents CVA from being configured in a classical recursive algorithm for tracking time varying processes. (Note that this is also true with least squares identification of ARMAX models.) On the other hand, it is argued (Larimore, 1990) that the batch CVA method gives rise to an optimal procedure for tracking time varying processes and avoids the need for heuristic choice of exponential forgetting factors.

Earlier investigations by McDermott (1984) have indicated that the RLS algorithm is not able to provide an accurate model for T_{991}. He hypothesized that the reason for this problem was the severe stiffness existing at that axial location. A fast mode is present due to material transport and a slow mode is present due to thermal transfer. Because of the inability to model T_{991} adequately, a measurement had to be taken at a different axial location and alternate model-based control strategy had to be developed. According to McDermott (1984), better control could have been achieved if a model was identified which could accurately describe the dynamics of T_{991}. By using T_{991} as the controlled variable, one would be able to modify the temperature profile along the entire reactor and maintain a sufficient margin of safety. Although RLS was not able to model T_{991} accurately, CVA did produce a suitable model, capturing the severe stiffness evident from cross-validation studies and spectral analysis.

3.3. Experimental distillation column

3.3.1. Process description. A 12-stage plus reboiler, 6 in. dia. pilot-scale distillation column at UCSB is used to separate a ternary mixture of butanols (Marchetti *et al.*, 1985). A dedicated gas chromatograph, measuring top and bottom composition periodically, and a microcomputer-based control system performing low-level control functions are interfaced to the high-level control computer. The asymmetric (nonlinear) and coupled dynamics of the column have furnished an excellent test of 2×2 system identification and adaptive and predictive control techniques in previous studies.

3.3.2. Experiment design. The manipulated inputs are the distillate flowrate and the reboiler steam pressure. The measured outputs are the distillate and bottom compositions. Pseudo-random binary sequences were applied to the two inputs simultaneously. A portion of the measurements of the actual input to the distillation column is displayed in Figs 9 and 10 at a much higher resolution than the 5 min sampling period used for system identification. For identification, the input is regarded as a sequence of step changes at the sampling times. However, Figs 9 and 10 demonstrate that the actual inputs deviate

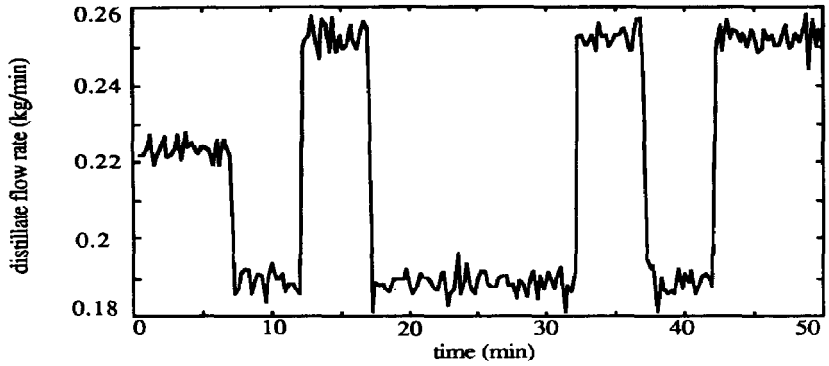

Fig. 9. Measurement of a portion of the input sequence for the distillate flowrate.

from the idealized step inputs used for identification, thereby introducing additional high-frequency colored process noise into the system.

3.3.3. Results and model validation. A MIMO model of the process was identified. The AICs are presented for different model orders in Table 5. A sixth-order model was selected. Orders 5, 6 and 7 provide essentially similar descriptions of the process from the viewpoint of statistical significance with order 6 having a slight edge. Beyond order 7, the AIC behavior is consistent with the hypothesis of no additional dynamics. Comparing orders 4 and 5 gives a χ^2-squared statistic of 17 on 6 *df* which is significant at the 0.01 level.

Note that the higher order state space model is required to account for time delays and also the interactions inherent to the system. The Φ, Γ and C matrices of (6) and (7) are given by:
The outputs for this model are distillate and bottoms composition of *n*-butanol. The inputs are distillate flowrate and reboiler steam pressure. The eigenvalues for the model are $0.0089 \pm 0.63j$, 0.40, $0.76 \pm 0.33j$ and 0.83.

The capability of the CVA identified model to simulate the process was tested by comparing the measured process responses with simulated responses of the identified model using the same input sequences. The comparisons between the actual and model responses are shown in Figs 11 and 12.

4. CONCLUSIONS

A new identification strategy has been applied to estimate models of chemical processes. This approach is based on canonical variate analysis to select a state coordinate system that relates inputs to future outputs in an optimal manner. The Akaike Information Criterion is used to determine an appropriate model order. Regression techniques are then used to determine a multi-input, multi-output process model. The CVA identification methodology showed satisfactory results for the estimation of process models for three case studies—a simulated CSTR, a simulated tubular autothermal counterflow reactor, and an experimental distillation column. The

$$\Phi = \begin{bmatrix} 0.89 & -0.42 & -0.059 & 0.040 & 0.047 & 0.046 \\ 0.31 & 0.72 & -0.58 & -0.14 & 0.0078 & 0.011 \\ 0.17 & 0.16 & 0.44 & -0.64 & 0.064 & 0.095 \\ 0.20 & 0.31 & 0.42 & -0.20 & 0.038 & -0.36 \\ 0.068 & 0.18 & 0.12 & 0.42 & 0.74 & -0.33 \\ -0.084 & 0.11 & 0.32 & 0.057 & 0.16 & 0.19 \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} -0.039 & -0.036 \\ 0.010 & -0.029 \\ 0.48 & -0.23 \\ -0.36 & 0.63 \\ 0.079 & 0.16 \\ -0.67 & -0.56 \end{bmatrix},$$

$$C = \begin{bmatrix} 0.030 & -0.0016 & 0.0028 & 0.0021 & 0.0010 & 0.00029 \\ -0.0047 & 0.0066 & 0.00088 & 0.00030 & 0.00017 & -0.00001 \end{bmatrix}.$$

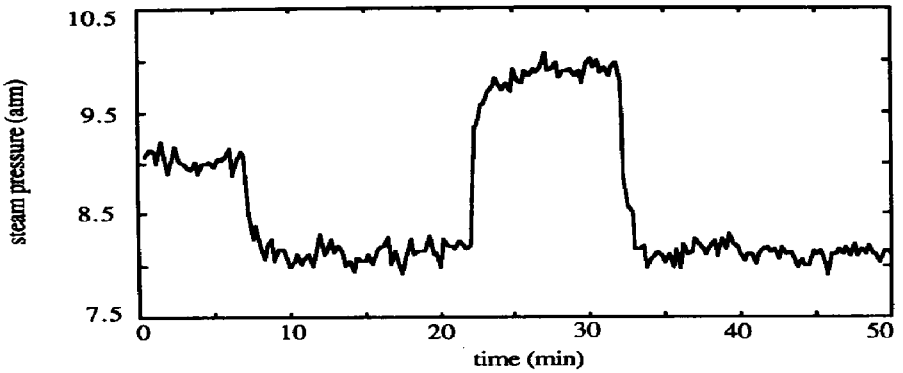

Fig. 10. Measurement of a portion of the input sequence for the steam pressure.

Table 5. Comparison of AIC with model order for the distillation column

Model order	AIC
0	−2034
1	−2076
2	−2101
3	−2120
4	−2130
5	−2135
6	−2136
7	−2135
8	−2128
9	−2121
10	−2112

identified models were evaluated under a variety of conditions that emphasized the presence of process nonlinearities and interactions, small sample sizes relative to the process settling time, stiff systems and evaluation at operating conditions different from those used for identification.

The ADAPT$_x$ computer software for automated system identification using the CVA method is available running in C^{++}, or running the MATRIX$_X$ or X-math computer aided control design software (Larimore, 1992).

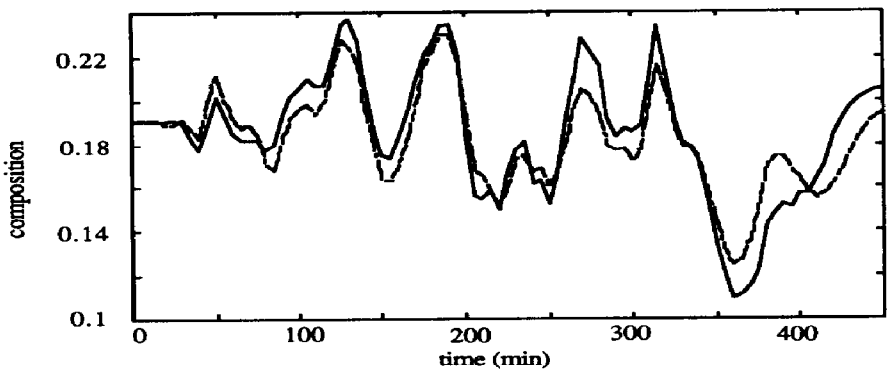

Fig. 11. Model prediction and measurement of the distillate compositon of n-butanol: (——) measurement; and (– – –) CVA model prediction.

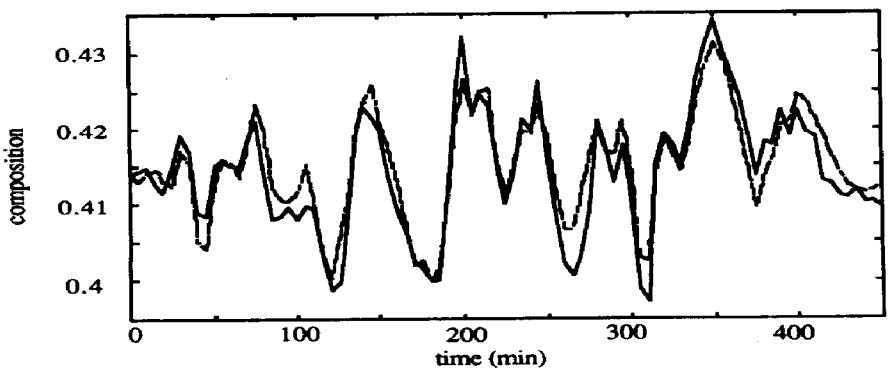

Fig. 12. Model prediction and measurement of the bottoms composition of n-butanol: (——) measurement; and (– – –) CVA model prediction.

Acknowledgement—The authors wish to acknowledge support of this research work by the National Science Foundation through a Small Business Innovation Research Contract, Number ISI-8761192, to Computational Engineering, Inc.

REFERENCES

Akaike H., A new look at statistical model identification. *IEEE Trans. Automat. Control* **AC-19,** 667 (1974).

Akaike H., Canonical correlation analysis of time series and the use of an information criterion. *System Identification: Advances and Case Studies* (R. K. Mehra and D. G. Lainoitis, Eds), p. 27. Academic Press, New York (1976).

Anderson T. W., *An Introduction to Multivariate Statistical Analysis*, 2nd Edn. Wiley, New York (1984).

Benallou A., D. E. Seborg and D. A. Mellichamp, Dynamic compartmental models for stagewise separation processes. *AIChE Jl* **32,** 1067 (1986).

Bonvin D. and D. A. Mellichamp, A unified derivation and critical review of modal approaches to model reduction. *Int. J. Contr.* **35,** 829 (1982).

Bonvin D., R. G. Rinker and D. A. Mellichamp, Dynamic analysis and control of a tubular autothermal reactor at an unstable steady state. *Chem. Engng Sci.* **35,** 603 (1980).

Bonvin D., R. G. Rinker and D. A. Mellichamp, On controlling an autothermal reactor at an unstable steady state: I—steady state and dynamic modelling. *Chem. Engng Sci.* **38,** 233 (1983a).

Bonvin D., R. G. Rinker and D. A. Mellichamp, On controlling an autothermal reactor at an unstable steady state: II—model reduction and control strategies which avoid state estimation. *Chem. Engng Sci.* **38,** 607 (1983b).

Box G. E. P. and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco (1976).

Camuto E. and G. Menga, Approximate realization of discrete time stochastic processes. *Sixth IFAC Symp. Ident. System Param. Estimation*. Washington, DC (1982).

Candy J. V., T. E. Bullock and M. E. Warren, Invariant description of the stochastic realization. *Automatica* **15,** 493 (1979).

Desai U. B. and D. Pal, A transformation approach to stochastic model realization. *IEEE Trans. Automat. Control* **AC-29,** 1097 (1984).

Gelfand I. M. and A. M. Yaglom, Calculation of the amount of information about a random function contained in another such function: *Am. Math. Soc. Trans. Ser.* **12,** 199 (1959).

Gevers M. and V. Wertz, On the problem of structure selection for the identification of stationary stochastic processes. *Sixth IFAC Symp. Ident. System Param. Estimation*, Washington, DC (1982).

Heij C. and B. Roorda, A modified canonical correlation approach to approximate state space modeling. *Proc. 30th Conf. Decision and Control*, Vol. 2, p. 1343, Brighton, U.K. (1991).

Hoskuldsson A., PLS regression methods. *J. Chemomet.* **2,** 211 (1988).

Hotelling H., Relations between two sets of variates. *Biometrika* **28,** 321 (1936).

Hurvich C. M., R. Shumway and C. L. Tsai, Improved estimators of Kullback–Leibler information for autoregressive model selection in small samples. *Biometrika* **77,** 709 (1990).

Kailath T. A view of three decades of linear filter theory. *IEEE Trans. Inform. Theory* **IT-20,** 146 (1974).

Kung S. Y. and D. W. Lin, Optimal Hankel-norm model reductions: multivariable systems. *IEEE Trans. Automat. Control* **26,** 832 (1981).

Larimore W. E., Predictive inference, sufficiency, entropy, and an asymptotic likelihood principle. *Biometrika* **70,** 175 (1983a).

Larimore W. E., System identification, reduced order filtering and modeling via canonical variate analysis. *Proc. Am. Control Conf.*, p. 445 (1983b).

Larimore W. E., Simultaneous confidence bands for efficient parametric multivariate spectral estimation. Unpublished manuscript (1989).

Larimore W. E., Canonical variate analysis for system idenification, filtering, and adaptive control. *Proc. 29th IEEE Conf. on Decision and Control*, Vol. 1, p. 635 Honolulu, Hawaii, (1990).

Larimore W. E., *ADAPT$_X$ Automated System Identification Software Users Manual*. Adaptics, Inc., 40 Fairchild Drive, Reading, MA (1992).

Larimore W. E. Accuracy confidence bands including the bias of model under fitting. *Proc. 1993 Am. Control Conf.*, San Francisco, CA (1993).

Larimore W. E. and R. K. Mehra, Technical assessment of adaptive flutter suppression research. Air Force Wright Aeronautical Lab Report No. AFWAL-TR-84-3052 (1984).

Larimore W. E. and R. K. Mehra, The problem of overfitting data. *Byte* **10,** 167 (1985).

Larimore W. E., S. Mahmood and R. K. Mehra, Multivariable adaptive model algorithmic control. *Proc. 23rd IEEE Conf. on Decision and Control*, Las Vegas, Nevada, p. 675 (1984).

Ljung L., *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ (1987).

Ljung L. and T. Soderstrom, *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, MA (1983).

MacGregor J. F., T. E. Marlin, J. Kresta and B. Skagerberg, Multivariate statistical methods in process analysis and control. *CPC-IV Conf.*, CACHE. South Padre Island, TX (1991).

Marchetti J. L., A. Benallou, D. E. Seborg and D. A. Mellichamp, A pilot scale distillation facility for digital computer control research. *Computers chem. Engng* **9,** 301 (1985).

Maurath P. R., Predictive controller design with applications to two-point composition control of distillation columns. Ph.D. Dissertation, University of California, Santa Barbara (1985).

Maurath P. R., A. J. Laub, D. E. Seborg and D. A. Mellichamp, Predictive controller design by principle component analysis. *IEC Res.* **27,** 1204 (1988).

McDermott P. E. Self-tuning control of a fixed-bed autothermal reactor. Ph.D. Dissertation, University of California, Santa Barbara (1984).

McDermott P. E., D. A. Mellichamp and R. G. Rinker, Pole-placement self-tuning control of a fixed-bed autothermal reactor—I. Single variable control. *AIChE Jl* **32,** 1004 (1986a).

McDermott P. E., D. A. Mellichamp and R. G. Rinker, Pole-placement self-tuning control of a fixed-bed autothermal reactor—II. Multivariable control. *AIChE Jl* **32,** 1015 (1986b).

Peloubet R. P., R. L. Haller and R. M. Bolding, On-line adaptive control of unstable aircraft wing flutter. *Proc. 29th IEEE Conf. on Decision and Control*, Honolulu, Hawaii (1990).

Prett D. M., T. A. Skrovanek and J. F. Pollard, Process identification—past, present, future. In *Shell Process Control Workshop* (D. M. Prett and M. Morari, Eds). Butterworths, Stoneham, MA (1987).

Schaper C. D., Ph.D. Dissertation, University of California, Santa Barbara (1990).

Seborg D. E., T. F. Edgar and D. A. Mellichamp, *Process Dynamics and Control*. Wiley, NY (1989).

Van der Klauw A. C., M. Verheagen and P. J. van der Bosch, Subspace identification of closed loop systems.

Proc. 30th Conf. Decision and Control, Vol. 2, p. 1321, Brighton, U.K. (1991).

Van Overschee P. V. and B. De Moor, Subspace algorithms for the stochastic identification problem. *Proc. 30th Conf. Decision and Control*, Vol. 2, p. 1327, Brighton, U.K. (1991).

Viberg M., B. Ottersten, B. Wahlberg and L. Ljung, A statistical perspective on state-space modeling using subspace methods. *Proc. 30th Conf. Decision and Control*, Vol. 2, p. 1337, Brighton, U.K. (1991).

Wold S., A. Ruhe, H. Wold and W. Dunn III, The colinearity problem in linear regression. The partial least squares (pls) approach to generalized inverses. *SIAM J. Sci. Statist. Comput.* **5**, 735 (1984).

Yaglom A. M., Outline of some topics in linear extrapolation of stationary random processes. *Proc. Fifth Berkeley Symp. Math. Stat. and Prob.*, p. 259. Berkeley, California Press (1970).

Young R. E., Adaptive supervisory control of a fixed-bed autothermal reactor. Ph.D. Dissertation, University of California, Santa Barbara (1988).