Proceedings of the 29th Conference
on Decision and Control
Honolulu, Hawaii • December 1990

IDENTIFICATION OF CHEMICAL PROCESSES USING CANONICAL VARIATE ANALYSIS

Charles D. Schaper*, Wallace E. Larimore**, Dale E. Seborg, Duncan A. Mellichamp

Department of Chemical and Nuclear Engineering
University of California
Santa Barbara, CA 93106

1 INTRODUCTION

In order for process identification to be performed on a 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 from the user 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 from 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., the possible energy storing relationships that are statistically significant, are first determined from the observations, and thus a mathematical basis for describing them is provided. Using this result, the detailed mathematical description of the system dynamics is easily determined by regression methods. This method thus provides a statistical basis, as opposed to an *ad hoc* basis, for performing model identification.

The objective of this paper is to describe the CVA process identification method and to provide demonstrations that the technique is suitable for a wide range of process control 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, 1958). 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 (SVD's), 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 in Akaike's canonical correlation method have been made by Larimore (1983b). These include generalization to systems with additive measurement noise and feedback control inputs. A major departure of the approach from previous work is the use of a single canonical variate analysis which yields the optimal choice of k linear combinations of the past input-output data 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). To discuss the related research more concisely, the problems of identification, reduced-order modeling and filtering, and predictive control can be described as follows (Larimore, 1983b).

The CVA approach identifies a dynamic process model from 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,

*Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305
**Adaptics Inc., 186 South Street, Reading, MA 01867

$$p_t^T = (y_{t-1}, y_{t-2}, \cdots, u_{t-1}, u_{t-2}, \ldots)^T \qquad (1a)$$

$$f_t^T = (y_t, y_{t+1}, \cdots)^T \qquad (1b)$$

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

The identification problem is formulated as 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×1 vector m_t and is 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

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

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 is possibly singular. The optimal prediction is to determine an optimal k-order memory

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

by 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

$$J \Sigma_{pp} J^T = I_m \,, \; L \Lambda L^T = I_n \,,$$

$$J \Sigma_{pf} L^T = \mathrm{Diag}(\gamma_1 \geq \cdots \geq \gamma_r, 0, \cdots, 0) \qquad (4)$$

The solution is given by choosing the rows of J_k as the first k rows of J if the k-th singular value satisfies $\gamma_k \geq \gamma_{k+1}$. 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

$$\min_{\mathrm{rank}(J_k \Sigma_{pp} J_k^T) = k} \mathbb{E}\{ \| f_t - \hat{f}_t \|^2_{\Lambda^\dagger} \}$$

$$= \mathrm{tr}\, L \Sigma_{ff} L^T - \gamma_1^2 - \cdots - \gamma_r^2 \qquad (5)$$

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

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).

The CVA method in conjunction with information based multiple decision procedures allows the determination of the fit of the various state-space models and the selection of the best model state order. 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

$$x_{t+1} = \Phi x_t + G u_t + w_t \qquad (6)$$

$$y_t = H x_t + A u_t + B w_t + v_t \qquad (7)$$

where x_t is the state and w_t and v_t are independent white noise processes with covariance matrices, Q and R, respectively. 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 finite data in the computational problem, the past and future of the process are taken to be finite lags of length d with

$$p_t^T = (y_{t-1}, \cdots, y_{t-d}, u_{t-1}, \cdots, u_{t-d})^T \,,$$

$$f_t^T = (y_t, \cdots, y_{d+1})^T \qquad (8)$$

Akaike (1976) proposed choosing the number d of lags by least squares autoregressive modeling using recursive least squares algorithms to minimize the AIC criterion discussed below. This procedure insures that a sufficient number of lags is used to capture all of the statistically significant behavior in the data and is easily generalized to include the case with inputs u. 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 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

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

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 $\hat{\theta}_k$ using a k-order model with M_k parameters. The model order k is chosen corresponding to 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.

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, hence 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 computations. From the CVA method theory (Larimore *et al.*, 1984), it can be shown that no difficulties, such as biased estimates, will be 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 both a continuous stirred-tank reactor (CSTR) and an autothermal reactor and through tests with an experimental distillation column.

3.1 Continuous stirred-tank reactor

Process description: The continuous stirred-tank reactor (CSTR) is a well- known example in the process control literature (Seborg *et al.*, 1989). The process typically utilizes two manipulated inputs (feed flow rate and coolant 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) constant physical properties of the inlet and outlet streams, the material and energy balances can be described by

$$\frac{dC_A}{dt} = \frac{Q_f}{V}(C_{AF} - C_A) - k_0 C_A e^{-E/RT} \quad (10)$$

$$\frac{dT}{dt} = \frac{Q_f}{V}(T_F - T) - \frac{k_0 C_A}{C_P}(-\Delta H)e^{-E/RT} - \frac{UA}{VC_P\rho}(T - T_c \quad (11)$$

where

T = temperature of reactor content and exit stream (°R)

C_A = concentration of reactor content and exit stream (lb mol/ft^3)

T_c = temperature of coolant (°R)

Q_f = volumetric flow rate (ft^3/hr)

The nominal operating values and constants for this example are given by Schaper (1990). Equations (1) and (2) can be linearized about an operating point and arranged into state-space form (see Seborg *et al.*, 1989).

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

Results and model validation: A multi-input, multi-output (MIMO) state-space model was identified with outputs C_A and T and inputs F and T_c. The minimum AIC value corresponded 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, G, and H matrices of (6) and (7) are given by Schaper (1990).

Model validation analysis included simulation of the system outputs for an input sequence and comparison of the transfer function spectrum of the identified state-space model with that computed from fast Fourier transforms (FFT) of the input and output data.

A rigorous test of the model is to use an input sequence which is different than the one used for the identification. This procedure is known as cross-validation. For this example, 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/hr were separately applied to the coolant temperature and flow rate. The magnitudes of these step tests were chosen large enough to emphasize the system's nonlinearities.

The results are given in Figures 1 through 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 indicates that the linearized model is constrained to a particular operating point while CVA can potentially obtain a more accurate linear approximation of the process dynamics over the range of operating points corresponding to the actual response. Schaper (1990) shows through spectral analysis that the CVA model and the linearized model are in good agreement.

A time delay was also incorporated in the model by delaying the concentration measurement prior to model fitting. CVA was shown to accommodate the time delay easily by adding a state variable to the state-space model. Validation results equivalent to those above were obtained.

3.2 Autothermal countercurrent reactor

Process description: The autothermal reactor utilizes energy freed in an exothermic chemical reaction to autoheat its own feed stream to reaction conditions in a counterflow arrangement. The process, which exhibits virtually every nonideality encountered in practice (Bonvin et al, 1983a,b), can be modeled by one ordinary and four partial differential equations to account for convection of material and energy. These equations can be approximated by a 37th-order o.d.e. model which exhibits, both qualitatively and quantitatively, all features of the actual process.

Experiment design and results: A 2x2 process model was developed utilizing heat fluxes in the top and bottom bed heaters as inputs, and temperatures at the top and bottom of the catalyst bed as outputs. Schaper (1990) reports these studies in detail and compares the CVA results with alternative identification results obtained earlier by McDermott (1984) and Young (1988).

3.3 Experimental distillation column

Process description: A twelve-stage plus reboiler, six-inch diameter pilot scale distillation column at UCSB is used to separate a ternary mixture of butanols (Marchetti *et al.*, 1985). The asymmetric (nonlinear) and coupled dynamics of the column have furnished excellent tests of 2x2 system identification and adaptive and predictive control techniques in previous studies.

Experiment design: The manipulated inputs are the distillate flow rate and the reboiler steam pressure. The measured outputs are the distillate and bottom compositions of n-butanol. Pseudo random binary sequences were applied to the two inputs simultaneously.

Results and model validation: A MIMO model of the process was identified. Inspection of the AIC's indicated that a sixth-order model was best [Schaper (1990)]. Note that the high order state-space model is required to account for time delays and also the interactions inherent in the system. The Φ, G, and H matrices of (6) and (7) are given by Schaper (1990). The eigenvalues for the model are $0.0089 \pm 0.63j$, $0.40 \pm 0.33j$, 0.76, and 0.83.

The capability of the CVA identified model to simulate the process was tested by comparing the measured process responses with the simulated responses using the same input sequences. The comparisons between the actual process and model responses are shown in Figures 5 and 6.

4 CONCLUSIONS

A new identification strategy has been used to estimate process models of three chemical processes. The approach is based on canonical variate analysis to select a state coordinate system that relates inputs to future outputs. Regression techniques are then used to determine a multi-input, multi-output process model. The Akaike Information Criterion is used to determine an appropriate model order. The CVA identification methodology showed satisfactory results for the estimation of process models for three case studies -- a simulated continuous stirred-tank reactor, a simulated autothermal counterflow reactor, and an experimental distillation column. The identified models were evaluated under a variety of conditions that emphasized the presence of process nonlinearites and interactions, small sample sizes relative to the process settling time, stiff dynamics, and evaluation at operating conditions different than those used for identification.

ACKNOWLEDGEMENT

The authors wish to acknowledge support of this research work by the National Science Foundation through an SBIR contract.

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., Academic Press, NY, 27 (1976).

Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, NY (1958).

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. Eng. 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. Eng. Sci.*, **38**, 607 (1983b).

Camuto, E. and G. Menga, "Approximate realization of discrete time stochastic processes," *Sixth IFAC Symp. Ident. System Param. Estimation*, Washington D.C. (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," *Amer. Math. Soc. Trans.*, Series (2), **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 D.C. (1982).

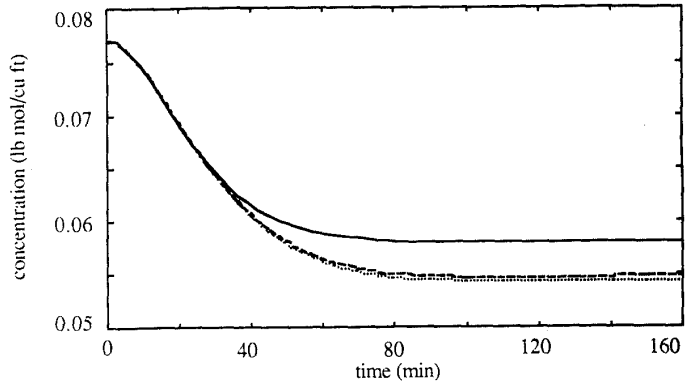

Fig.1. The response of the reactor concentration for a step change of 30 °R in the coolant temperature (___ nonlinear model; - - - CVA model; ... linearized model).

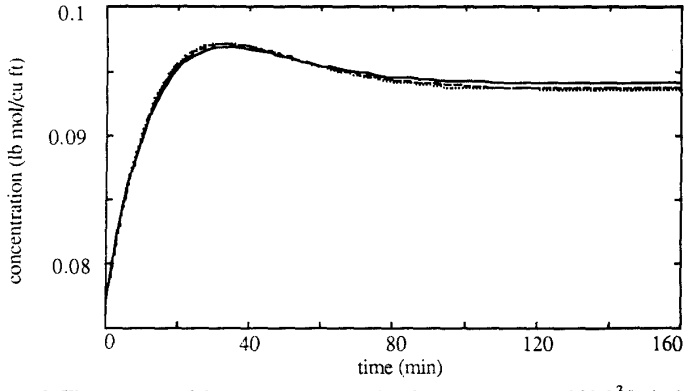

Fig.2. The response of the reactor concentration for a step change of 30 ft^3/hr in feed flow rate (___ nonlinear model; - - - CVA model; ... linearized model).

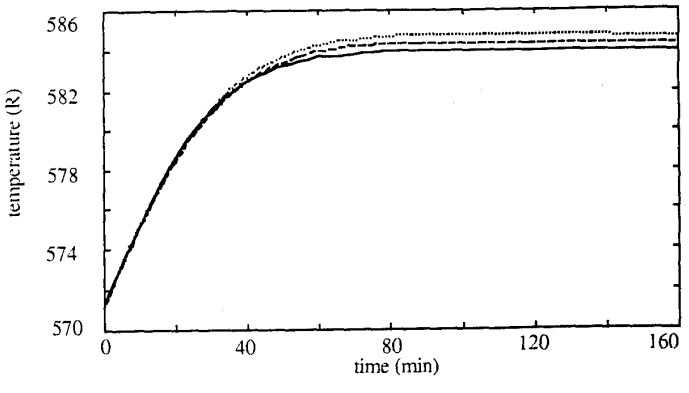

Fig.3. The response of the reactor temperature for a step change of 30 °R in the coolant temperature (___ nonlinear model; - - - CVA model; ... linearized model).

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

Kailath, T. "A view of three decades of linear filter theory," *IEEE Trans. Info. 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. Amer. Control Conf.*, 445 (1983b).

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. Conf. Dec. Control*, 675 (1984).

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

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

McDermott, P. E., Ph.D. Dissertation, Univ. of California, Santa Barbara (1984).

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

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

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

Young, R. E., Ph.D. Dissertation, Univ. of California, Santa Barbara (1988).

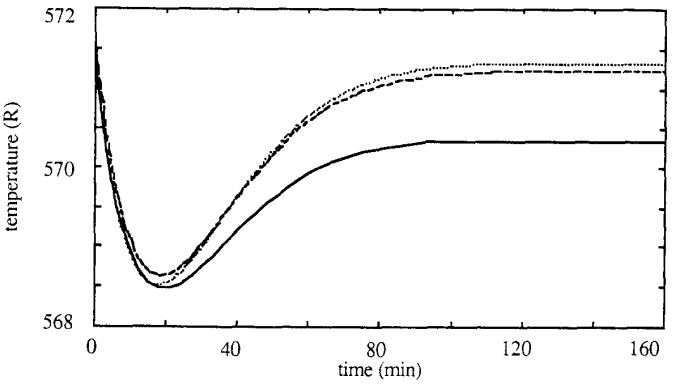

Fig.4. The response of the reactor temperature for a step change of 30 ft^3/hr in the feed flow rate (___ nonlinear model; - - - CVA model; ... linearized model).

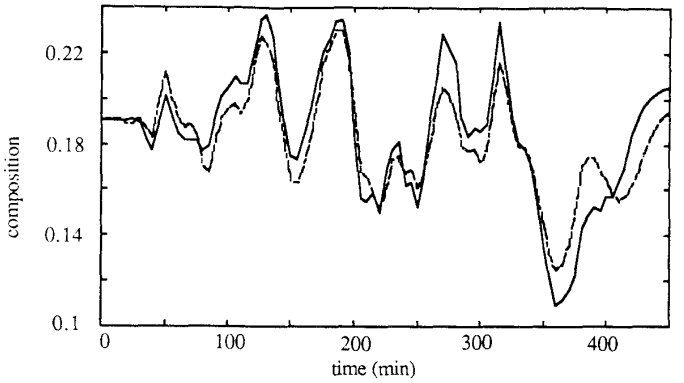

Fig. 5. Model prediction and measurement of the distillate composition of n-butanol (— measurement; --- CVA model prediction).

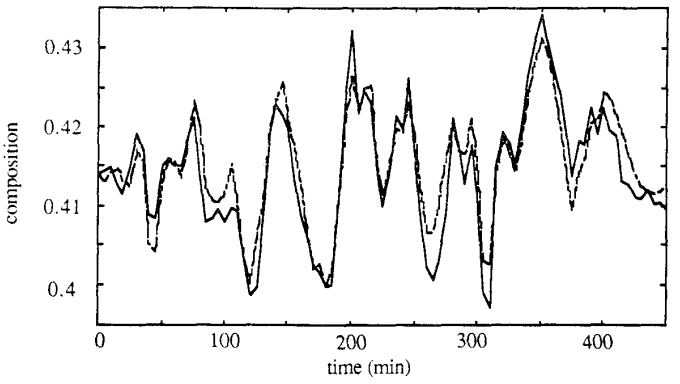

Fig. 6. Model prediction and measurement of the bottoms composition of n-butanol (— measurement; --- CVA model prediction).