NONLINEAR SUBSPACE MODEL IDENTIFICATION

Ali Cinar and Jeffrey DeCicco [1]

Illinois Institute of Technology
Department of Chemical and Environmental Engineering
10W 33rd Street, Chicago, IL, 60616 USA

Abstract: Canonical variates state space (CVSS) modeling is a popular subspace linear model identification technique. A nonlinear extension of CVSS modeling approach was proposed (DeCicco and Cinar, 2000). The modeling procedure consists of two steps: development of a multivariable nonlinear model for a set of latent variables and the linking of the latent variables to outputs of the process. The nonlinear model is structured like a Generalized Additive Model (GAM) and is estimated with CANALS, a nonlinear canonical variate analysis algorithm. This communication presents the methodology and an illustrative example of chemical reactor modeling using data generated from a detailed polymerization reactor model.

Keywords: subspace identification, nonlinear autoregressive model, chemical reactor model identification

1. INTRODUCTION

Canonical variate analysis (CVA) and canonical variate (CV) regression are powerful methods used for developing linear dynamic models. Most notably they are used in subspace modeling to estimate linear state space models (Larimore, 1990*b*). Subspace methods are attractive because of their ease in which they can model multivariate systems. An extension of linear CVA for finding nonlinear state space models was examined (Larimore, 1990*a*) where use of alternating conditional expectation (ACE) algorithm (Breiman and Friedman, 1985) was suggested as the nonlinear CVA method. The examples used linear CVA to model a system by augmenting the linear system with polynomials of past outputs.

Subspace modeling can be cast as a reduced rank regression (RRR) of collections of future outputs on past inputs and outputs after removing the effects of future inputs. CVA performs this RRR. In the case of a linear system, an approximate Kalman filter sequence is recovered from this regression. The state space coefficient matrices are recovered from the state sequence. The nonlinear approach extends this regression to allow for possible nonlinear transformations of the past inputs and outputs, and future inputs and outputs before RRR is performed. The model structure consists of two sub-models. The first model is a multivariable dynamic model for a set of latent variables, the second relates these latent variables to outputs. The latent variables are linear combinations of nonlinear transformations of past inputs and outputs. These nonlinear transformations or functions are found using CANALS (van der Burg and de Leeuw, 1983). Using nonlinear CVA to fit dynamic models is not new. ACE algorithm was used to visually infer nonlinear functions for single output additive models (Chen and Tsay, 1993). This work differs in that the nonlinear functions estimated are directly utilized for prediction. Also, a collection of multiple future outputs is consid-

[1] Corresponding author (e-mail : *cinar@iit.edu).*

ered, which leads to the latent variables model structure. The latent variables are then linked to the outputs using linear projection type nonlinear model structures such as projection pursuit regression (PPR) (Friedman and Stuetzel, 1981) or a linear model through least squares regression.

2. NONLINEAR MODEL STRUCTURE

Let $\mathbf{y}_t \in \mathcal{R}^l$, $\mathbf{u}_t \in \mathcal{R}^m$ and $\mathbf{x}_t \in \mathcal{R}^n$ represent outputs, inputs and latent variables that are collections of individual variables y_t^i, u_t^i and x_t^i,

$$\mathbf{y}_t = \begin{bmatrix} y_t^1 \\ \vdots \\ y_t^l \end{bmatrix}, \ \mathbf{u}_t = \begin{bmatrix} u_t^1 \\ \vdots \\ u_t^m \end{bmatrix}, \ \mathbf{x}_t = \begin{bmatrix} x_t^1 \\ \vdots \\ x_t^n \end{bmatrix}. \quad (1)$$

The model structure for a single latent variable is

$$x_{t+\beta}^i = \sum_{j=1}^{\beta} \sum_{k=1}^{m} h_{i,j+k-1}^* \theta_{k,j}^p \left(u_{t+\beta-j}^k\right) + \sum_{j=1}^{\beta} \sum_{k=1}^{l} h_{i,\beta m+j+k-1}^* \phi_{k,j}^p \left(y_{t+\beta-j}^k\right) \quad (2)$$

where h^* are scalar coefficients, θ^p and ϕ^p are nonlinear functions, and β is the past window length. The model structure linking the latent variables to a single output is

$$y_t^i = y_{ss}^i + \sum_{j=1}^{M} n_{i,j} \psi_j \left(\mathbf{l}_j^T \mathbf{x}_t\right), \quad (3)$$

where $n_{i,j}$ are scalar coefficients, $\mathbf{l}_j$ are $n \times 1$ coefficient vectors, ψ_j is some nonlinear function, and y_{ss}^i is a steady state operating point.

The structure for the latent variables model in (2) is of a generalized additive model (GAM) (Hastie and Tibshirani, 1990). But it is developed using a nonlinear CVA method discussed in Section 4.

3. LINEAR MODELING

A number of methods fall under this framework including CVA (Larimore, 1990*b*) and N4SID (Van Overschee and De Moor, 1994). An estimate of Kalman filter states can be recovered from a RRR of a collection of future outputs on past collection of inputs and outputs, and future inputs in linear systems (Van Overschee and De Moor, 1994). "Past" and "future" discriminate previously observed historical data used in the estimation of a causal dynamic model.

Define the collection of past and future observed outputs:

$$\mathbf{y}_\beta(t) = [\mathbf{y}_t^T \ \mathbf{y}_{t+1}^T \ \cdots \ \mathbf{y}_{t+\beta-1}^T]^T, \quad (4)$$
$$\mathbf{y}_\gamma(t) = [\mathbf{y}_{t+\beta} \ \mathbf{y}_{t+\beta+1}^T \ \cdots \ \mathbf{y}_{t+\alpha-1}^T]^T,$$

where $\alpha = \beta + \gamma$ and γ is the future window length. The collections of past and future inputs ($\mathbf{u}_\beta$ and $\mathbf{u}_\gamma$) are defined in the same manner. Let $t+\beta-1$ represent the present time. Observations of the above collections are

$$\mathbf{Y}_\beta = [\mathbf{y}_\beta(1) \ \cdots \ \mathbf{y}_\beta(N)] \quad (5)$$
$$\mathbf{Y}_\gamma = [\mathbf{y}_\gamma(1) \ \cdots \ \mathbf{y}_\gamma(N)]$$

Collections of observed past and future inputs are $\mathbf{U}_\beta$ and $\mathbf{U}_\gamma$, respectively. The objective is to extract a causal model that predicts future outputs using assigned future input values:

$$\mathbf{Y}_\gamma = \mathbf{L_1}\mathbf{U}_\beta + \mathbf{L_2}\mathbf{Y}_\beta + \mathbf{L_3}\mathbf{U}_\gamma \quad (6)$$

where $\mathbf{L}_1$, $\mathbf{L}_2$, and $\mathbf{L}_3$ are coefficient matrices.

The RRR problem that leads to approximate state variables sequence is formulated as

$$\min_{\mathbf{L}_1, \mathbf{L}_2, \mathbf{\Phi}_\gamma} \| \mathbf{Y}_\gamma - [\mathbf{L_1}\mathbf{U}_\beta + \mathbf{L_2}\mathbf{Y}_\beta + \mathbf{L_3}\mathbf{U}_\gamma] \|_{\mathbf{W}}^2 \quad (7)$$

where $\| \mathbf{M} \|_{\mathbf{W}} = \mathrm{tr}\left(\mathbf{M}^T\mathbf{W}\mathbf{M}\right)$. The regression problem has a reduced rank structure in that $[\mathbf{L}_1 \ \mathbf{L}_2] = \mathbf{\Gamma}\mathbf{H}$, where $\mathbf{\Gamma}$ and $\mathbf{H}$ both have rank n. The CVA RRR solution is derived by setting $\mathbf{W} = \left(\mathbf{Y}_\gamma \mathbf{\Pi}_\gamma^\perp \mathbf{Y}_\gamma^T\right)^{-1}$ (Jansson and Wahlberg, 1999). Estimates of $\mathbf{\Gamma}$ and $\mathbf{H}$ are obtained from singular value decomposition (SVD)

$$\mathbf{W}^{\frac{1}{2}} \mathbf{Y}_\gamma \mathbf{\Pi}_\gamma^\perp \mathbf{P}_\beta^T \left(\mathbf{P}_\beta \mathbf{\Pi}_\gamma^\perp \mathbf{P}_\beta^T\right)^{-\frac{1}{2}} = \mathbf{Q}\mathbf{\Sigma}\mathbf{S}^T \quad (8)$$

$$\mathbf{\Pi}_\gamma^\perp = \mathbf{I} - \mathbf{U}_\gamma^T \left(\mathbf{U}_\gamma \mathbf{U}_\gamma^T\right) \mathbf{U}_\gamma, \quad \mathbf{P}_\beta = \begin{bmatrix} \mathbf{U}_\beta \\ \mathbf{Y}_\beta \end{bmatrix} \quad (9)$$

The real matrices $\mathbf{Q}$ and $\mathbf{S}$ contain the canonical variate vectors, and $\mathbf{\Sigma}$ contains the canonical correlations of the CVA decomposition (8). The estimates become

$$\mathbf{\Gamma} = \mathbf{W}^{-\frac{1}{2}} \times \left(\text{first } n \text{ columns of } \mathbf{Q}\mathbf{\Sigma}^{\frac{1}{2}}\right), \quad (10)$$
$$\mathbf{H} = \left(\text{first } n \text{ rows of } \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{S}^T\right) \left(\mathbf{P}_\beta \mathbf{\Pi}_\gamma^\perp \mathbf{P}_\beta^T\right)^{-\frac{1}{2}}.$$

The approximate state variables sequence is

$$\mathbf{X} = \mathbf{H} \begin{bmatrix} \mathbf{U}_\beta \\ \mathbf{Y}_\beta \end{bmatrix}, \quad (11)$$

with $\mathbf{X} = [\hat{\mathbf{x}}(\beta+1) \ldots \hat{\mathbf{x}}(\beta+N)]$. With the estimated state variables sequence (11) and the observed inputs and outputs it is possible to estimate the model coefficients of the linear state space model including the Kalman filter gain (Larimore, 1990*b*; Van Overschee and De Moor, 1994).

4. NONLINEAR MODEL IDENTIFICATION

The nonlinear model identification is an extension of the linear approach. First nonlinear transformations of observed data are sought in the regression

of future outputs ($\mathbf{Y}_\gamma$) on past inputs and outputs, and future inputs ($\mathbf{U}_\beta$, $\mathbf{Y}_\beta$, and $\mathbf{U}_\gamma$). This leads to a set of latent variables that are nonlinear functions of past inputs and outputs. Then, the reduced rank structure and coefficient matrices are estimated using CVA as in the linear case. Finally, the model structure between latent variables and outputs is developed using projection pursuit (Friedman and Stuetzel, 1981) or other methods.

4.1 *Latent Variable Model Identification*

The nonlinear transformations are estimated by a modified version of the nonlinear CVA technique CANALS (van der Burg and de Leeuw, 1983). Once the nonlinear transformations are found, linear CVA is used to find the linear combinations that form the latent variable model. The regression model (6) is first generalized as

$$\mathbf{Z}_\gamma = \mathbf{L}_1\mathbf{V}_\beta + \mathbf{L}_2\mathbf{Z}_\beta + \mathbf{L}_3\mathbf{V}_\gamma, \tag{12}$$

where $\mathbf{L}_1$, $\mathbf{L}_2$, and $\mathbf{L}_3$ are coefficient matrices, $\mathbf{Z}_\gamma$, $\mathbf{Z}_\beta$, $\mathbf{V}_\gamma$, and $\mathbf{V}_\beta$ are nonlinear transformations of the past and future inputs and outputs:

$$\begin{aligned}
\mathbf{Z}_\gamma &= [\mathbf{z}_\gamma(1) \ \dots \ \mathbf{z}_\gamma(N)]^T \\
\mathbf{Z}_\beta &= [\mathbf{z}_\beta(1) \ \dots \ \mathbf{z}_\beta(N)]^T \\
\mathbf{V}_\gamma &= [\mathbf{v}_\gamma(1) \ \dots \ \mathbf{v}_\gamma(N)]^T \\
\mathbf{V}_\beta &= [\mathbf{v}_\beta(1) \ \dots \ \mathbf{v}_\beta(N)]^T \\
\mathbf{z}_\gamma(t) &= \left[\Phi_1^f(\mathbf{y}_{t+\beta})^T \ \dots \ \Phi_\gamma^f(\mathbf{y}_{t+\alpha-1})^T\right]^T \\
\mathbf{z}_\beta(t) &= \left[\Phi_\beta^p(\mathbf{y}_t)^T \ \dots \ \Phi_1^p(\mathbf{y}_{t+\beta-1})^T\right]^T \\
\mathbf{v}_\gamma(t) &= \left[\Theta_1^f(\mathbf{u}_{t+\beta})^T \ \dots \ \Theta_\gamma^f(\mathbf{u}_{t+\alpha-1})^T\right]^T \\
\mathbf{v}_\beta(t) &= \left[\Theta_\beta^p(\mathbf{u}_t)^T \ \dots \ \Theta_1^p(\mathbf{u}_{t+\beta-1})^T\right]^T
\end{aligned}$$

Φ and Θ are vector valued functions of the form

$$\Phi_i^p(y_t) = \begin{bmatrix} \phi_{1,i}^p(y_t^1) \\ \vdots \\ \phi_{l,i}^p(y_t^l) \end{bmatrix}, \quad \Theta_i^p(u_t) = \begin{bmatrix} \theta_{1,i}^p(u_t^1) \\ \vdots \\ \theta_{m,i}^p(u_t^m) \end{bmatrix}.$$

The observed inputs and outputs have been centered around their means or some steady state operating point of the process.

CANALS was originally developed for analysis of categorical data (van der Burg and de Leeuw, 1983) and did not utilize a locally adaptive regression technique. A modified version of CANALS estimates the nonlinear functions with nonparametric regression (DeCicco and Cinar, 2000). The final nonparametric estimates are interpolated by Chebychev polynomials to allow a smooth interpolation. The modified CANALS is used to estimate the coefficient matrices and nonlinear transformations. CANALS seeks to minimize

$$\min_{\tilde{\mathbf{L}},\mathbf{L}_4,\mathbf{G},\mathbf{Z}_\gamma} \| \mathbf{L}_4\mathbf{Z}_\gamma - \tilde{\mathbf{L}}\mathbf{G} \|^2 \tag{13}$$

where $\mathbf{L}_4$ is a canonical variate coefficient matrix, $\tilde{\mathbf{L}} = [\mathbf{L}_1 \ \mathbf{L}_2 \ \mathbf{L}_3]$, and $\mathbf{G}^T = \left[\mathbf{V}_\beta^T \ \mathbf{Z}_\beta^T \ \mathbf{V}_\gamma^T\right]$. The loss function (13) is minimized such that the nonlinear functions have zero mean and unit variance with the constraints:

$$\begin{aligned}
\mathbf{L}_4\tilde{\mathbf{Z}}_\gamma\left(\mathbf{L}_4\tilde{\mathbf{Z}}_\gamma\right)^T &= (N+\alpha-1)\mathbf{I}, \\
\tilde{\mathbf{L}}\mathbf{G}\left(\tilde{\mathbf{L}}\mathbf{G}\right)^T &= (N+\alpha-1)\mathbf{I},
\end{aligned}$$

where $\mathbf{I}$ is the identity matrix.

CANALS uses alternating least squares (ALS) to estimate coefficient matrices and nonlinear transformations. The ALS method works iteratively until convergence of (13). $\mathbf{L}_4$ and $\tilde{\mathbf{L}}$ are estimated by CVA RRR. The nonlinear transformations are estimated using a back-fitting approach (Hastie and Tibshirani, 1990). The supersmoother of (Friedman, 1984) is used to estimate the nonlinear functions in the back-fitting step.

The number of latent variables and linear combinations of nonlinear functions that make up these latent variables are determined by using linear CVA between the $\mathbf{Z}_\gamma$ and $\tilde{\mathbf{P}}_\beta = \left[\mathbf{V}_\beta^T \ \mathbf{Z}_\beta^T\right]^T$ after the effect of $\mathbf{V}_\gamma$ is removed. Use SVD

$$\left(\mathbf{Z}_\gamma\tilde{\mathbf{\Pi}}_\gamma^\perp\mathbf{Z}_\gamma^T\right)^{-\frac{1}{2}}\mathbf{Z}_\gamma\tilde{\mathbf{\Pi}}_\gamma^\perp\tilde{\mathbf{P}}_\beta^T\left(\tilde{\mathbf{P}}_\beta\tilde{\mathbf{\Pi}}^\perp\tilde{\mathbf{P}}_\beta^T\right)^{-\frac{1}{2}} = \mathbf{Q}\mathbf{\Sigma}\mathbf{S}^T,$$

$$\tilde{\mathbf{\Pi}}_\gamma^\perp = \mathbf{I} - \mathbf{V}_\gamma^T\left(\mathbf{V}_\gamma\mathbf{V}_\gamma^T\right)^{-1}\mathbf{V}_\gamma. \tag{14}$$

The linear combinations of estimated states are

$$\mathbf{H}^* = \left(\text{first } n \text{ rows of } \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{S}^T\right) \times \left(\tilde{\mathbf{P}}_\beta\tilde{\mathbf{\Pi}}_\gamma^\perp\tilde{\mathbf{P}}_\beta^T\right)^{-\frac{1}{2}}.$$

The latent variable sequence becomes

$$\tilde{\mathbf{X}}(t) = \mathbf{H}^*\tilde{\mathbf{P}}_\beta, \tag{15}$$

where $\tilde{\mathbf{X}}(t) = [\tilde{\mathbf{x}}_{t+\beta} \ \tilde{\mathbf{x}}_{t+\beta+1} \dots \tilde{\mathbf{x}}_{t+\beta+N-1}]$. This leads to the latent variable model structure

$$\tilde{\mathbf{x}}(t+\beta) = \mathbf{H}^*\begin{bmatrix} \mathbf{v}_\beta(t) \\ \mathbf{z}_\beta(t) \end{bmatrix} \tag{16}$$

where $\mathbf{h}_i^*$ is the ith column of $\mathbf{H}^*$, and (16) is a generalization of (2). The number of latent variables n is chosen by inspecting the singular values of $\mathbf{\Sigma}$ in (14). Significant latent variables have relatively large singular singular values.

4.2 *Link Function*

The relationship between latent variables and outputs is generalized by a PPR model. This type

of model structure includes linear least squares and GAM model structures. We build a linear model estimated by CVA regression and a PPR model, and compare their performances. The linear model is a special case of the PPR structure. If a linear model is adequate the overall model structure is simpler. The multivariable version of (3) relating latent variables to outputs is

$$\mathbf{y}_t = \mathbf{y}_{ss} + \sum_{i=1}^{M} \mathbf{n}_i \psi_i \left(\mathbf{l}_i^T \mathbf{x}_t \right). \qquad (17)$$

PPR seeks to minimize the loss function

$$SSE = \sum_{j=1}^{l} \sum_{t=1}^{N+\alpha-1} \left(y_t^j - \sum_{i=1}^{M} n_{i,j} \psi_i (\mathbf{l}_i^T \mathbf{x}_t) \right)^2 \qquad (18)$$

with respect to $n_{i,j}$, $\mathbf{l}_i$, ψ_i, and M. A back-fitting procedure is utilized to perform the regression. First, $\mathbf{l}_i$ is found by minimizing (18) using numerical optimization. Next, ψ_i is found by nonparametric regression and lastly $\mathbf{n}_i$ is estimated with CVA regression. The number of terms M is chosen by using (18) (See Figure 1).

With the estimated latent variable sequence $\tilde{\mathbf{X}}(1)$ of (15) the loss function (18) may be written as

$$\min_{\mathbf{N},\mathbf{\Psi}\left(\mathbf{L}_x(\bar{\mathbf{X}}(1))\right)} \| \mathbf{Y}_\gamma[1,:] - \mathbf{N}\mathbf{\Psi}\left(\mathbf{L}_x^T \tilde{\mathbf{X}}(1)\right) \|^2 \qquad (19)$$

$$\mathbf{N} = [\mathbf{n}_1 \ \mathbf{n}_2 \ \ldots \ \mathbf{n}_M], \quad \mathbf{L}_x = [\mathbf{l}_1 \ \mathbf{l}_2 \ \ldots \ \mathbf{l}_M],$$

$$\mathbf{\Psi}\left(\mathbf{L}_x^T \tilde{\mathbf{X}}(1)\right) = \begin{bmatrix} \psi_1(\mathbf{l}_1^T \mathbf{x}_{t+\beta}) & \ldots & \psi_1(\mathbf{l}_1^T \mathbf{x}_{t+\beta+N-1}) \\ \psi_2(\mathbf{l}_2^T \mathbf{x}_{t+\beta}) & \ldots & \psi_2(\mathbf{l}_2^T \mathbf{x}_{t+\beta+N-1}) \\ \vdots & & \vdots \\ \psi_M(\mathbf{l}_M^T \mathbf{x}_{t+\beta}) & \ldots & \psi_M(\mathbf{l}_M^T \mathbf{x}_{t+\beta+N-1}) \end{bmatrix}.$$

4.3 *Nonlinear Function Estimation*

ALS and back-fitting approach in CANALS and PPR algorithms for estimating ϕ, θ, and ψ involve use of a regression technique capable of capturing nonlinear relationships. There are several regression techniques such as orthogonal polynomials, neural networks, local polynomial regression, smoothing splines, and kernel smoothers that may be grouped in terms of their common properties such as parametric/ nonparametric, fixed/adaptive, or local/global. No one technique is strictly superior to the others.

The modeling procedure requires a flexible, automated, and robust regression technique. Kernel regression, smoothing splines, and local polynomial regression techniques are flexible since they easily adapt to data because they are not constrained to any global parametric structure. Automating such techniques is a difficult task especially for serially correlated data. These techniques require the selection of some parameter value which determines the degree of smoothing such as bandwidth. Cross-validation is used frequently to determine the degree of smoothing, but for serially correlated data it may over-smooth or under-smooth. To avoid such difficulties, a fixed bandwidth is used throughout the regression.

These techniques are also sensitive to outliers. Robust techniques exist such as locally weighted scatter plot smoothing (LOWESS) (Cleveland, 1979). LOWESS iteratively smoothes with local polynomials. At each iteration, weights that are inversely proportional to the magnitude of residuals from the previous iteration are assigned to data and the regression is repeated. This greatly reduces sensitivity to outliers. In this work, assignment of weights to data and iterative smoothing are carried out by a supersmoother (Friedman, 1984) which is a local, adaptive, nonparametric regression technique. The functions ϕ, θ and ψ based on supersmoother estimates are local in the sense they are only defined within the domain of the data from which they are developed. Outside this domain, a linear relationship can be assumed. Let $\psi(x)$ represent the estimated supersmoother function where $l \leq x \leq u$. The final function including extrapolation is $\varphi(x)$

$$\varphi(x) = \begin{cases} a^l + b^l x & : \quad x < l \\ \psi(x) & : \quad l \leq x \leq u \\ a^u + b^u x & : \quad x > u \end{cases} \qquad (20)$$

Initial estimates of a^l, b^l, and a^u, b^u are found by regressing the lower and upper quartiles of observed $\psi(x)$ on x, respectively. The intercept terms a^l and a^u are then adjusted such that $a^l + b^l \min(x) = \psi(\min(x))$ and $a^u + b^u \max(x) = \psi(\max(x))$. Shifting of the intercepts allows for a smooth transition between domains. The functions estimated by the supersmoother are not considered continuous functions. Interpolation is done by regressing observed $\psi(x)$ on x using orthogonal Chebychev polynomials.

5. CSTR POLYMERIZATION MODELING

Data from a poly-vinyl acetate CSTR simulation (Teymour, 1989) is used to illustrate model identification. The outputs are reactor temperature (y_T) and number average molecular weight (MWn) of the polymer (y_M), and the manipulated input is residence time. The steady state gain of the system is not constant in the region of operation selected.

The input and outputs are sampled at 5 *min* intervals, and there is a time delay of 5 *min* for the input. The base input residence time levels are set at random from a uniform distribution between 10 and 90 *min* with a switching probability of 0.95. Added to this input is a signal with levels between -5 and 5. The range of input was chosen to exaggerate the nonlinearity of the system. Input switch levels are based on a uniform distribution with a switching probability of 0.95 and 0.80. 3000 samples of inputs and outputs are collected. To avoid numerical round off errors, MWn is divided by 1000 before model development. A known steady state operating point at a residence time of 50*min* was used to center the data. Gaussian random measurement noise was added to outputs prior to model development.

Latent Variable Model Identification. The latent variable model identification requires the specification of γ, β, and n. For this example, future and past horizons of 10 is used ($\gamma = \beta = 10$). The number of latent variables n is chosen by investigating the singular values of (14). For comparison, a linear model is also developed by linear CVA subspace identification.

Figure 1 shows the canonical correlation squared for linear and nonlinear models. The singular values for the linear and nonlinear model drop off significantly after 5 and 10 latent variables, respectively. The singular values can be interpreted as the canonical correlation between the future outputs and past inputs and outputs, after the effect of future inputs is removed. The canonical correlations of the nonlinear model are greater than the linear model. This is expected because the CANALS nonlinear CVA algorithm seeks to find nonlinear transformations of the original variables that maximize the canonical correlation. **Model Comparison**. A linear state space model

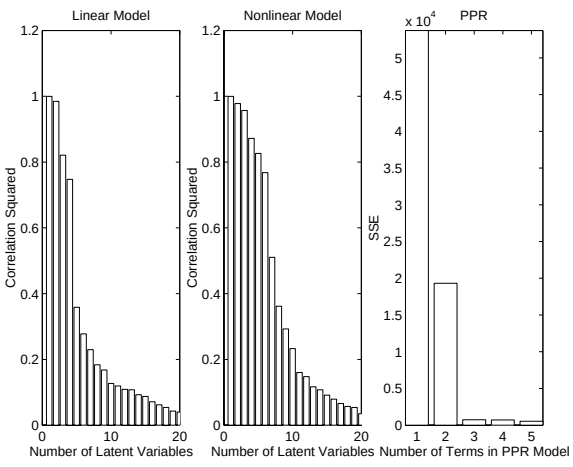

Fig. 1. Latent variable and PPR order selection.

found with CVA subspace modeling was compared to the nonlinear model with a linear or PPR link function. In-sample and out-of-sample prediction is evaluated by sum of squared error (SSE). Prediction consists of initializing the models from observed data then recursively simulating future outputs based on actual measured inputs and either past predicted states in the case of the state space model, or past predicted outputs in the case of the nonlinear model. To determine the form of the link function PPR was compared to linear least squares. The PPR model developed had 5 terms (M=5). The relative magnitudes of SSE in Figure 1 indicates that 3 terms are sufficient. A plot of the in-sample prediction of reactor temperature for various models (Figure 2) indicate that nonlinear models outperform the linear model. The SSE used for comparison is

$$SSE = \sum_t \left[\left(\frac{y_M(t) - \hat{y}_M(t)}{\sigma_M} \right)^2 + \left(\frac{y_T(t) - \hat{y}_T(t)}{\sigma_T} \right)^2 \right]$$

where y and $\hat{y}$ are the actual and predicted outputs respectively, and σ is the standard deviation of the observed variable. The SSE is 530 for the nonlinear model with linear link function, 558 for the nonlinear model with PPR link function, and 930 for the linear state space model. For the

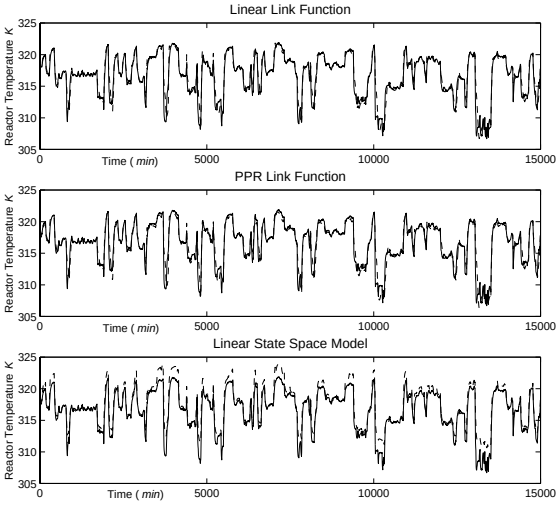

Fig. 2. Actual (-) and predicted (- -) reactor temperature. Top: nonlinear model with linear link; Center: nonlinear model with PPR link, and Bottom: linear model.

out-of-sample comparison 20 runs with different inputs and noise sequences were simulated in the same manner as the in-sample case. Nonlinear models clearly outperform the linear model with respect to SSE (Figure 3). The nonlinear models with linear and PPR link function have comparable performance with the linear link function model with slightly better performance. **Steady-State Analysis.** Numerical continuation is implemented with AUTO (Doedel *et al.*, 1998) to determine the steady-state characteristics of the

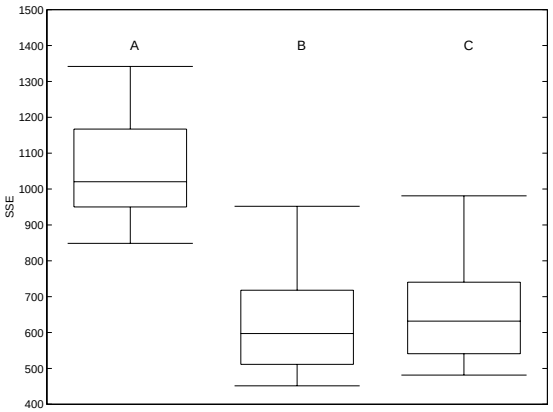

Fig. 3. SSE for 20 runs for (A) linear model, (B) nonlinear model with linear link, and (C) nonlinear model with PPR link.

empirical model. AUTO uses numerical continuation to trace out the fixed point solution given an initial steady state. The resulting steady-state curve is then compared to the steady-state curve of the physical model. For this comparison the linear link model was used. Let y_M^{ss} and y_T^{ss} be the steady state values of the outputs that correspond to the input u^{ss}. For the case of a least squares linear link function the model at steady state is

$$y_M^{ss} = \sum_{j=1}^{\beta} a_j^1 y_M^{ss} + \sum_{j=1}^{\beta} b_j^1 y_T^{ss} + \sum_{j=1}^{\beta} c_j^1 u^{ss}, \quad (21)$$

$$y_T^{ss} = \sum_{j=1}^{\beta} a_j^2 y_M^{ss} + \sum_{j=1}^{\beta} b_j^2 y_T^{ss} + \sum_{j=1}^{\beta} c_j^2 u^{ss}. \quad (22)$$

Figure 4 compares the actual and predicted fixed point steady state solutions. The nonlinear model predicts the fixed point solution inside the domain of the experimental data and the linear approximation works well outside.

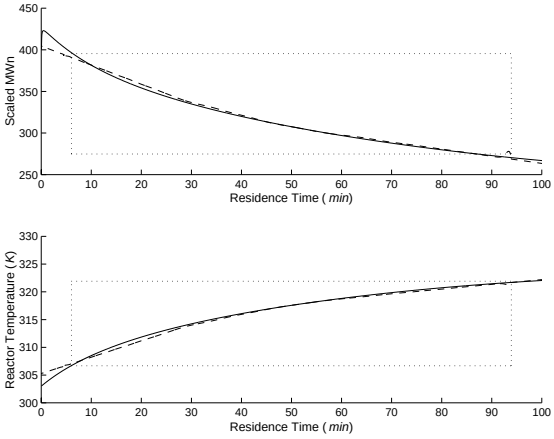

Fig. 4. Fixed point steady state solution of the actual physical model (-) and nonlinear empirical model (- -). The domain of experimental data is inside the dashed box (:).

6. CONCLUSION

The multivariate nonlinear empirical dynamic modeling technique is the extension of linear CVA subspace identification. A case study on modeling a polymerization in a CSTR illustrates the modeling approach and the dynamic and steady state performance of the nonlinear model which are better than the linear model performance.

7. REFERENCES

Breiman, L. and J. H. Friedman (1985). Estimating Optimal Transformations for Multiple Regression and Correlation. *J. Amer. Statist. Assoc.* **80**, 580–598.

Chen, R. and R. S. Tsay (1993). Nonlinear Additive ARX Models. *J. Amer. Statist. Assoc.* **88**(423), 955–967.

Cleveland, W.S. (1979). Locally-weighted Regression and Smoothing Scatterplots. *J. Amer. Stat. Assoc.* **74**, 829–836.

DeCicco, J. and A. Cinar (2000). Empirical Modeling of Systems with Output Multiplicities by Multivariate Additive NARX Models. *Indust. Eng. Chem. Research* **39**(6), 1747–1755.

Doedel, E. J., A. R. Champneys, T. F. Fairgrieve, Y. A. Kuznetsov, B. Sandstede and X. Wang (1998). AUTO 97: Continuation and Bifurcation Software for ODEs).

Friedman, J. and W. Stuetzel (1981). Projection Pursuit Regression. *J. Am. Stat. Assoc.* **76**, 817–823.

Friedman, J. H. (1984). A Variable Span Smoother. Technical Report No. 5. Dept. of Statistics, Stanford University.

Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models.* Monographs on Statistics and Applied Probability. Chapman & Hall.

Jansson, M. and B. Wahlberg (1999). A Linear Regression Approach to State-Space Subspace System Identification. *Signal Processing* **52**(2), 103–129.

Larimore, W. (1990*a*). Identification and filtering of nonlinear systems using canonical variate analysis. In: *Nonlinear Modeling and Forecasting.* Addison-Wesley.

Larimore, W. (1990*b*). Canonical Variate Analysis in Identification, Filtering, and Adaptive Control. In: *Proc IEEE Conf. Decision Control.*

Teymour, F. (1989). The Dynamic Behavior of Free Radical Solution Polymerization Reactions in a Continuous Stirred Tank Reactor. PhD thesis. University of Wisconsin.

van der Burg, E. and J. de Leeuw (1983). Nonlinear Canonical Correlation. *British J. Math. Statist. Psychol.* **36**, 54–80.

Van Overschee, P. and B. De Moor (1994). N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems. *Automatica* **30**(1), 75–93.