

# Identification of Multivariable, Linear, Dynamic Models: Comparing Regression and Subspace Techniques

**Ben C. Juricek**

*Toyon Research Corporation, 75 Aero Camino Road, Goleta, California 93117*

**Dale E. Seborg\***

*Department of Chemical Engineering, University of California, Santa Barbara, California 93106*

**Wallace E. Larimore**

*Adaptics, Inc., 1717 Briar Ridge Road, McLean, Virginia 22101*

Dynamic models identified as ARX or FIR models using two regression techniques, PLS and CCR, are compared with state-space models identified using a predictive error method and two subspace algorithms, CVA and N4SID. The objective functions for PLS and CCR are shown to be related. A comprehensive simulation study of several processes with different characteristics and noise properties is used to compare the identification methods. The results indicate that, if the time delay structure is known or estimated accurately, the identified subspace models tend to be more accurate than the models identified using regression. The state-space models identified using the CVA algorithm are especially accurate.

## Introduction

A key requirement for many advanced control and monitoring techniques is the availability of an accurate process model. Developing dynamic models on the basis of fundamental physicochemical relationships is often prohibitively difficult for industrial applications on account of unknown chemical reactions, poorly known or unknown kinetic coefficients, etc. Thus, decades of research have been devoted to the development of models from empirical input–output data.

Most industrial processes, and almost all found in the chemical industry, are multivariable and nonlinear, as well as constantly responding to disturbances that are unmeasurable and occur at unknown times. Although almost all processes are nonlinear, in practice, linear models are commonly used for control and monitoring. This practice is not entirely unjustified, as many processes are operated within a localized region and the nonlinearities are quite mild within this operating region (e.g., refinery processes). The important issue for industrial processes is therefore estimating a multivariable, linear model that is appropriate for control and monitoring applications. The word “appropriate” is subjective but reflects the ability of the model to describe the system, and not just fit the data—what Ljung<sup>1</sup> refers to as the difference between *system identification* and *curve fitting*.

The field of system identification is quite mature. The textbooks of Box et al.,<sup>2</sup> Ljung,<sup>3</sup> and Söderström and Stoica<sup>4</sup> contain the fundamental theory for identifying discrete, stochastic, linear models. Each book describes the necessary steps and decisions for identifying empirical models: selecting the model structure, the model order, and the parameter estimation method. For multi-input–multi-output (MIMO) models, these decisions (in particular, the model structure) are especially challeng-

ing and generally left for an “expert” to specify. Several researchers have since studied strategies that systematically identify large-dimensional models. A systematic methodology is particularly important for, e.g., model predictive control applications, where an expert in MIMO system identification is not always available. This paper compares methods that purport to identify accurate linear models for multivariable systems with little input from the user—true “black-box” models.

In the next section, the generic identification problem is presented, followed by a discussion of methods from multivariable statistics and subspace algorithms for identifying dynamic models. Then, a comprehensive simulation study compares the techniques and demonstrates the significant issues for several representative examples.

## Identifying Linear Models

This paper is concerned with linear, discrete-time, stochastic models. The books by Ljung<sup>3</sup> and Söderström and Stoica<sup>4</sup> are excellent references for classical system identification theory. A summary of the relevant issues is presented in this section.

The identification of an empirical process model requires excitation of the process to collect experimental data and then use of these data to identify a dynamic model. After collecting the experimental data, the user makes three important selections, first choosing whether to use a parametric or nonparametric model, then identifying the model structure (i.e., the type of model and the model order), and finally selecting the parameter estimation method. All three factors have significant effects on the identified model, and the relative importance of each factor will be examined.

Consider a linear, discrete-time, stochastic model of the form

$$y(k) = P(q^{-1})u(k) + H(q^{-1})e(k) \quad (1)$$

\* Corresponding author. Fax: 805 893 3352. Phone: 805 893 4731. E-mail: seborg@engineering.ucsb.edu

where  $y(k)$  is an  $(n_y \times 1)$  vector of measured outputs at sample  $k$ ,  $u(k)$  is an  $(n_u \times 1)$  vector of measured inputs, and  $e(k)$  is the  $(n_e \times 1)$  vector of unmeasured disturbances, assumed to be white noise. The backward shift operator,  $q^{-1}$ , shifts time backward by one sample, i.e.,  $q^{-1}w(k) = w(k-1)$ . The system model,  $P(q^{-1})$ , and the noise model,  $H(q^{-1})$ , are functions of  $q^{-1}$ ; they can be transfer functions, polynomials in  $q^{-1}$ , or the equivalent state-space model.

The finite impulse response (FIR) model is written as

$$y(k) = \sum_{i=1}^{N_{\text{FIR}}} G_i u(k-i) + v(k) \quad (2)$$

where  $v(k)$  describes the effect of all of the unmeasurable disturbances affecting the system.

The matrix of FIR coefficients,  $G_i$ , shows the response of process outputs at sample  $k$  to a pulse input at  $k-i$ . Only the model order  $N_{\text{FIR}}$  is specified. Although the coefficient matrices must be estimated, the FIR model is labeled "nonparametric" because the FIR model acts as a function for describing the complete system response to an impulse input.<sup>4</sup>

Finite impulse response models are frequently used for three reasons: (i) there are few restrictions about the processes to be modeled, (ii) the FIR parameters are estimated using least-squares regression, and (iii) the FIR structure is intuitive yet meaningful to control engineers. The major disadvantage of FIR models is that the model order can be quite large, even for very simple processes. This is particularly true for MIMO models of stiff processes that have a wide range of time constants. For high-order models with large input and output dimensions, the estimate of  $G_i$  can be poor as a result of a poorly conditioned regressor matrix and the large number of estimated parameters.

Parametric models, such as autoregressive models with exogenous inputs (ARX) or autoregressive moving average models with exogenous inputs (ARMAX), assume that the true system and noise process are described by a particular model structure. The properties of the model (e.g., the numbers of poles and zeros, and the form of the frequency response) are determined by this structure. A parametric model will generally estimate the true system with fewer parameters than a FIR model, hence yielding a more parsimonious model. Furthermore, the variance of the estimated model parameters is proportional to the number of estimated parameters.<sup>5</sup>

The ARX model can be written as

$$A(q^{-1})y(k) = B(q^{-1})u(k) + e(k) \quad (3)$$

where  $A(q^{-1})$  and  $B(q^{-1})$  are matrix polynomials given by

$$A(q^{-1}) = I + A_1 q^{-1} + \dots + A_{N_A} q^{-N_A} \quad (4)$$

$$B(q^{-1}) = B_1 q^{-1} + \dots + B_{N_B} q^{-N_B} \quad (5)$$

The ARX coefficient matrices,  $A_i$ ,  $B_j$ , can be estimated via least-squares regression. In general, the ARX model will have fewer coefficient matrices than the FIR model, i.e.,  $N_A + N_B < N_{\text{FIR}}$ . However, the ARX model structure assumes that the white noise  $e(k)$  is filtered by  $A^{-1}(q^{-1})$ ,

i.e.,  $H(q^{-1}) = A^{-1}(q^{-1})$  in eq 1. If this assumption is incorrect, the least-squares estimate will be biased from the true values, even if the model structure and order for  $P(q^{-1})$  in eq 1 are correct.<sup>3,6</sup>

The ARMAX model is a more general version of the ARX model,

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})e(k) \quad (6)$$

where  $A(q^{-1})$ ,  $B(q^{-1})$ , and  $C(q^{-1})$  are the matrix polynomials

$$A(q^{-1}) = I + A_1 q^{-1} + \dots + A_{N_A} q^{-N_A} \quad (7)$$

$$B(q^{-1}) = B_1 q^{-1} + \dots + B_{N_B} q^{-N_B} \quad (8)$$

$$C(q^{-1}) = I + C_1 q^{-1} + \dots + C_{N_C} q^{-N_C} \quad (9)$$

The output error (OE) structure is a special case of the ARMAX model structure where  $C(q^{-1}) = A(q^{-1})$

$$y(k) = A^{-1}(q^{-1})B(q^{-1})u(k) + e(k) \quad (10)$$

The ARMAX structure provides a flexible description of the noise process with few parameters; i.e., the ARMAX structure is parsimonious. Because neither eq 6 nor eq 10 can be written as a linear regression problem, nonlinear optimization methods must be used to estimate the parameters. In the unlikely event that the noise model is known,  $H(q^{-1})$  could filter the residuals, and linear regression could be used to estimate model parameters.<sup>3</sup> Several estimation methods have been proposed for ARMAX models including pseudolinear regression; correlation methods (e.g., instrumental variables); and more recently, subspace methods.

The subspace methods identify a state-space model

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Gw(k) \\ y(k) &= Cx(k) + v(k) \end{aligned} \quad (11)$$

where  $x$  is an  $(n_x \times 1)$  vector of state variables,  $w$  is the  $(n_x \times 1)$  vector of state noise variables, and  $v$  is the  $(n_y \times 1)$  vector of measurement errors. The state and measurement noise vectors are assumed to be uncorrelated with one another and distributed as normal random variables with covariance matrices  $Q$  and  $R$ , respectively. Kailath<sup>7</sup> and Ljung<sup>3</sup> show that every state-space model is parametrically equivalent to an ARMAX model in eq 6, and vice versa. The state-space model can be identified using nonlinear optimization and a prediction error criterion, i.e., by using a prediction error method (PEM). However, the structure of the state-space matrices can be difficult to specify, and estimation of the matrices requires that a nonconvex optimization problem be solved. For MIMO models in particular, the computational burden is large, and there is no guarantee of a global minimum, i.e., the model identified using the PEM might not converge to the "true" model even if the correct model order is specified.

The choice between parametric and nonparametric methods and the selection of the model order and model structure (for both the process and noise models) both depend on the end use of the model. For example, if the properties of the noise model are important, a parametric model is required. More practically, several model

predictive control software packages use nonparametric models,<sup>9</sup> whereas many modern control and monitoring applications have been developed for state-space models.

After the model structure has been specified, the model parameters are estimated. For multivariable models, the estimation method must be robust to a large degree of correlation among the measured variables. The next sections describe two classes of estimation methods that are robust to correlated, measured variables. In the next section, multivariable statistical regression methods are described. Then, subspace algorithms for identifying state-space models are considered.

### Multivariable Regression

For a linear relation between predicted variables  $y$  and regressor variables  $x$ , the  $i$ th sample can be represented as

$$y(i) = B^T x(i) + e(i) \quad (12)$$

where  $y$  is dimension  $n_y$ ,  $x$  is dimension  $n_x$ , and the noise  $e(i)$  is assumed to be uncorrelated and normally distributed with zero mean and constant covariance matrix  $\Sigma$ . Additionally,  $x$  and  $y$  are both assumed to be mean-centered. Equation 12 can be written for  $N$  measurements of  $y$  and  $x$  as

$$Y = XB + E \quad (13)$$

where the rows of matrices  $Y$  (dimension  $N \times n_y$ ),  $X$  (dimension  $N \times n_x$ ), and  $E$  contain the individual observations of  $y^T$ ,  $x^T$ , and  $e^T$ . The ordinary least-squares (OLS) estimate of  $B$  is

$$\hat{B}_{OLS} = (X^T X)^{-1} X^T Y \quad (14)$$

For ordinary least squares, the noise covariance is  $\Sigma = I$ . Generalized least squares (GLS)<sup>10</sup> is appropriate if  $\Sigma \neq I$ . The GLS solution is equivalent to scaling both  $x$  and  $y$  by  $\Sigma^{-1/2}$ ; thus,  $\Sigma$  must be known and invertible.

The OLS solution can be inaccurate if  $X^T X$  is ill-conditioned. If there are collinearities within the matrix  $X$ , then the inverse of  $X^T X$  does not exist. If there are large correlations among the  $X$  variables, then the inverse might be poorly conditioned, resulting in an inaccurate estimate of  $\hat{B}_{OLS}$ . Additionally, the OLS solution ignores any correlations among the predicted variables,  $Y$ , treating each variable independently. A number of methods have been proposed for multivariable regression that are appropriate for high correlation in both the regressor and response matrices: principal component regression (PCR), ridge regression, "curds & whey", canonical correlation regression (CCR), and partial least squares (PLS). Note that there are two common PLS methods: one that treats each output independently (PLS1) and one that incorporates correlations among outputs (PLS2). This research focuses on PLS2, so herein, the term PLS refers to the PLS2 algorithm). Several excellent review articles analyze the various properties of these regression methods.<sup>11-13</sup> The two methods of interest in this research are PLS and CCR because they are the most commonly used techniques in the control engineering literature.

Beginning in the late 1980s, process control researchers investigated the use of chemometric methods for

modeling industrial processes. Chemometric methods such as principal component analysis (PCA), PCR, and PLS were developed to handle the large amounts of data that were stored as part of the "information age". These methods identify the model coefficient matrices by forming reduced-rank matrices of the regressors and predictors. That is, the correlations and collinearities within the input or output matrices are used to estimate  $B$  more accurately than the ordinary least-squares (OLS) solution.

Canonical correlation regression is a similar technique for reduced-rank regression, but it is less frequently mentioned in the process control and chemometrics communities. Canonical correlation regression is described by Tso<sup>14</sup> and has recently been examined by Stoica and Viberg.<sup>15</sup> Canonical correlation regression was derived as the maximum-likelihood estimate for the rank-deficient regression problem (i.e., when the  $X$  and  $Y$  matrices might not be full-rank).

An important point sometimes overlooked in applications of multivariate statistics is that several methods are available to handle an ill-conditioned regressor matrix, a well-known problem in the numerical literature. The ill-conditioned problem can be solved using methods based on a QR or SVD decomposition of the  $X^T X$  matrix.<sup>3,17</sup> The multivariate statistical approaches are similar to the numerical methods computationally, but they offer the advantages of a statistical interpretation that allows, e.g., hypothesis testing. Also, CCR and PLS incorporate correlations among the outputs, as well as among the regressors, in the solution. Ljung<sup>1</sup> suggested that the multivariate statistics methods might be an appropriate choice if there are measurement errors in both inputs and outputs, the "errors-in-variables" problem.

The significant features of the PLS and CCR methods are presented below. The complete algorithms are included in Appendices A and B. Before each method is discussed individually, a basis for comparing the two is presented in the next section.

**Regression on Orthogonal Components.** Estimating  $B$  in eq 13 by either PLS or CCR can be thought of as regression on orthogonal components.<sup>18,19</sup> Define the orthogonal components  $F$  as the projection of the regressor matrix,  $X$ , on a matrix  $O$

$$F = XO \quad (15)$$

The column rank of  $F$ ,  $n_F$ , will be less than or equal to the column rank of  $X$ ,  $n_X$ . For "reduced-rank regression",  $n_F < n_X$ ; i.e., a linear combination of the regressors is used to predict the outputs. Usually,  $O$  is defined such that the columns of  $F$  are orthogonal. The dimensions of  $O$  are thus  $n_X \times n_F$ .

The regression problem in eq 13 with  $F$  (rather than  $X$ ) as the regressor matrix is

$$Y = FB_F + E \quad (16)$$

The coefficient matrix in terms of  $F$ ,  $B_F$ , is related to the original coefficient matrix,  $B$ , by

$$Y = XOB_F + E \quad (17)$$

$$B = OB_F \quad (18)$$

Therefore,  $B$  can be estimated by first estimating  $B_F$  and then multiplying it by  $O$

$$\hat{B}_F = (F^T F)^{-1} F^T Y \quad (19)$$

$$\hat{B} = O \hat{B}_F = O(F^T F)^{-1} F^T Y$$

Each regression method defines  $O$  differently. In principal component regression (PCR),  $O$  is made up of the eigenvectors corresponding to the significant eigenvalues of the matrix  $X^T X$ . As will be discussed below, PLS and CCR are eigenvalue problems with different weightings on  $X^T X$ .

**Partial Least Squares (PLS).** Partial least squares was developed by Herman Wold<sup>18</sup> as a multivariate regression method suitable for econometrics applications where correlated and noisy inputs and outputs are common. Several reviews of PLS are available, including tutorials<sup>20</sup> and theoretical analyses.<sup>11,18,21</sup> Several chemometrics applications, e.g., predicting species concentrations from IR spectra, have used PLS.<sup>22</sup> Process applications of PLS has been reported by several authors.<sup>23–25</sup> Ricker,<sup>26</sup> MacGregor et al.,<sup>27</sup> and Qin<sup>28</sup> used PLS to estimate FIR models. Kaspar and Ray<sup>29</sup> used PLS to identify linear models with a dynamic inner relationship, a different method than the regression approach taken here. More recently, Dayal and MacGregor<sup>16</sup> compared the use of PLS and CCR for identifying FIR models.

Although there are different numerical procedures for PLS, each algorithm solves the same set of cost functions. The traditional NIPALS algorithm is given in Appendix A. The NIPALS algorithm is iterative; the steady-state solution of the NIPALS algorithm can be written as a set of eigenvalue problems,<sup>18</sup> including the following equation

$$X^T Y Y^T X = O \Lambda O^T \quad (20)$$

where  $\Lambda$  is a diagonal matrix of real, positive eigenvalues, ordered from largest to smallest. The dimension of the regression is reduced by selecting only those columns of  $O$  that correspond to the significant eigenvalues. The eigenvalue problem in eq 20 can be interpreted as calculating the eigenvalues of  $X^T X$  weighted by  $Y Y^T$ . Intuitively, this approach weights the projection matrix toward large values of  $Y$ . More formally, PLS calculates a projection matrix based on maximization of the covariance between  $X$  and  $Y$ .

**Canonical Correlation Regression (CCR).** Tso<sup>14</sup> introduced canonical correlation regression in the statistics literature, as the maximum-likelihood solution to reduced-rank multivariable regression. Compared with the large number of applications using PLS, CCR is used less frequently within the chemometrics and statistics communities, perhaps because it is not as well-known. The CCR solution has recently resurfaced in the engineering literature, where Stoica and Viberg<sup>15</sup> analyzed the numerical and asymptotic properties of maximum-likelihood parameter estimation in reduced-rank multivariable linear regression. The algorithm and derivation for CCR are given in Appendix B.

When viewed as a regression on orthogonal components, CCR defines the matrix  $O$  by the eigenvalue problem

$$(X^T X)^{-1/2} X^T Y (Y^T Y)^{-1} Y^T X (X^T X)^{-1/2} = O D O^T \quad (21)$$

with the constraint

$$O^T X^T X O = I \quad (22)$$

or

$$F^T F = I \quad (23)$$

Equation 21 defines one set of canonical variates;<sup>30</sup> the name “canonical correlation regression” arises from the canonical variates that define the orthogonal components used in the regression. The diagonal matrix  $D$  has the canonical correlation coefficients as its diagonal elements. The coefficients can be used to select the column rank of  $F$  by choosing columns of  $O$  that correspond to the statistically significant correlation coefficients. Equation 21 is not used directly in the CCR calculation but is useful for understanding and analyzing CCR.

The similarity between eqs 21 and 20 is striking. PLS weights elements of  $X$  by the values of  $Y$ ; CCR normalizes this weighting by  $(X^T X)^{-1/2}$  and  $(Y^T Y)^{-1/2}$ . In PLS applications, it is common practice to normalize each variable by its standard deviation. This is equivalent to multiplying  $x$  by the diagonal elements in  $(X^T X)^{-1/2}$  and  $y$  by the diagonal elements of  $(Y^T Y)^{-1/2}$ . Thus, if the off-diagonal elements are included in data normalization, the PLS and CCR solutions should be theoretically equivalent, i.e., the only differences will be due to the sample estimation properties. Because of its weighting, CCR is independent of scaling operations and is appropriate for an unknown and general noise covariance  $\Sigma$ . PLS will remain sensitive to scaling operations and the noise matrix, particularly when there are off-diagonal elements in  $\Sigma$ .

In the context of dynamic model identification, these details regarding scaling and the noise covariance matrix are less significant because of the presence of a noise model; i.e., the choice of model structure is more significant than the parameter estimation method that is used.

## Subspace Methods

Subspace identification methods are a recent development in the system identification field. The canonical variate analysis (CVA) algorithm was proposed by Larimore<sup>19,31–33</sup> and is based on the time series analysis methods developed by Akaike.<sup>34,35</sup> The N4SID algorithm was proposed by van Overschee and De Moor<sup>36,37</sup> and is more closely related to linear systems theory in the engineering literature. Both algorithms identify a stochastic state-space model of the form

$$x(k+1) = Ax(k) + Bu(k) + w(k) \quad (24)$$

$$y(k) = Cx(k) + Du(k) + v(k)$$

where  $x$  is the vector of state variables ( $n_x \times 1$ ),  $u$  is the vector of measured inputs ( $n_u \times 1$ ), and  $y$  is the vector of measured outputs ( $n_y \times 1$ ). In general, the state noise  $w$  and the measurement noise  $v$  are correlated if a minimal-order state-space model is identified. For subspace algorithms, the state vector has a very particular meaning that will be described below. As mentioned earlier, any linear model structure (e.g., ARX, ARMAX, OE) can be written as a state-space model.

The derivations of subspace algorithms are rather complicated compared with the derivations of traditional

prediction error methods. Furthermore, the different approaches used in the derivations make comparing subspace algorithms a challenging task. The derivation of CVA is cast in a mathematical statistics framework. The derivation of N4SID uses geometric arguments and a system theoretic approach, similar to realization theory.<sup>7</sup>

For subspace algorithms, the state vector  $x(k)$  is defined to be a linear combination of the past inputs and outputs

$$x(k) = Jp(k) \tag{25}$$

where

$$p(k) = [u^T(k-1) \ \dots \ u^T(k-N) \ y^T(k-1) \ \dots \ y^T(k-N)]^T \tag{26}$$

and  $p(k)$  is referred to as the “past” for sample  $k$ . The dimension of the past is the number of lags  $N$ . The state vector  $x(k)$  is calculated from data and is not specified to be a physical state of the system. After  $J$  has been calculated (as will be described below), the state vector can be estimated by eq 25. The state-space model matrices can then be estimated via linear regression

$$\begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix} = \text{Cov} \left( \begin{bmatrix} x(k+1) \\ y(k) \end{bmatrix}, \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \right) \times \text{Cov}^{-1} \left( \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}, \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \right) \tag{27}$$

The parameter estimation step of subspace algorithms can vary, but all algorithms proceed in the same general fashion: estimate the state vector from the past, and then estimate the state-space matrices using “current” values for state, input, and output vectors.

The calculation of  $J$  distinguishes the various subspace algorithms from one another. In the N4SID approach,  $J$  results from a series of geometric arguments (or linear algebraic arguments) based on a set of matrix equations for the evolution of a linear system.<sup>36</sup> The language and derivation for CVA are different, but the final set of equations is similar.

For CVA,  $J$  represents the canonical loadings<sup>30</sup> that relate the past,  $p(k)$ , and the “conditional future”, denoted by  $\tilde{f}$ . The canonical future is given by

$$\tilde{f}(k) = f(k) - Mq(k) \tag{28}$$

where  $f(k)$  and  $q(k)$  are vectors of “future” outputs and inputs

$$f(k) \triangleq [y^T(k) \ \dots \ y^T(k+N-1)]^T \tag{29}$$

$$q(k) \triangleq [u^T(k) \ \dots \ u^T(k+N-1)]^T \tag{30}$$

The conditional future in eq 28 represents the future measurements with the effect of future inputs removed. For state-space models,  $M$  is a block Toeplitz matrix with the FIR coefficient matrices along the diagonals

$$M = \begin{bmatrix} D & 0 & \dots & \dots \\ CB & D & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ CA^{N-1}B & CA^{N-2}B & \dots & D \end{bmatrix} \tag{31}$$

The canonical loadings that define  $J$  for CVA are calculated by a singular value decomposition (SVD). For all subspace methods, the calculation of  $J$  can be written as a weighted SVD<sup>38</sup>

$$\text{svd}(W_1 \mathcal{O} W_2) = [U_1 \ U_2] \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \tag{32}$$

where  $W_1$  and  $W_2$  are weighting matrices and  $\mathcal{O} = \Sigma_{p,\tilde{f}}$ . In CVA,  $\mathcal{O}$  is known as the covariance matrix of  $p$  and  $\tilde{f}$ . In N4SID terminology,  $\mathcal{O}$  is the oblique projection of the future outputs along future inputs on the past inputs and outputs. For either method,  $J$  is the lower-dimensional subspace defined by  $U_1$ . In the N4SID algorithm,  $W_1 = W_2 = I$ . For the CVA algorithm,  $W_1 = \Sigma_{p,p}^{-1/2}$  and  $W_2 = \Sigma_{\tilde{f},\tilde{f}}^{-1/2}$ , so that  $J$  is the matrix of canonical loadings corresponding to  $p(k)$ , the past inputs and outputs.

The  $W_1$  and  $W_2$  scaling matrices for CVA are particularly noteworthy because they result from the maximum-likelihood (ML) solution of a constrained rank regression problem for multivariate regression,<sup>19</sup> exactly as in CCR. When applied to time series data, a ML procedure incorporates the correct delay structure of the regression coefficients (i.e., the precise structure of the ARMAX polynomials). In CVA, this unknown shift structure is not explicitly imposed, but several simulation studies demonstrate the accuracy of CVA to be essentially equal to that of ML.<sup>5</sup> A formal proof for this phenomenon has yet to be obtained. ML accuracy permits the computation of confidence bands on model accuracy and precise tests of hypotheses concerning the model structure, e.g., detection of the presence of bias or trends, feedback, and delays. Near-ML accuracy in quite small sample sizes has also been demonstrated for a number of complex systems by Larimore.<sup>5</sup>

In principle, different subspace algorithms could identify equivalent state-space models if scaled appropriately. However, other important differences exist among the subspace algorithms: the estimates of  $\mathcal{O}$  using a finite amount of sampled data and the numerical procedure for estimating the state-space matrices, determining the dimension of the past ( $N$  in eq 26), and selecting the model order. Larimore<sup>5</sup> has shown that the CVA weighting can identify models that are nearly equivalent to maximum-likelihood (ML) methods, which require that the exact model structure be known a priori.

Ideally, the model order corresponds to the singular values in eq 32 that are greater than zero or some very small value  $\epsilon$ . For N4SID, the order is selected by the user, usually by looking for a “knee” in the plot of singular values vs model order or by excluding the singular values that fall below a specified critical value. Often, there is more than one knee or no knee at all. For CVA, the model order is selected via Akaike’s information criterion (AIC), a statistical method for model selection. The singular values (or equivalently, the canonical correlations) can also be used for order selection and hypothesis testing.

Subspace algorithms provide a simple and fast method for identifying a MIMO state-space model. The calculations are noniterative and computationally efficient. Thus, a rich model structure can be identified without the need for optimization or an expert in system identification to specify the structure of the state-space matrices and the model order.

## Discussion of Identification Techniques

The multivariable regression techniques are numerical methods for identifying models that can be written in the form of eq 13. That is, no matter which regression method is used, the identified model is an ARX or FIR model. These structures are useful, but they are sensitive to the noise process. For example, if there is "output error noise" present, the ARX coefficients will be biased from their true values. The ARX model compensates for this bias by using a larger-order model. Wahlberg<sup>6</sup> has described the statistical properties of high-order ARX models.

The subspace techniques are not sensitive to the noise process. Because the resulting state-space model compactly describes the noise and system, it can be easier to interpret than high-order FIR or ARX models. This issue is particularly important if the model is used for designing a control or fault-detection algorithm, because a more parsimonious model simplifies the algorithm.

The calculation for  $J$  in eq 25 is the major difference between the various subspace algorithms. Similar to the difference between CCR and PLS, the difference between solutions can be interpreted as different weightings in a cost function (eq 32). The cost function for subspace models is less intuitive than those defining PLS and CCR; however, CVA, like CCR, is scale-invariant and thus appropriate for general noise covariance matrices. By contrast, the N4SID algorithm is sensitive to scale changes. Additionally, Juricek et al.<sup>41</sup> have demonstrated that including the CVA weighting in the N4SID algorithm does not produce a model equivalent to that identified by the ADAPT<sub>X</sub> CVA procedure.

## Simulation Examples

Three simulation examples provide a comparison of the various system identification techniques. The first example is a static regression problem that illustrates the differences between CCR and PLS. The second example examines identified models of a continuous stirred-tank reactor (CSTR) for different disturbances. The identified models are ARX and FIR models estimated by CCR and PLS, subspace models estimated by CVA and N4SID, and a state-space model estimated by PEM. The same models are used in the third example, the quality-control problem of Dayal and MacGregor.<sup>16</sup> The quality-control example contains collinear outputs and time delays.

**Multivariate Regression Example.** Consider a coefficient matrix,  $B$  in eq 13, given by

$$B_0 = \begin{bmatrix} 5\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -5\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \quad (33)$$

Because the rank of  $B_0 = 1$  (the two variables in  $Y$  are collinear), the reduced-rank techniques should be able to identify  $B_0$  with one orthogonal component. The noise,  $e(i)$  in eq 12, has covariance  $\Sigma$ . Two different values of  $\Sigma$  are considered:

$$\begin{array}{ll} \text{Case 1} & \Sigma = \begin{bmatrix} 10 & 0 \\ 0 & 0.4 \end{bmatrix} \\ \text{Case 2} & \Sigma = \begin{bmatrix} 10 & 1.99 \\ 1.99 & 0.4 \end{bmatrix} \end{array}$$

**Table 1. Kullback Information Results for the Regression Example**

Method	Case	Scaling	$\bar{I}$	No. Rejections
PLS	1	auto	1.69	6
CCR	1	—	1.68	6
PLS	2	auto	2.41	11
PLS	2	$\Sigma^{-1/2}$	1.60	5
CCR	2	—	1.55	5

Note that matrix  $\Sigma$  is usually unknown in practical applications; for this analysis, it is used to demonstrate the significance of scaling.

The CCR and PLS methods were used to estimate  $B_0$  for 100 trials. For PLS and case 1, each output was scaled to unit variance, and input vector  $x$  was a normally distributed random vector with mean  $\mu_x = 0$  and covariance matrix  $\Sigma_{x,x} = I$ . For both CCR and PLS, one latent variable was used (the obvious selection based on AIC or Bartlett's test for CCR<sup>30</sup> and cross-validation for PLS). For case 2, two different scaling matrices for PLS were used: one scaling the outputs to unit variance and the other scaling the outputs by  $\Sigma^{-1/2}$ . According to the discussion following eq 23, the latter scaling should produce asymptotically equivalent results for PLS and CCR.

Different noise sequences and inputs were used for 100 trials, with each trial consisting of 1000 samples. The Kullback information<sup>42</sup> was computed for each trial. It measures the distance between the probability distribution of the estimated model,  $p(z|\theta_1)$ , and the probability distribution of the true system,  $p(z|\theta_0)$ , for  $z = (y_1, \dots, y_N)$  and a sample of size  $N$

$$I(\theta_1, \theta_0) = \int p(z|\theta_0) \log \frac{p(z|\theta_1)}{p(z|\theta_0)} dz \quad (34)$$

where  $\theta_1$  denotes the estimated parameters of a proposed model and  $\theta_0$  denotes the true parameters for the true model. For this multivariable regression example, the proposed and true models have the same structure (i.e., each model is a linear relationship between the same regressors and outputs). The Kullback information for a multivariate regression where the estimated coefficient matrix is  $\theta_1 = B_1$  and true coefficient matrix is  $\theta_0 = B_0$  is given by<sup>42</sup>

$$I(B_1, B_0) = \text{trace}[\Sigma^{-1}(B_1 - B_0)^T X^T X (B_1 - B_0)] \quad (35)$$

Theoretically,  $I(B_1, B_0)$  follows a  $\chi_k^2$  distribution with  $k$  degrees of freedom, where  $k$  is the number of independently estimated parameters. The expected value of  $\chi_k^2$ , and thus  $I(B_1, B_0)$ , equals  $k$ . For the current example,  $k = 2$ .

The mean value for the Kullback information is shown in the fourth column of Table 1 for each regression method. The slightly smaller values for CCR in both cases indicate that, on average, the regression matrices estimated by CCR are closer to the true matrix  $B_0$ . As expected, the mean values for PLS are very close to those of CCR for case 1 and for case 2 when the outputs are scaled by  $\Sigma^{-1/2}$ . For each trial, the following hypothesis test was used to determine whether the estimated regression matrix was statistically different from the true value,  $B_0$ , for the observed values of  $X$

$$H_0: B_1 = B_0 \quad H_1: B_1 \neq B_0$$

**Table 2. Nominal Values for the CSTR Simulations**

Variable	Value	Variable	Value
$F$	100 L/min	$E/R$	8750 K
$C_F$	1 mol/L	$k_0$	$7.2 \times 10^{10} \text{ min}^{-1}$
$T_F$	350 K	$UA$	$5 \times 10^4 \text{ J/(min K)}$
$V$	100 L	$T_J$	309.9 K
$\rho$	1 kg/L	$C_p$	239 J/(kg K)
$-\Delta H$	$5 \times 10^4 \text{ J/mol}$		

The rejections when  $I(B_1, B_0)$  was tested against the  $\chi^2_2(\alpha)$  limit for  $\alpha = 0.05$  are shown in the fifth column of Table 1. Note that, for case 2, using PLS with the usual scaling rejects the hypothesis test 11 times, indicating that this procedure can perform poorly.

The above example shows that, for multivariable regression problems, the PLS estimate of  $B_0$  is close to the CCR estimate of  $B_0$  when the noise matrix is either known or diagonal. Otherwise, CCR will produce significantly better results. The next two examples compare the dynamic models from PLS and CCR with models estimated by the subspace techniques.

**CSTR Example.** A continuous stirred-tank reactor (CSTR) with a first-order reaction  $A \rightarrow B$  is described by<sup>43</sup>

$$\frac{dC_A}{dt} = \frac{F}{V}(C_0 - C_A) - k_0 \exp\left(-\frac{E_A}{RT}\right)C_A \quad (36)$$

$$\frac{dT}{dt} = \frac{F}{V}(T_0 - T) + \frac{-\Delta H}{\rho C_p} k_0 \exp\left(-\frac{E_A}{RT}\right)C_A + \frac{UA}{\rho C_p V}(T_J - T) \quad (37)$$

The concentration of A,  $C_A$ , and the temperature,  $T$ , are the measured (output) variables, and the manipulated (input) variables are the jacket temperature,  $T_J$ , and the flow rate,  $F$ . Nominal values for the CSTR parameters are given in Table 2. The equations were linearized, discretized, and put into the state-space form

$$\begin{aligned} x(k+1) &= \Phi x(k) + Gu(k) \\ y(k) &= Hx(k) + v(k) \end{aligned} \quad (38)$$

where

$$y(k) = [C_A \ T]^T \quad (39)$$

$$u(k) = [F \ T_J]^T \quad (40)$$

and the disturbance,  $v(k)$ , is a stochastic process. The model variables  $y$  and  $u$  are expressed as deviations from the nominal conditions in Table 2. The equivalent transfer function representation and matrix fraction decomposition<sup>7</sup> were used to define the appropriate noise model, i.e.

$$\begin{aligned} y(k) &= P(q^{-1})u(k) + v(k) \\ P(q^{-1}) &= H(qI - \Phi)^{-1}G = A_0^{-1}(q^{-1})B_0(q^{-1}) \end{aligned} \quad (41)$$

where  $A_0(q^{-1})$  and  $B_0(q^{-1})$  are the corresponding matrix polynomials as in eqs 7 and 8. Processes with three different noise models were examined: (a) the OE model  $v(k) = v_0(k)$ , where  $v_0(k)$  is normally distributed with mean zero and covariance matrix  $R$ ; (b) the ARX model  $v(k) = A_0^{-1}(q^{-1})w_0(k)$ , where  $w_0(k)$  is normally distrib-

**Table 3. Covariance Matrices for the CSTR Simulations**

$$\begin{aligned} R &= \begin{bmatrix} 1 \times 10^{-6} & 0 \\ 0 & 4 \times 10^{-2} \end{bmatrix} \\ Q_1 &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \times 10^{-2} \end{bmatrix} \\ Q_2 &= \begin{bmatrix} 1 \times 10^{-6} & 0 \\ 0 & 1 \times 10^{-2} \end{bmatrix} \\ Q_{\text{ARX}} &= \begin{bmatrix} 1 \times 10^{-5} & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

uted with mean zero and covariance matrix  $Q$ ; and (c) the ARMAX model  $v(k) = v_0(k) + A_0^{-1}(q^{-1})C(q^{-1})w_0(k)$ , where  $v_0(k)$  and  $w_0(k)$  are uncorrelated, normally distributed sequences with covariance matrices  $R$  and  $Q$ , respectively. Furthermore, two matrices for  $Q$  were examined,  $Q_1$  and  $Q_2$ .

The elements of  $R$  were selected to yield a signal-to-noise ratio (SNR) of approximately 9 for the OE simulation. The covariance matrices are shown in Table 3. In each case, 50 trials with different noise seeds were generated. For each trial, 1000 samples were generated; 500 samples were used for identification, the other 500 for validation. The inputs were pseudo-random binary signals (PRBSs)<sup>3</sup> with a clock period of  $\Delta = 10$  samples. The PRBS amplitudes were  $\delta T_J = 1 \text{ K}$  and  $\delta F = 0.6 \text{ L/min}$ . The sampling period was 3 s.

A fourth simulation examined identification during closed-loop operation. For the closed-loop simulations, white noise was added to each output with variances  $\sigma_{C_A}^2 = 10^{-6}$  and  $\sigma_T^2 = 0.0411$ . Even though white noise was added, the OE model is incorrect because the true noise process includes the feedback path. That is, if the transfer function for the controller is denoted by  $C(q^{-1})$ , then  $H(q^{-1})$  in eq 1 is given by

$$H(q^{-1}) = [I + P(q^{-1})C(q^{-1})]^{-1} \quad (42)$$

The identification data were generated by exciting each set point as a PRBS with a clock period of 10 samples. The PRBS set-point amplitudes were  $C_{A,sp} = 0.01 \text{ mol/L}$  and  $T_{sp} = 1 \text{ K}$ .

For each simulation the following models were identified: (a) FIR models with 30 lags identified by PLS using the Matlab PLS Toolbox<sup>44</sup> and by CCR using the algorithm developed by Tso<sup>14</sup> (these models will be referred to as the PLS-FIR and CCR-FIR models, respectively); (b) ARX models identified by PLS and CCR, with the number of lags for the ARX model determined by the AIC (these models will be referred to as the PLS-ARX and CCR-ARX models, respectively); (c) subspace models identified by N4SID using the MATLAB System Identification Toolbox<sup>45</sup> and CVA using ADAPT<sub>X</sub><sup>46</sup>; (d) state-space models identified by PEM using the MATLAB System Identification Toolbox, version 4.0.5, with the true model order correctly specified, the PEM models identified using the "modstruc" and "pem" commands, and the structures of the state-space coefficient matrices were not specified; and (e) ARX models (using the MATLAB System Identification Toolbox) with the true model orders, i.e., the orders of  $A(q^{-1})$  and  $B(q^{-1})$  set equal to  $A_0(q^{-1})$  and  $B_0(q^{-1})$  (these models will be referred to as the ARX models). In analyzing the identification results, it is important

**Table 4. CSTR Results for an OE Model**

Method	$R^2(y_1)$ (%)	$R^2(y_2)$ (%)
PLS-FIR	83.7	85.1
PLS-ARX	89.8	90.1
CCR-FIR	83.4	85.2
CCR-ARX	88.9	89.3
N4SID	88.9	88.7
CVA	90.1	90.3
ARX	85.8	81.1
PEM	90.6	90.8

**Table 5. CSTR Results for an ARX Noise Process**

Method	$R^2(y_1)$ (%)	$R^2(y_2)$ (%)
PLS-FIR	(-) <sup>a</sup>	(-)
PLS-ARX	98.9	97.5
CCR-FIR	(-)	(-)
CCR-ARX	98.9	97.5
N4SID	98.2	97.3
CVA	98.9	97.5
ARX	98.9	97.5
PEM	98.9	97.5

<sup>a</sup>  $R^2$  values less than zero are denoted by (-).

**Table 6. CSTR Results for an ARMAX Process**

Method	$R^2(y_1)$ (%)	$R^2(y_2)$ (%)
PLS-FIR	40.8	42.1
PLS-ARX	92.5	88.9
CCR-FIR	43.0	43.5
CCR-ARX	92.6	89.0
N4SID	90.6	87.8
CVA	92.8	89.3
ARX	91.5	85.6
PEM	93.0	89.6

to distinguish between the effect of the model structure and the effect of the parameter estimation method.

The standard metric for comparing different models is the normalized sum of squared errors, or  $SSE_N$ , defined by

$$SSE_N = \frac{\sum_i [y(i) - \hat{y}(i)]^2}{\sum_i [y(i) - \bar{y}]^2} \quad (43)$$

which is often reported as the coefficient of determination  $R^2$

$$R^2 = (1 - SSE_N) \times 100\% \quad (44)$$

Because the true process was known in the simulation study, the difference between the identified and true frequency responses is a more meaningful criterion than the  $R^2$  values for comparing the different models. This assertion is made because the frequency response shows the accuracy of the identification method for all frequencies whereas the coefficient of determination provides a single measure of how well the identified model fits the data.

Tables 4–7 show the  $R^2$  values of each identified model for the ARX, OE, ARMAX and closed-loop simulations. The tables and figures present results for the validation data rather than for the identification data. The  $R^2$  values provide a simple metric for discriminating between bad and good models. For example, the results for the ARX and ARMAX cases in Tables 5 and 6, respectively, indicate that the FIR models are quite poor ( $R^2 < 0$  in Table 5!) and that the other models are all of about the same accuracy; for the ARX case in Table 5, the ARX, CCR-ARX, PLS-ARX, CVA, and PEM models have identical  $R^2$  values to three significant figures. For

**Table 7. CSTR Results for Closed-loop Identification**

Method	$R^2(y_1)$ (%)	$R^2(y_2)$ (%)
PLS-FIR	72.7	84.6
PLS-ARX	86.4	89.6
CCR-FIR	84.0	88.0
CCR-ARX	86.3	89.6
N4SID	81.4	85.7
CVA	87.1	90.3
ARX	81.9	82.2
PEM	87.3	90.4

the OE and closed-loop cases in Tables 4 and 7, the  $R^2$  values are very similar. For example,  $R^2(y_2)$  in Table 4 varies from 81.1 for the ARX model to 90.8 for the PEM model. The differences among the PLS-ARX, CCR-ARX, N4SID, CVA, and N4SID  $R^2$  values are small, although, for each case, the PEM and CVA models have the largest  $R^2$  values.

A more complete description of the model accuracies is provided by the frequency response magnitudes. Figures 1–3 show the frequency response magnitude of each identified model for the OE, ARX, and ARMAX processes, respectively. The magnitude plots for the  $C_A - F$  transfer function,  $g_{11}$ , are shown. Although only  $|g_{11}|$  is shown, the significant observations are essentially the same if the magnitudes for the other transfer functions or the phase angles of the transfer functions are considered, so that the use of  $|g_{11}|$  does not favor a particular identification method.

The solid line in each figure is the true frequency response magnitude; the dashed line represents the mean for the 50 trials. The dotted lines are the 95% confidence limits for the mean (calculated from the sample standard deviations for the 50 trials). The different examples demonstrate the effect of a noise process on the identification method. The major differences among the identified models are mostly explained by the model *structure*, not the identification *method*.

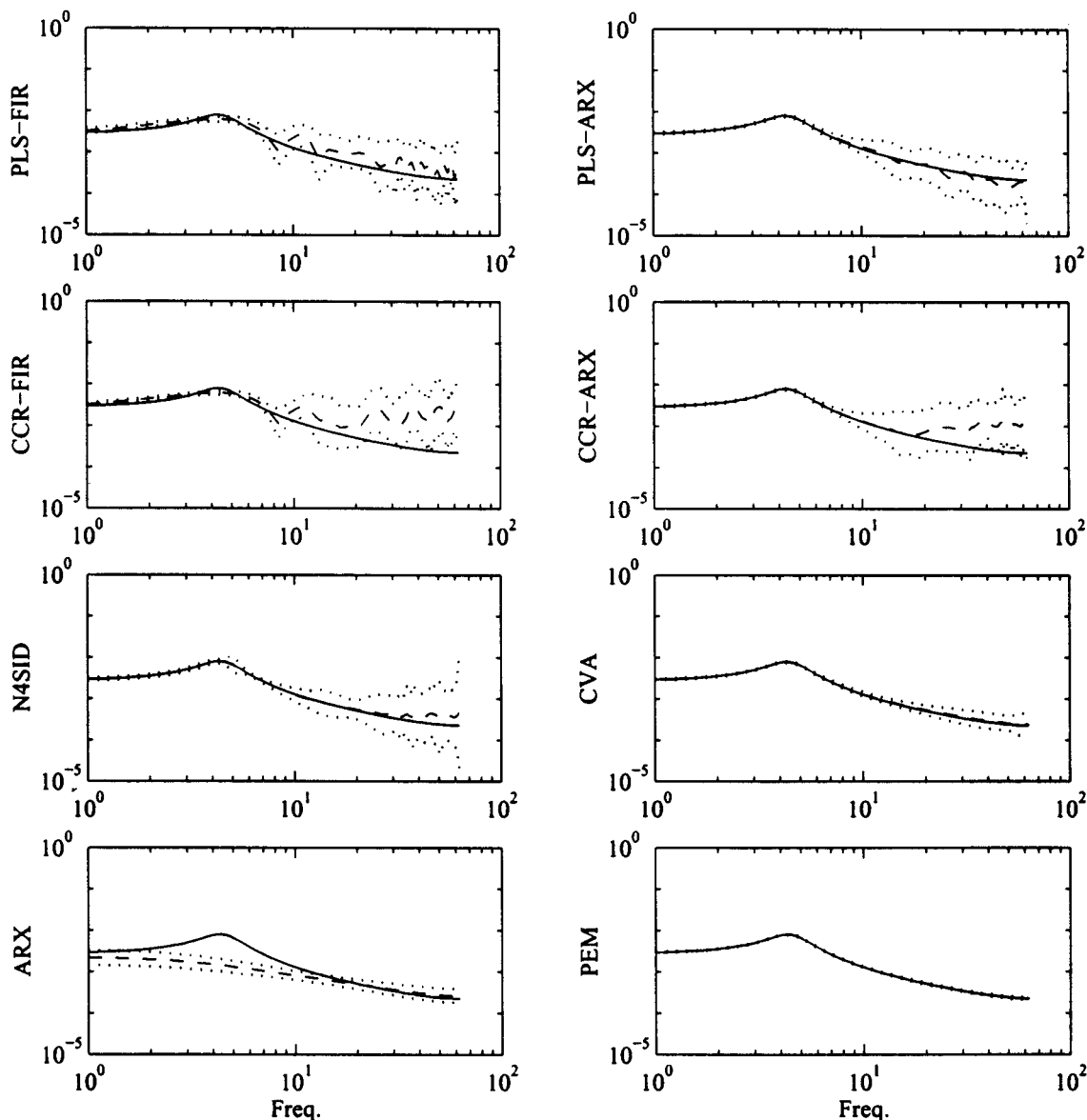
For the OE case in Figure 1, several methods perform poorly, especially at higher frequencies. This might be the result of three factors:

(a) *The frequencies excited by the PRBS signal.* The clock period of the PRBS signal is 10 samples; hence, the change in model accuracy at  $\omega_0 = 6.28 \text{ min}^{-1}$  (i.e.,  $\omega_0 = \pi/\Delta\delta t$ , where the clock period,  $\Delta = 10$  samples and the sampling period is  $\delta t = 0.05 \text{ min}$ ). Based on the power spectral density of a PRBS signal,<sup>47</sup> most of the power occurs for  $\omega < \omega_0$ ; i.e., for  $\omega > \omega_0$ , the signal-to-noise ratio is poor. Thus, for frequencies higher than  $\omega_0$ , models will be accurate only if the system and noise model structures are correct.

(b) *The frequencies that are weighted by the model structure.* Approximating an OE process by a high-order ARX model asymptotically (i.e., with an infinite amount of data) yields the true model (see section 10.4 in ref 3). For finite data, however, this approach weights the model error differently at different frequencies, leading to a “waterbed effect”—improved accuracy at some frequencies is achieved at the cost of less accuracy at other frequencies.

(c) *The frequencies that are most heavily weighted by the regression method.* The frequency weighting of PLS or CCR is not well understood; the importance of this issue will be discussed below.

The ARX model (with the true model orders) is biased at almost all frequencies, demonstrating its sensitivity to the noise model and motivating the use of a high-order ARX, FIR, or subspace model.



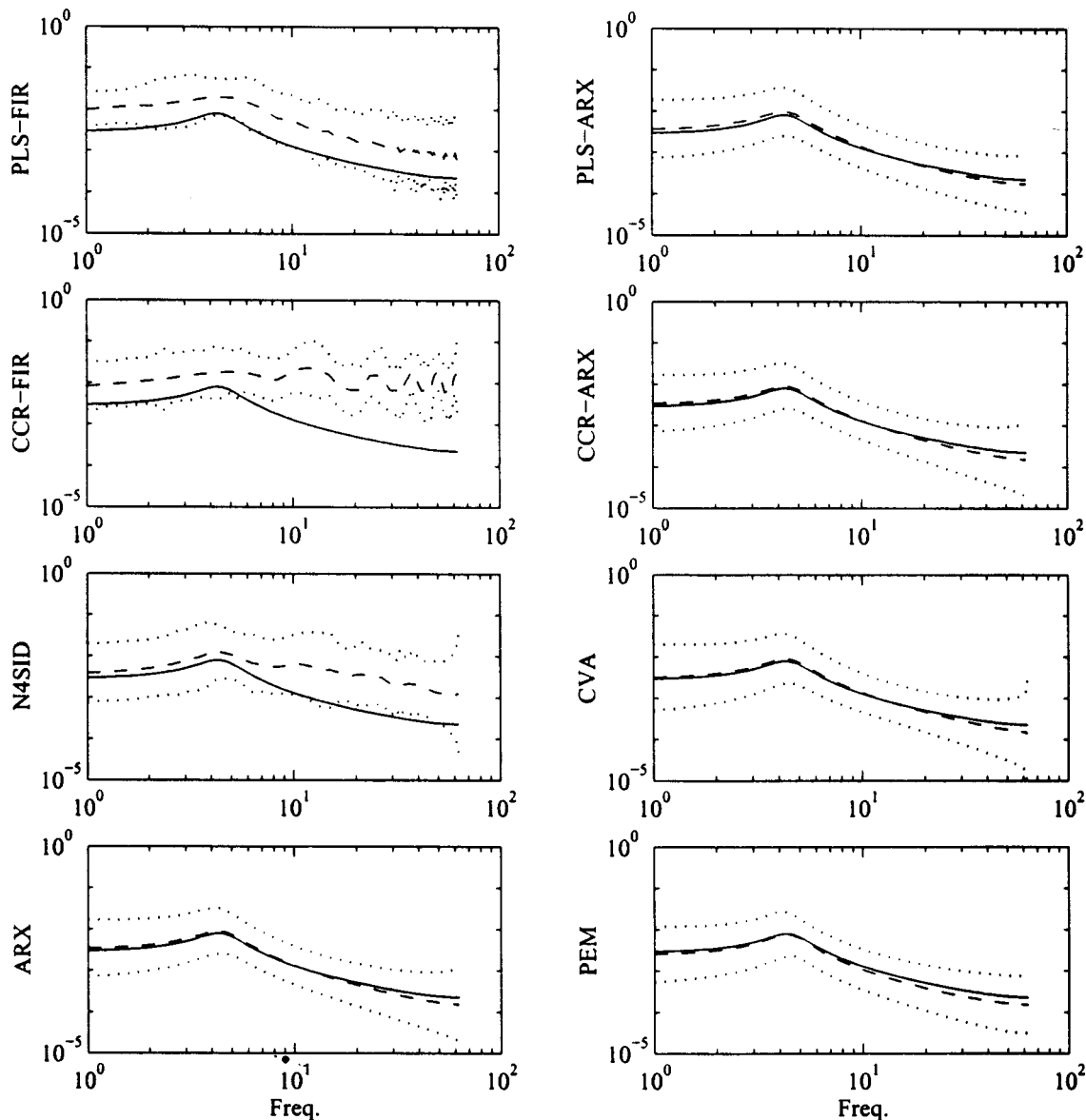
**Figure 1.** Frequency response magnitudes for  $v(k) = v_0(k)$ ,  $v_0(k) \in N(0, R)$ ; an OE model describes the noise process (— process, --- estimated, ... 95% confidence limits).

The CCR-FIR and PLS-FIR models are similar, being fairly accurate up to  $\omega_0$  and poor for  $\omega > \omega_0$ . The CCR-ARX and PLS-ARX models are very accurate for  $\omega < \omega_0$ , and the mean value of the PLS-ARX model follows the true process for  $\omega > \omega_0$ . Somewhat surprisingly, the CCR-FIR and CCR-ARX models are quite poor (measured by the bias in the mean value and the wide confidence limits) at the higher frequencies. Similar simulation studies<sup>16,48</sup> found that CCR-FIR models were more accurate than PLS-FIR models when the PRBS clock period was  $\Delta = 1$ , so that the input power was larger at high frequencies. The results of this study show the opposite behavior when the signal-to-noise ratio is small at high frequencies. Dayal and MacGregor<sup>16</sup> argued that the smaller singular values of the covariance matrix contain the high-frequency information and that PLS ignores these singular values. For the current investigation, when the clock period is  $\Delta = 10$ , there is less information about the system at frequencies larger than  $\omega_0$ . By ignoring these high frequencies, the PLS-ARX and PLS-FIR models do not attempt to fit the noise.

The subspace models, and CVA in particular, produce accurate models (unbiased with small confidence limits) for all frequencies, demonstrating the advantage of the state-space structure. The PEM model (generated using the OE command in MATLAB) is the most accurate.

The results for the ARX noise process are shown in Figure 2. The results are poor for the PLS-FIR, CCR-FIR, and N4SID models. The PLS-ARX, CCR-ARX, and ARX models are equivalent because the model structures are correct, and all latent vectors are used by PLS and CCR; i.e., the solution from the reduced-rank methods converges to the ordinary least-squares solution. The CVA and PEM models are also quite accurate.

Because an OE model can be considered to be a special case of an ARMAX model,  $C(q^{-1}) = A(q^{-1})$  in eq 6, the results for the ARMAX case in Figure 3 are similar to those for the OE case in Figure 1. There is very little difference among the ARX, subspace, and PEM methods for  $\omega < \omega_0$ . The FIR models are again fairly accurate for  $\omega < \omega_0$  and poor at the higher frequencies, although the FIR models are slightly better



**Figure 2.** Frequency response magnitudes for  $v(k) = A(q^{-1})w_0(k)$ ,  $w_0(k) \in N(0, Q_{ARX})$ ; an ARX model describes the noise process. (— process, - - - estimated, ... 95% confidence limits).

for the ARMAX case than for the OE case. The magnitudes for the ARX model have improved relative to the OE case because of the addition of process noise, but they are still poor. Once again, the mean magnitudes are biased at high frequencies for the PLS-ARX, CCR-ARX, and N4SID models, although there is a range,  $\omega_0 < \omega < \omega^*$ , where the mean frequency response magnitudes are close to the true value. The CVA and PEM models are accurate for all frequencies.

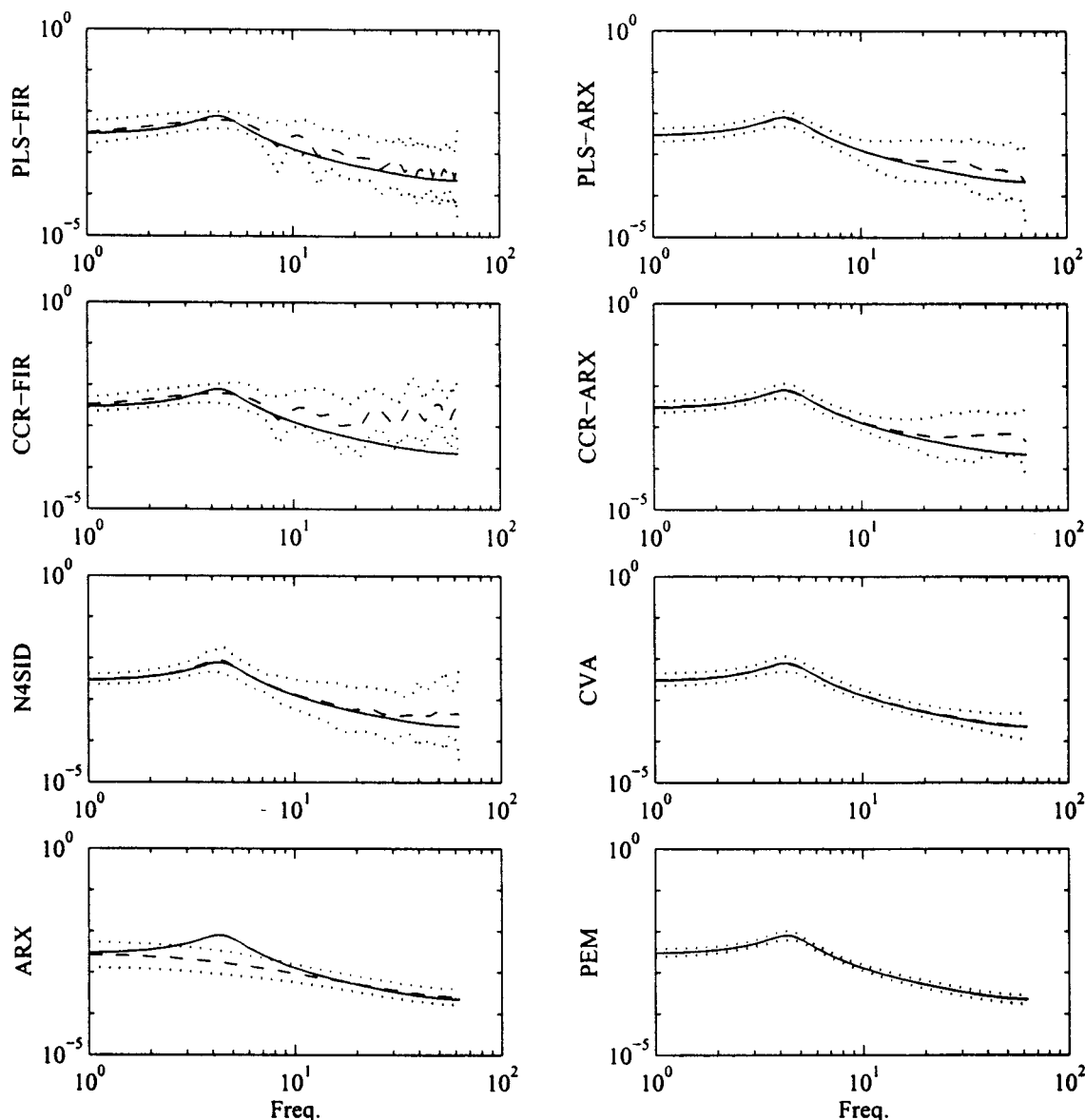
By increasing the process noise variance, the robustness of each method to process noise can be demonstrated. Figure 4 shows the plots for the larger process noise covariance matrix,  $Q_2$ . The results for the PLS-ARX, CCR-ARX, N4SID, CVA, and PEM models are comparable for  $\omega < \omega_0$ , but the CVA and PEM models are more accurate for  $\omega > \omega_0$ , even though the signal-to-noise ratio is small. The FIR models are especially poor.

The  $R^2$  results for the closed-loop identification are given in Table 7, and the magnitude plots are shown in Figure 5. The ARX models are clearly biased for the

entire range of frequencies. The CCR-FIR and PLS-FIR models are poor, particularly for  $\omega > \omega_0$ ; the CCR-ARX models also are biased for  $\omega > \omega_0$  but are very accurate at low frequencies. The PLS-ARX, N4SID, CVA, and PEM models are unbiased for all frequencies; the CVA and PEM models are especially accurate.

It is interesting to compare the  $R^2$  values with the frequency responses, because they do not necessarily agree. For the OE process, the PLS-ARX, CVA, and PEM models all have similar  $R^2$  values. However, the frequency response magnitude plots in Figure 1 show that the CVA and PEM methods produce models that are very accurate over all frequencies and that the PLS-ARX models are inaccurate at higher frequencies. Although  $R^2$  is a useful metric that describes how well the model fits data and does not require knowledge of the true process, it is not necessarily related to how well the model represents the actual process at all frequencies.

*A Robust Control Metric.* The singular values of the error in the frequency response can be used as another



**Figure 3.** Frequency response magnitudes for  $v(k) = v_0(k) + C(q^{-1})w_0(k)$ ,  $v_0(k) \in N(0, R)$ ,  $w_0(k) \in N(0, Q_1)$ ; an ARMAX model describes the noise process (— process, --- estimated, ... 95% confidence limits).

metric for analyzing the identified models.<sup>16</sup> The additive perturbation error is defined as

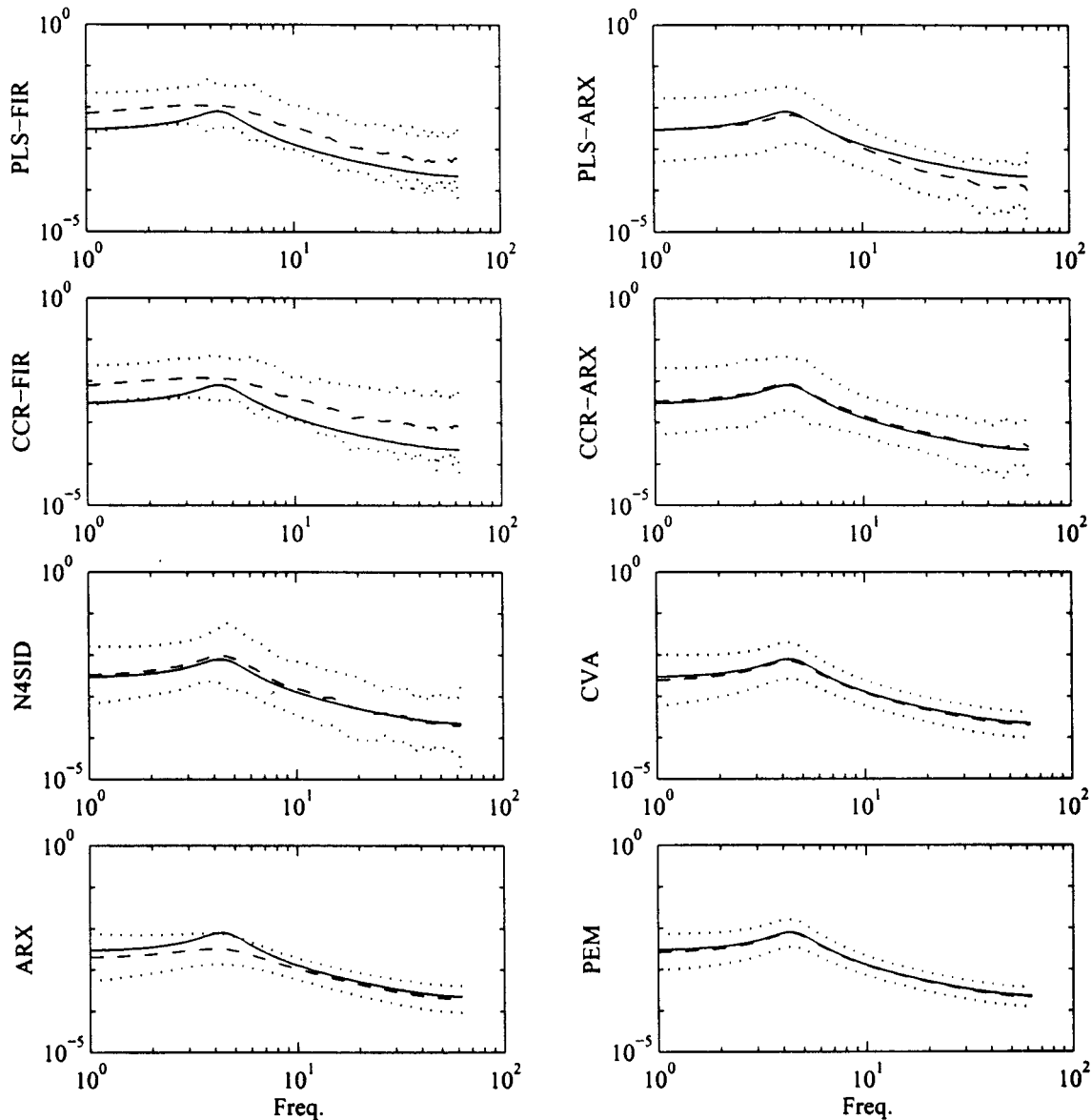
$$\Delta_a(\omega) = P(e^{-i\omega}) - \hat{P}(e^{-i\omega}) \quad (45)$$

where  $P(e^{-i\omega})$  is the true process frequency response and  $\hat{P}(e^{-i\omega})$  is the frequency response of the identified model. As described by Zhou et al.,<sup>49</sup> the achievable controller performance is limited by a norm of the perturbation,  $\|\Delta_a(\omega)\|$ . For the design of robust controllers,  $\|\Delta_a(\omega)\|$  can be specified using process knowledge or experimental data, or it can be calculated using the techniques proposed by Smith<sup>50</sup> and Smith and Poola.<sup>51</sup> It is important to note that the true perturbation is not given by eq 45, because the effect of noise is ignored. Nevertheless, eq 45 can be used to compare the accuracy of the various identification methods. In this research, the mean value of  $\|\Delta_a(\omega)\|$ , the maximum singular value of  $\Delta_a$  (the  $H^\infty$  norm), for the Monte Carlo trials is calculated to compare the different identification methods.

The mean values of  $\|\Delta_a\|$  for 50 trials are shown in Figures 6–8 for the PLS-FIR, CCR-FIR, PLS-ARX,

CCR-ARX, and subspace models from the OE case.  $\|\Delta_a(\omega)\|$  for the PEM model is shown in each figure for comparison. The plots support several of the observations made using the frequency response magnitude plots: For  $\omega > \omega_0$ , the CCR models are worse than the PLS models. The FIR models are quite poor across all frequencies, even though the FIR results in Figure 1 are fairly accurate at lower frequencies. This anomaly might occur because the frequency response magnitude plots consider only the (1,1) element whereas  $\|\Delta_a(\omega)\|$  is sensitive to all of the transfer function errors. The PEM and CVA methods result in the smallest values of  $\|\Delta_a\|$ , as indicated in Figure 8.

*Discussion.* A comparison of the subspace and PEM models puts the subspace methods, and CVA in particular, in perspective. The performances of the N4SID, CVA, and PEM methods are generally very good, and the PEM models that have the correct model orders are always the most accurate (in terms of either the  $R^2$  values, the frequency response magnitudes, or  $\|\Delta_a(\omega)\|$ ). Therefore, one might ask why select a subspace method rather than a PEM to identify a state-space model?



**Figure 4.** Frequency response magnitudes for  $v(k) = v_0(k) + C(q^{-1})w_0(k)$ ,  $v_0(k) \in N(0, R)$ ,  $w_0(k) \in N(0, Q_2)$ ; an ARMAX model describes the noise process (— process, --- estimated, ... 95% confidence limits).

Theoretically, there is no guarantee that the PEM method will converge to a unique solution because it relies on nonlinear optimization for a solution, unlike the subspace methods. This issue is particularly important for larger-dimensional problems where the nonlinear optimization needs to optimize a large number of parameters. Also, selecting the model order for a PEM model can be challenging. If several model orders are examined, each PEM model requires a new optimization; by contrast, the subspace methods are noniterative, and hence, the models are identified much faster. Defining the model structure of a PEM model (i.e., the model order and the state-space matrix structures) can also be challenging. For subspace methods, only the model order is specified. Furthermore, metrics such as AIC can be used to select the model order automatically, and plots of the singular values of  $\mathcal{O}$  in eq 32 aid the user for a manual model order selection. The closed-loop simulations show that CVA also performs very well for closed-loop identification. Thus, subspace methods are robust to different noise models, are not affected by

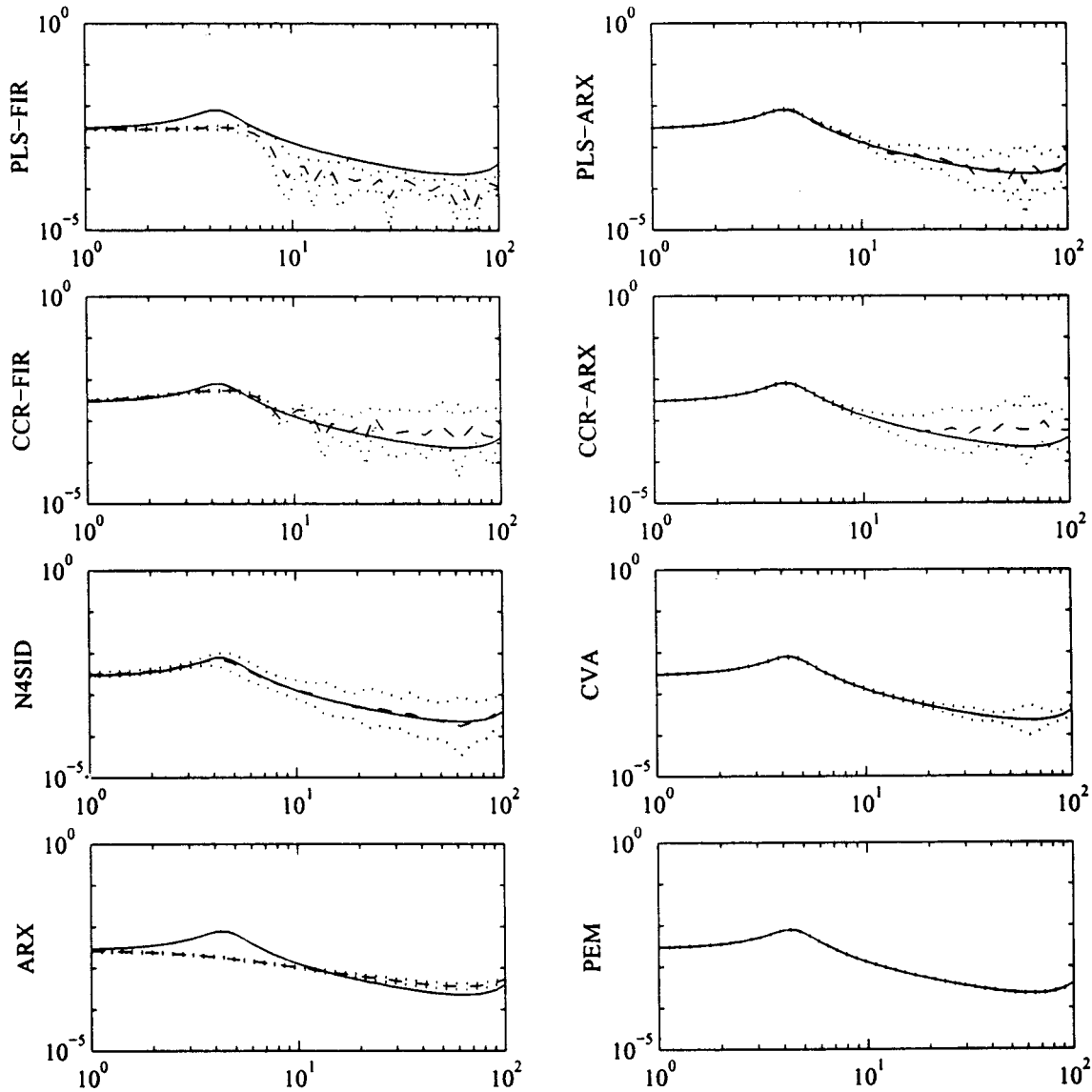
feedback, do not require a priori knowledge of the model structure, and identify models that are nearly as accurate as the PEM model while being much simpler to compute.

**Quality-Control Example.** The process examined in this section is the quality-control example from Dayal and MacGregor<sup>16</sup>

$$x(k) = \frac{0.2q^{-5}}{1 - 0.8q^{-1}} \left( \begin{bmatrix} 1 & 0.7 \\ 0.7 & -1 \end{bmatrix} u(k) + \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} w(k) \right) \quad (46)$$

$$y(k) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} x(k) + v(k)$$

Because of the collinearity among outputs, a reduced-rank method that is robust to output collinearities is



**Figure 5.** Frequency response magnitudes for the closed-loop simulations. (— process, --- estimated, ... 95% confidence limits).

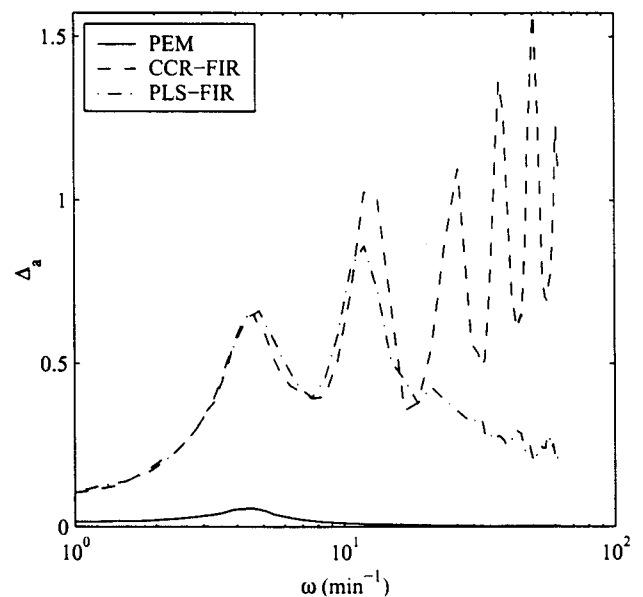
necessary. Two simulations were performed:

Case 1  $w(k) = 0$ , or an output error process

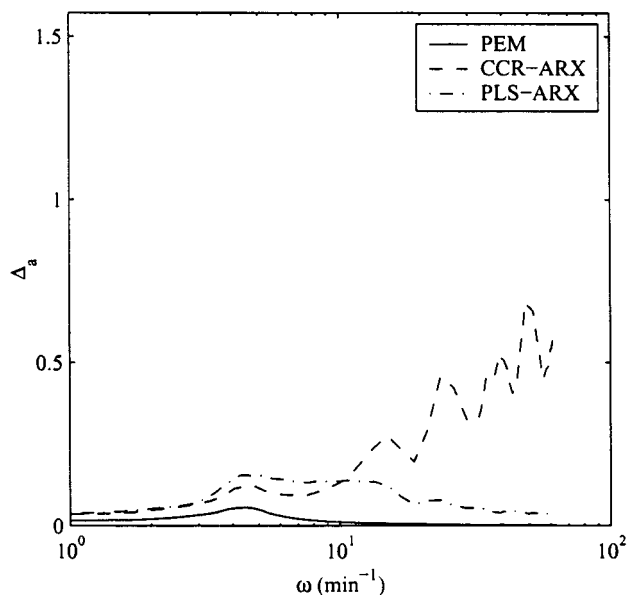
Case 2  $w(k) \in N(0, Q)$ , or an ARMAX process

The measurement noise,  $v(k)$ , was uncorrelated with the process noise,  $w(k)$ , and was normally distributed with covariance matrix  $R$  chosen to provide a signal-to-noise ratio of approximately 3 for each output. For each case, 100 Monte Carlo trials with different noise seeds were generated. For each trial, 1000 samples were generated; 500 samples were used for identification, and the other 500 samples for validation. The excitation data were generated by changing each input simultaneously and independently as a PRBS with clock period  $\Delta = 10$ .

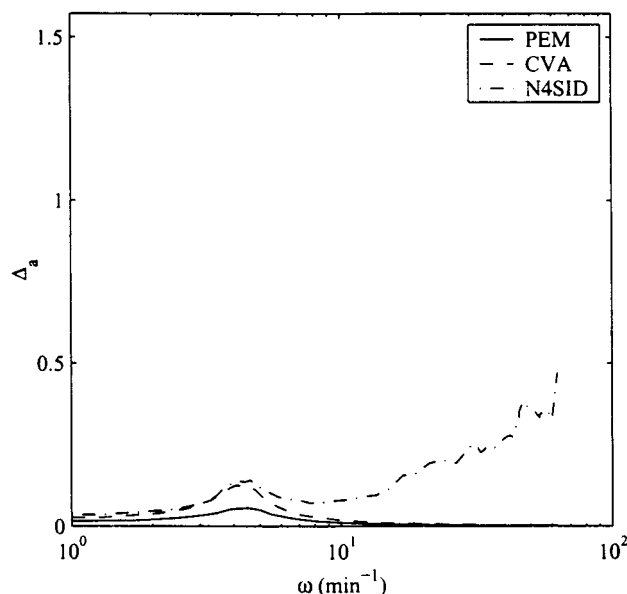
The following model structures were identified: (a) FIR models with 30 lags identified by CCR and PLS, (b) ARX models identified by CCR and PLS (model order determined by AIC), (c) subspace models identified by N4SID and CVA, and (d) a second set of subspace models that incorporated a priori knowledge of the four sample time delay (these models are referred to as CVA-2 and N4SID-2).



**Figure 6.** Singular values of the additive perturbation error,  $\Delta_a$ , for the estimated FIR and PEM models.



**Figure 7.** Singular values of the additive perturbation error,  $\Delta_a$ , for the estimated ARX and PEM models.



**Figure 8.** Singular values of the additive perturbation error,  $\Delta_a$ , for the subspace and PEM models.

A preliminary investigation of PEM models indicated that the results were extremely sensitive to the model structure specification (i.e., the free parameters to be estimated, as specified in the "modstruc" command). The "brute-force" approach used in the CSTR simulations (i.e., using a state-space model and not specifying a structure for the state-space matrices) produced extremely inaccurate results, even when the correct time delay was specified a priori. It is likely that a more careful approach (e.g., explicitly modeling the collinearities among the outputs a priori) could produce accurate PEM models. However, it would be difficult to make meaningful comparisons between "black-box" FIR, ARX, and subspace models, and a PEM method that requires precise a priori knowledge. Hence, PEM models were not considered for this simulation study.

**Results for Quality-Control Example.** The coefficient of determination,  $R^2$  (based on one-step-ahead prediction for each output), and the numbers of parameters

**Table 8. Quality-Control Example,  $Q = 0$**

Method	$R^2(y_1)$ (%)	$R^2(y_2)$ (%)	$R^2(y_3)$ (%)	$R^2(y_4)$ (%)	$R^2(y_5)$ (%)	Parameters <sup>a</sup>
PLS-FIR	70.0	72.1	71.6	70.2	72.1	300
PLS-ARX	69.0	71.3	70.4	68.9	71.2	350
CCR-FIR	70.4	72.9	72.3	70.6	72.9	300
CCR-ARX	69.8	72.1	71.6	69.9	72.3	350
N4SID	57.2	60.0	63.5	59.5	61.9	51
N4SID-2	68.7	67.8	67.9	67.7	66.8	87
CVA	68.4	71.4	71.6	69.1	71.8	123
CVA-2	71.4	73.9	72.8	71.6	73.8	39

<sup>a</sup> For the FIR models, the number of parameters is  $n_y \times n_{lag} \times n_u$ , where  $n_{lag} = 30$ . For the ARX models, the number of parameters is  $n_y \times n_{lag} \times (n_y + n_u)$ , where  $n_{lag} = 10$ . For the state-space models, the number of parameters is determined by<sup>5</sup>  $M = \bar{n}_x(2n_y + n_u) + 1/2\bar{n}_x(n_y + 1)$ , where  $\bar{n}_x$  is the average state vector dimension for all 100 Monte Carlo trials.

**Table 9. Quality-Control Example,  $Q \neq 0$**

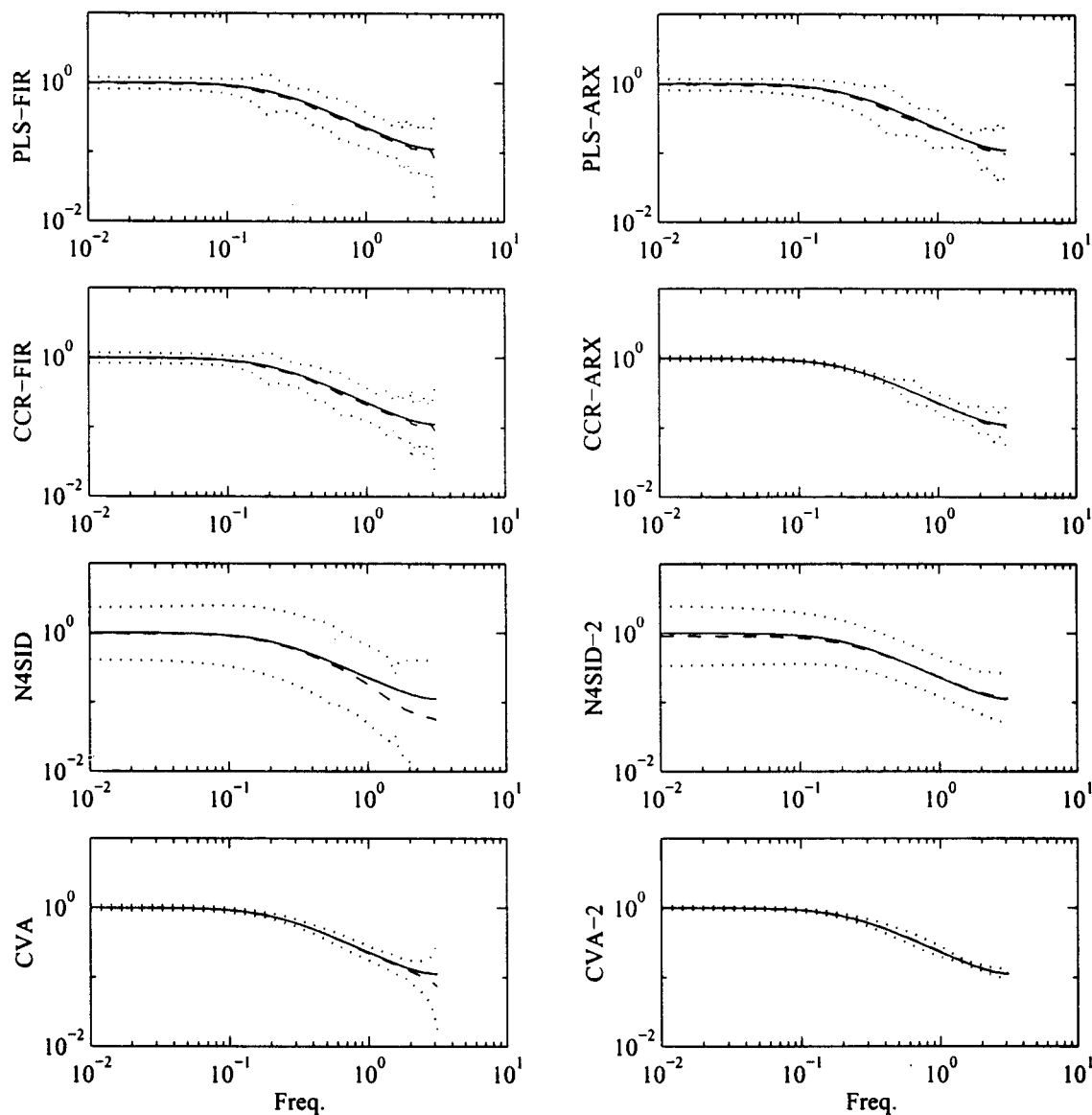
Method	$R^2(y_1)$ (%)	$R^2(y_2)$ (%)	$R^2(y_3)$ (%)	$R^2(y_4)$ (%)	$R^2(y_5)$ (%)	Parameters <sup>a</sup>
PLS-FIR	40.9	41.9	35.1	38.1	38.9	300
PLS-ARX	61.4	63.9	63.9	62.1	64.2	280
CCR-FIR	41.6	42.9	36.2	39.0	39.9	300
CCR-ARX	62.6	64.8	64.1	63.0	64.8	280
N4SID	51.7	54.6	56.7	53.4	55.7	111
N4SID-2	62.4	65.3	64.0	62.9	65.1	99
CVA	51.3	52.6	52.0	51.6	52.3	75
CVA-2	65.0	67.5	66.4	65.4	67.2	39

<sup>a</sup> For the FIR models, the number of parameters is  $n_y \times n_{lag} \times n_u$ , where  $n_{lag} = 30$ . For the ARX models, the number of parameters is  $n_y \times n_{lag} \times (n_y + n_u)$ , where  $n_{lag} = 8$ . For the state-space models, the number of parameters is determined by<sup>5</sup>  $M = \bar{n}_x(2n_y + n_u) + 1/2\bar{n}_x(n_y + 1)$ , where  $\bar{n}_x$  is the average state vector dimension for all 100 Monte Carlo trials.

estimated for each model are shown in Tables 8 and 9. For the OE case ( $Q = 0$ ) in Table 8, the  $R^2$  values are approximately equal for all but the N4SID and N4SID-2 models; the results for N4SID and N4SID-2 models are poorer. For the ARMAX case ( $Q \neq 0$ ) in Table 9, the  $R^2$  results are poor for the PLS-FIR and CCR-FIR models and, somewhat surprisingly, also for the CVA and N4SID models. The results for CCR-ARX and PLS-ARX are better, and the CVA-2 and N4SID-2 models, which assume that the time delay is known, are the most accurate. The mean values and 95% confidence limits for the frequency response magnitudes are shown in Figures 9 and 10 for  $g_{11} = y_1/u_1$ .

The results for the OE case ( $Q = 0$ ) in Figure 9 indicate that the N4SID models have quite large confidence limits. Plots of each Monte Carlo trial indicated that most frequency response magnitudes from the identified models appeared to be normally distributed. However, a number of the trials (approximately 10 out of 100 for both N4SID and N4SID-2) clearly were outliers. The occurrence of these outliers is likely due to the linearly correlated outputs. Among the methods considered, only the N4SID algorithm is not designed to handle collinear outputs.<sup>36</sup> The state orders for those N4SID and N4SID-2 models corresponding to the apparent outlier trials (selected by Akaike's FPE in MATLAB) were always much larger than the state dimension of the true process. Interestingly, the CVA models in Figure 9 do not appear to be greatly affected by either the time delay or the correlated outputs.

However, when state noise was added to the simulation ( $Q \neq 0$ ), the resulting CVA models in Figure 10 were significantly less accurate than the other methods;



**Figure 9.** Frequency response magnitudes for  $Q = 0$  for the quality-control example.

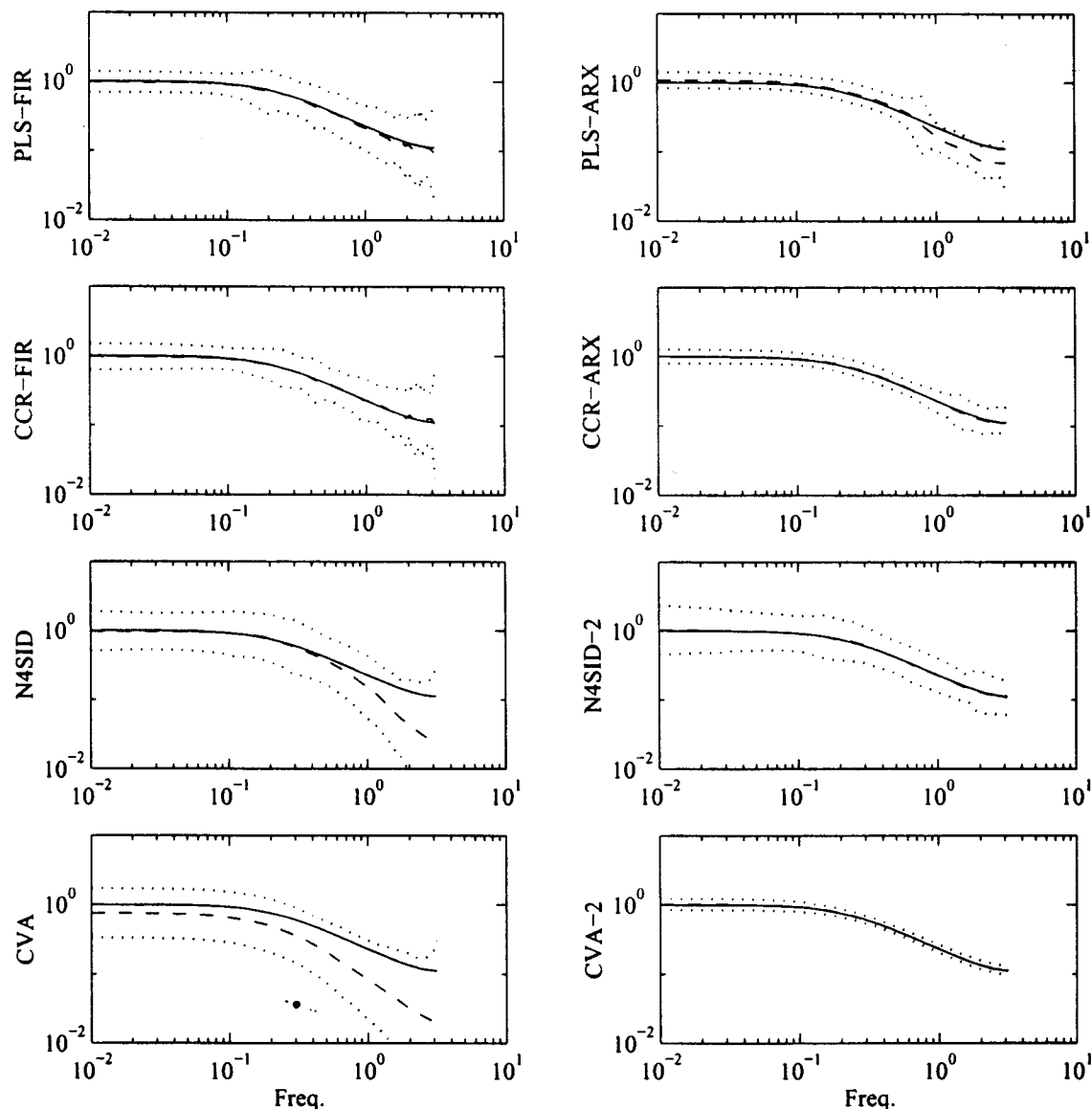
even the N4SID model was more accurate. The large confidence limits for the magnitudes in Figure 10 reflect a large variability in the identified models. Unlike the N4SID models for the  $Q = 0$  case, the large variability of the CVA magnitudes was not due to outliers. Knowledge of the time delay significantly improves the accuracy, as shown by the tight 95% confidence limits for the CVA-2 models.

Another anomaly for the results in both cases ( $Q = 0$  and  $Q \neq 0$ ) is the rapid increase in the variance of the mean magnitude at high frequencies for many of the methods. The frequency response magnitudes appeared to be normally distributed, but with a significantly larger variance. This might be due, in part, to the poor signal-to-noise ratio at these frequencies, but the rapid increase in the variance was unexpected.

During the latter part of this research, version 3.5 of the ADAPT<sub>X</sub> software was released.<sup>46</sup> It allows the time delay to be estimated as part of the system identification procedure. The time delay estimation procedure is discussed briefly in Appendix C. The CVA models based on the estimated time delays for each input are labeled CVA-delay. The CVA-delay results for 100 Monte Carlo trials are shown in Table 10. The 95% confidence limits

for the CVA-delay models are shown in Figure 11. The accuracy of the identified model improves significantly using the estimated time delays and approaches the accuracy of the CVA-2 models in Figures 9 and 10 where the time delays are assumed to be known.

Among the FIR and ARX models in Figures 9 and 10 (which are not adjusted for the time delay), the CCR-ARX models are particularly accurate. Interestingly, the FIR and ARX models tend to be more accurate than the subspace methods that do not account for the time delay. This result suggests that subspace models are particularly sensitive to the assumed time delay. When the time delay is unknown, subspace models compensate by adding additional state variables that require more parameters to be estimated in a very coordinated way. Ideally, the  $A$  matrix should be structured to account for the delay with the corresponding blocks of the  $B$  and  $C$  matrices in eq 24 set equal to zero. Thus, the FIR and ARX models "explicitly" model a time delay because one set of FIR or ARX coefficients equals zero. The state-space model "implicitly" models the delay because several parameters, block coefficient matrices within the state-space matrices, must all be identified accurately. Because of this difference in the model structures, the



**Figure 10.** Frequency response magnitudes for  $Q \neq 0$  for the quality-control example.

**Table 10. Distribution of the Estimated Time Delays for the 100 CVA-delay Models**

Estimated Delay	$Q \neq 0$		$Q = 0$	
	$u_1$	$u_2$	$u_1$	$u_2$
0	0	0	1	0
1	0	0	0	0
2	0	0	0	0
3	0	0	3	3
4	99	99	95	96
5	1	0	1	0
6	0	1	0	1

subspace methods will be more sensitive to time delays than the FIR or ARX models.

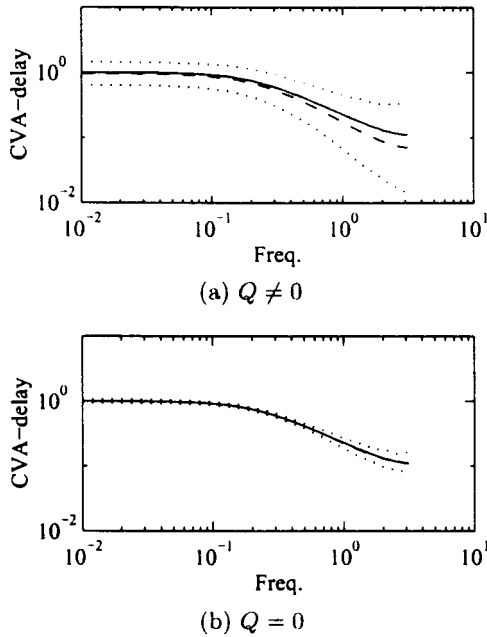
### Comparison with Previous Studies

The CSTR results in this study differ from the results published by Juricek et al.<sup>48</sup> Notably, the PLS models in Juricek et al.<sup>48</sup> appear to be biased for low frequencies. The noise processes for each study were slightly different and caused minor differences in the results. More importantly, the smaller PRBS clock period used in the earlier study significantly reduced the accuracy

of the PLS models. In the current study, the PRBS clock period is  $\Delta = 10$  samples. In Juricek et al.,<sup>48</sup> the PRBS clock period was  $\Delta = 2$  samples. Based on the power spectral density for a PRBS signal,<sup>47</sup> a smaller PRBS clock period decreases the signal power at low frequencies and increases the signal power at high frequencies.

As discussed in the CSTR example, a frequency domain interpretation of CCR and PLS is not rigorously understood. Dayal and MacGregor<sup>16</sup> assert that the smaller singular values of the covariance matrix contain the high-frequency information and that PLS ignores these singular values while CCR does not. That is, PLS tends to ignore the higher frequencies and model the low frequencies; CCR does the opposite. If, at low frequencies, the signal power is large, the PLS models benefit from this property as the PLS models do not "fit the noise". Conversely, the CCR models are more appropriate for signals that have large power at high frequencies.

These assertions are empirical observations based on simulation studies. Future research could provide a rigorous theoretical foundation to support these obser-



**Figure 11.** Frequency response magnitudes for CVA models incorporating the estimated delay for the quality-control example: (a)  $Q \neq 0$ , (b)  $Q = 0$ .

variations and to better understand the SNR–frequency effects on CCR and PLS.

### Conclusions

The results of extensive simulation studies demonstrate that the subspace identification methods, particularly CVA, are very adept at producing accurate, parsimonious models and are quite robust to different noise processes. However, the quality-control example shows that subspace methods are particularly sensitive to unknown time delays and that FIR and ARX models can be more accurate when time delays are not known or estimated. Although the accuracy of any identified model will improve if time delays are known and not estimated, time delay estimation methods, such as that provided by the ADAPT<sub>X</sub> software, are especially important for subspace methods. If the subspace methods incorporate an estimated time delay, the subspace methods are very accurate compared to the other model parametrizations.

It has been shown that the estimated regressor matrices and the objective functions for the CCR and PLS methods are asymptotically equivalent if scaled appropriately. For dynamic models, the scaling issue is further complicated by the presence of the noise process. For FIR or ARX model identification, these simulation results and those of previous studies<sup>16,48</sup> suggest that the choice between CCR or PLS could depend on the input excitation and the frequency range of interest (i.e., the signal-to-noise ratio for different frequencies), because CCR is sensitive to noise at high frequencies.

### Acknowledgment

An earlier version of this paper was presented at the 1998 IFAC DYCOPS Symposium.<sup>48</sup> Funding provided by the UCSB Process Control Consortium is gratefully appreciated.

### Appendix A: PLS Algorithm

1. Set  $u$  to first column of  $Y$
2.  $w = X^T u / u^T u$

3. Scale  $w$  to be length 1

4.  $t = Xw$

5.  $c = Y^T t / t^T t$

6. Scale  $c$  to be length 1

7.  $u = Y^T c / c^T c$

8. If convergence continue, else go to step 2

9.  $X$  loadings:  $p = X^T t / t^T t$

10.  $Y$  loadings:  $q = Y^T u / u^T u$

11. Regression ( $u$  upon  $t$ ):  $b = u^T t / t^T t$  "Inner Relation"

12. Residual matrices:  $X \rightarrow X - tp^T$ ,  $Y \rightarrow Y - btc^T$

The procedure is repeated using the new residual matrices. After the  $X$  and  $Y$  loadings have been calculated, the regression matrix between  $Y$  and  $X$  calculated by PLS is given by<sup>25</sup>

$$\hat{B}_{\text{PLS}} = W(P^T W)^{-1}(T^T T)^{-1}T^T Y \quad (47)$$

where  $W$ ,  $P$ , and  $T$  are the matrices with columns calculated from the above algorithm.

### Appendix B: CCR Algorithm

The CCR algorithm has been derived using several different approaches. Larimore<sup>32</sup> provides a geometric interpretation of the canonical variables in terms of a generalized singular value decomposition. The traditional derivation can be found in refs 14 and 15 and is based on maximizing the likelihood function (or the logarithm of the likelihood) for the multivariate regression equation in eq 13

$$2 \log L = k + N \log |\Sigma^{-1}| - \text{tr}(Y - XB)\Sigma^{-1}(Y - XB)^T \quad (48)$$

As explained in the Regression on Orthogonal Components section, factor  $B$  into the product of two matrices

$$B = OB_F \quad (49)$$

where  $O$  is full column rank and  $B_F$  is the regression matrix with  $F$  as the regressor matrix, where  $F = XO$ . Normalize  $O$  such that

$$F^T F = I_s \quad (50)$$

Decompose  $X$  and  $Y$  into the product of two matrices

$$X = US, \quad Y = VT \quad (51)$$

where  $V$  ( $N \times t$ ) and  $U$  ( $N \times r$ ) have orthonormal columns ( $U^T U = I_r$ ,  $V^T V = I_t$ ) and  $S$  and  $T$  are invertible. Define

$$R \triangleq U^T V \quad (52)$$

The singular values of  $R$  are the canonical correlations between  $X$  and  $Y$ . For the following eigenvalue problem

$$(I_r - RR^T)L = LA \quad (53)$$

the eigenvectors corresponding to the  $s$  smallest eigenvalues,  $L_s$ , define the subspace of  $O$ .<sup>14</sup> The normalization condition ( $F^T F = I$ ) is satisfied by

$$O = S^{-1}L_s \quad (54)$$

Having defined  $O$ , BF is calculated from

$$B_F = F^T Y = O^T X^T Y \quad (55)$$

### Appendix C: Time Delay Estimation Procedure in ADAPT<sub>X</sub>

To simplify the discussion, assume that the model is for a SISO process. The time delay estimation procedure is similar for the MISO and MIMO problems, but requires the specification of whether delays act on inputs or outputs for the hypothesis test. The delay estimation procedure begins by fitting an ARX model without incorporating any time delays

$$y(t) = \sum_{k=1}^p A_k y(t-k) + \sum_{k=1}^p B_k u(t-k) + e(t) \quad (56)$$

If the true delay is  $\delta$ , the corresponding ARX model would be given by

$$y(t) = \sum_{k=1}^p A_k y(t-k) + \sum_{k=\delta+1}^p B_k u(t-k) + e(t) \quad (57)$$

For each hypothesized delay  $d$  with  $1 \leq d \leq p$ , the  $B_k$  coefficients are set to zero for  $k \leq d$ , and the remaining nonzero ARX coefficients are estimated. The AIC statistic corresponding to the model with delay  $d$  is denoted  $AIC(d)$ .

For a time delay of  $\bar{d}$ , a hypothesis test can be used to decide whether  $d = \bar{d}$  or  $d = \bar{d} - 1$ , i.e., whether the corresponding ARX coefficient  $B_{\bar{d}-1}$  is significantly different from zero

$$\begin{aligned} H_0: d = \bar{d} \quad (B_k = 0, k \leq \bar{d}) \\ H_a: d = \bar{d} - 1 \quad (B_k = 0, k < \bar{d}) \end{aligned} \quad (58)$$

An optimal test of the hypothesis is given by the maximum-likelihood ratio statistic

$$\lambda = \frac{p(y|\hat{\theta}_0)}{p(y|\hat{\theta}_a)} \quad (59)$$

where  $p(y|\hat{\theta})$  is the likelihood for  $y$  with maximum-likelihood estimate for the model parameters  $\hat{\theta}$  under each hypothesis. For the hypothesis test in eq 58,  $\hat{\theta}_0$  denotes the estimated  $\theta$  when  $d = \bar{d}$  and  $\hat{\theta}_a$  denotes  $\theta$  when  $d = \bar{d} - 1$ . The likelihood ratio can be evaluated directly in terms of the AIC as<sup>52</sup>

$$\log \lambda = \frac{1}{2} [AIC(\bar{d}) - AIC(\bar{d}-1)] + 1 \quad (60)$$

Assuming that  $y$  is normally distributed, eq 60 is distributed as a noncentral  $\chi^2$  variable with one degree of freedom. Thus, for a proposed delay  $\bar{d}$ ,  $\alpha(\bar{d})$  denotes the probability of rejecting  $H_0$ , i.e., the probability of rejecting  $d = \bar{d}$  in favor of  $d = \bar{d} - 1$ . The estimated delay for the model is the smallest  $\bar{d}$  that is significant, i.e., when  $\alpha(\bar{d})$  is less than a specified value. An alternative approach to this multiple comparison prob-

lem (i.e., choosing the optimal delay from the multiple estimates) is to select the delay with the minimum AIC statistic.

### Literature Cited

- (1) Ljung, L. Issues in System Identification. *IEEE Control Syst. Mag.* **1991**, *11* (1), 25–29.
- (2) Box, G.; Jenkins, G.; Reinsel, G. *Time Series Analysis: Forecasting and Control*; Prentice Hall: Englewood Cliffs, NJ, 1994.
- (3) Ljung, L. *System Identification*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, 1999.
- (4) Söderström, T.; Stoica, P. *System Identification*; Prentice Hall: Hertfordshire, U.K., 1989.
- (5) Larimore, W. Statistical Optimality and Canonical Variate Analysis System Identification. *Signal Process.* **1996**, *52*, 131–144.
- (6) Wahlberg, B. Model Reductions of High-Order Estimated Models: The Asymptotic ML Approach. *Int. J. Control* **1989**, *49*, 169–192.
- (7) Kailath, T. *Linear Systems*; Prentice Hall: Englewood Cliffs, NJ, 1980.
- (8) Hannan, E.; Deistler, M. *The Statistical Theory of Linear Systems*; John Wiley: New York, 1988.
- (9) Qin, S.; Badgwell, T. An Overview of Industrial Model Predictive Control Technology. *AIChE Symp. Ser.* **1996**, *316*, 232–256.
- (10) Mardia, K.; Kent, J.; Bibby, J. *Multivariate Analysis*; Academic Press: New York, 1979.
- (11) Burnham, A.; Viveros, R.; MacGregor, J. Frameworks for Latent Variable Multivariate Regression. *J. Chemometrics* **1996**, *10*, 31–45.
- (12) Breiman, L.; Friedman, J. Predicting Multiple Responses in Multiple Linear Regression. *J. R. Stat. Soc. B* **1997**, *59*, 3–54.
- (13) Frank, I.; Friedman, J. A Statistical View of Some Chemometrics Regression Tools. *Technometrics* **1993**, *35*, 109–148.
- (14) Tso, M. Reduced-Rank Regression and Canonical Analysis. *J. R. Stat. Soc. B* **1981**, *43*, 183–189.
- (15) Stoica, P.; Viberg, M. Maximum Likelihood Parameter and Rank Estimation in Reduced-Rank Multivariate Linear Regressions. *IEEE Trans. Signal Process.* **1996**, *44*, 3069–3078.
- (16) Dayal, B.; MacGregor, J. Multi-output Process Identification. *J. Process Control* **1997**, *7*, 269–282.
- (17) Laub, A. *Computational Matrix Analysis: Course Notes for ECE 234*. University of California: Santa Barbara, CA, 1995.
- (18) Höskuldsson, A. PLS Regression Methods. *J. Chemometrics* **1988**, *2*, 211–228.
- (19) Larimore, W. Optimal Reduced Rank Modeling, Prediction, Monitoring, and Control using Canonical Variate Analysis. In *IFAC 1997 International Symposium on Advanced Control of Chemical Processes*; Elsevier Science, Ltd.: Amsterdam, 1997; pp 61–66.
- (20) Beebe, K.; Kowalski, B. An Introduction to Multivariable Calibration and Analysis. *Anal. Chem.* **1987**, *59*, 1007–1017.
- (21) Di Ruscio, D. A weighted view on the partial least-squares algorithm. *Automatica* **2000**, *36*, 831–850.
- (22) Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley: New York, 1989.
- (23) Kresta, J.; Marlin, T.; MacGregor, J. Development of Inferential Process Models Using PLS. *Comput. Chem. Eng.* **1994**, *18*, 597–611.
- (24) Piovoso, M.; Kosanovich, K. Applications of Multivariate Statistical Methods to Process Monitoring and Controller Design. *Int. J. Control* **1994**, *59*, 743–765.
- (25) Wise, B.; Gallagher, N.; MacGregor, J. The Process Chemometrics Approach to Process Monitoring and Fault Detection. In *IFAC Workshop on On-line Fault Detection and Supervision in the Chemical Process Industries*, Newcastle, U.K.; Elsevier Science, Ltd.: Amsterdam, 1995; pp 1–21.
- (26) Ricker, N. L. The Use of Biased Least-Squares Estimators for Parameters in Discrete-Time Pulse-Response Models. *Ind. Eng. Chem. Res.* **1988**, *27*, 343–350.
- (27) MacGregor, J.; Kourti, T.; Kresta, J. Multivariable Identification: A Study of Several Methods. In *IFAC ADICHEM '91 Preprints*, Toulouse, France; Najim, K., Babary, J., Eds.; Elsevier Science, Ltd.: Amsterdam, 1991; pp 369–375.

- (28) Qin, S. Partial Least Squares Regression for Recursive System Identification. In *32nd IEEE Conference on Decision and Control*, San Antonio, TX; IEEE: Piscataway, NJ, 1993; pp 2617–2622.
- (29) Kaspar, M.; Ray, W. Chemometric Methods for Process Monitoring and High-Performance Controller Design. *AIChE J.* **1992**, *38*, 1593–1608.
- (30) Anderson, T. *An Introduction to Multivariate Statistical Analysis*; John Wiley: New York, 1984.
- (31) Larimore, W. Canonical Variate Analysis in Control and Signal Processing. In *Statistical Methods in Control and Signal Processing*; Katayama, T.; Sugimoto, S., Eds.; Marcel Dekker: New York, 1997.
- (32) Larimore, W. Automated Multivariable System Identification and Industrial Applications. In *Proceedings of the American Control Conference*, San Diego, CA; IEEE: Piscataway, NJ, 1999; pp 1148–1162.
- (33) Larimore, W. Automated and Optimal System Identification by Canonical Variables. In *Proceedings of the American Control Conference*, Baltimore, MD; IEEE: Piscataway, NJ, 1994; pp 1640–1644.
- (34) Akaike, H. Markovian Representation of Stochastic Processes by Canonical Variables. *SIAM J. Control* **1975**, *13*, 162–173.
- (35) Akaike, H. Canonical Correlation and Information. In *System Identification: Advances and Case Studies*; Mehra, R., Lainiotis, D., Eds.; Academic Press: **1976**.
- (36) van Overschee, P.; De Moor, B. *Subspace Identification for Linear Systems*; Kluwer Academic Publishers: Boston, MA, 1996.
- (37) van Overschee, P.; De Moor, B. N4SID: Subspace Algorithms for the Identification of Combined Deterministic–Stochastic Systems. *Automatica* **1994**, *30*, 75–93.
- (38) van Overschee, P.; De Moor, B. A Unifying Theorem for Three Subspace System Identification Algorithms. *Automatica* **1995**, *31*, 1853–1864.
- (39) Cho, Y. M.; Xu, G.; Kailath, T. Fast Identification of State-Space Models via Exploitation of Displacement Structure. *IEEE Trans. Autom. Control* **1994**, *39*, 2004–2017.
- (40) Basseville, M.; Nikiforov, I. *Detection of Abrupt Changes*; Prentice Hall: Englewood Cliffs, NJ, 1995.
- (41) Juricek, B.; Seborg, D.; Larimore, W. Identification of the Tennessee Eastman Challenge Process Using Subspace Methods. *Control Eng. Pract.*, **2001**, *9*, 1337–1351.
- (42) Kullback, S. *Information and Statistics*; John Wiley: New York, 1959.
- (43) Seborg, D.; Edgar, T.; Mellichamp, D. *Process Dynamics and Control*; John Wiley: New York, 1989.
- (44) Wise, B.; Gallagher, N. *PLS Toolbox*; Eigenvector Research, Inc.: Manson, WA, **1996**.
- (45) Ljung, L. *System Identification Toolbox, Version 4.0.5*; Mathworks, Inc.: Natick, MA, 1995.
- (46) Larimore, W. *ADAPT<sub>X</sub> Automated System Identification Software Users Manual*; Adaptics, Inc.: McLean, VA, 2001.
- (47) Godfrey, K., Ed. *Perturbation Signals for System Identification*; Prentice Hall: New York, 1993.
- (48) Juricek, B.; Larimore, W.; Seborg, D. Reduced Rank ARX and Subspace Identification for Process Control. In *IFAC DYCOPS Symposium*, Corfu, Greece; Georgakis, C., Ed.; Elsevier Science, Ltd.: Amsterdam, 1998; pp 247–252.
- (49) Zhou, K.; Doyle, J.; Glover, K. *Robust and Optimal Control*; Prentice Hall: Upper Saddle River, NJ, 1996.
- (50) Smith, R. Model Validation for Robust Control: An Experimental Process Control Application. *Automatica* **1995**, *31*, 1637–1647.
- (51) Smith, R.; Poolla, K. Recent Results on Model Validation. *J. Soc. Instrum. Control Eng. (Japan)* **1995**, *34*, 176–181.
- (52) Larimore, W. Accuracy Confidence Bands Including the Bias of Model Under-fitting. In *Proceedings of the American Control Conference*, San Francisco, CA; IEEE: Piscataway, NJ, 1993; pp 1995–1999.

Received for review August 9, 2000

Revised manuscript received September 5, 2001

Accepted February 11, 2002

IE000740G