Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods†

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · SEPTEMBER 1999

Impact Factor: 2.59 · DOI: 10.1021/ie990110i

3 AUTHORS, INCLUDING:

Joe S Qin
University of Southern California
215 PUBLICATIONS **9,959** CITATIONS

Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods†

Sergio Valle, Weihua Li, and S. Joe Qin*

Department of Chemical Engineering, The University of Texas at Austin, Austin, Texas 78712

One of the main difficulties in using principal component analysis (PCA) is the selection of the number of principal components (PCs). There exist a plethora of methods to calculate the number of PCs, but most of them use monotonically increasing or decreasing indices. Therefore, the decision to choose the number of principal components is very subjective. In this paper, we present a method based on the variance of the reconstruction error to select the number of PCs. This method demonstrates a minimum over the number of PCs. Conditions are given under which this minimum corresponds to the true number of PCs. Ten other methods available in the signal processing and chemometrics literature are overviewed and compared with the proposed method. Three data sets are used to test the different methods for selecting the number of PCs: two of them are real process data and the other one is a batch reactor simulation.

1. Introduction

Principal component analysis (PCA) has wide applications in signal processing,[1] factor analysis,[2] chemometrics,[3] and chemical processes data analysis.[4,5] Although the original concept of PCA was introduced by Pearson[33] and developed by Hotelling,[32] the vast majority of applications happened only in the last few decades with the popular use of computers. In the chemical engineering area. the applications of PCA include the following:

(1) monitoring of batch and continuous processes,[6,7] (2) extraction of active biochemical reactions,[8] (3) product quality control in the principal component subspace,[9] (4) missing value replacement[10] (5) sensor fault identification and reconstruction,[11] (6) process fault identification and reconstruction,[12] and (7) disturbance detection.[13] Early PCA applications to the quality control of chemical processes are documented in Jackson.[4]

A key issue in developing a PCA model is to choose the adequate number of PCs to represent the system in an optimal way. If fewer PCs are selected than required, a poor model will be obtained and an incomplete representation of the process results. On the contrary, if more PCs than necessary are selected, the model will be overparameterized and will include noise. Zwick and Velicer[20] compared five different approaches available in the literature. Himes et al.[34] provide further comparisons of several approaches using the Tennessee Eastman problem. Different approaches had been proposed in the past to select the optimal number of PCs:

(1) Akaike information criterion (AIC),[14] (2) minimum description length (MDL),[15] (3) imbedded error function (IEF),[16] (4) cumulative percent variance (CPV),[2] (5) scree test on residual percent variance (RPV),[17,18] (6) average eigenvalue (AE),[19] (7) parallel analysis (PA),[20] (8) autocorrelation (AC),[21] (9) cross validation based on the PRESS and *R ratio*,[3,22,23] and (10) variance of the reconstruction error (VRE).[29]

In the signal processing literature, PCA has been used to determine the number of independent source signals from noisy observations by selecting the number of significant principal components.[1] Under the assumption that measurement noise corresponds to the smallest equal eigenvalues of the *covariance matrix*, the Akaike information criterion and the minimum description length principle have been applied to selecting the number of PCs.[1] These criteria apply to PCA based on the *covariance matrix* of the data and do not work on the correlation-based PCA of the data, although correlation-based PCA is preferred in chemical processes to equally weight all variables by variance scaling. One advantage of the AIC and MDL criteria is that they have a solid statistical basis and are theoretically shown to have a minimum number of PCs.

To develop a selection criterion that works with both correlation-based and covariance-based PCA, Qin and Dunia[29] propose a new criterion to determine the number of PCs based on the best reconstruction of the variables. An important feature of this new approach is that the proposed index has a minimum (i.e.. nonmonotonic) corresponding to the best reconstruction. When the PCA model is used to reconstruct missing values or faulty sensors, the reconstruction error is a function of the number of PCs. Qin and Dunia[29] use the *variance of the reconstruction error* to determine the number of PCs. The VRE can be decomposed into a portion in the principal component subspace (PCS) and a portion in the residual subspace (RS). The portion in the RS is shown to decrease monotonically with the number of PCs and that, in the PCS in general, increases with the number of PCs. As a result, the VRE always has a minimum which points to the optimal number of PCs for best reconstruction. In the case of missing value reconstruction, the VRE for different missing patterns are combined with weights based on the frequency of occurrence. In the case of faulty sensor identification and reconstruction, the VREs are weighted based on the variance of each variable. In the special

* To whom correspondence should be addressed. Tel: (512)-471-4417. Fax: (512)471-7060. E-mail: qin@che.utexas.edu.

† An earlier version of the paper was presented at the AIChE Annual Meeting, Miami, FL, Nov 16−20, 1998.

10.1021/ie990110i CCC: $18.00
Published on Web 09/30/1999

case where each sensor is reconstructed once, the procedure can be thought of as a cross validation by *leaving sensors out*, in contrast to the traditional *leaving samples out* cross validation.

The ultimate purpose of this paper is 3-fold:

(1) Introduce the VRE, AIC, MDL, and other related methods for the selection of the number of principal components in chemical process data modeling.
(2) Provide further analysis and comparison of the VRE method with other methods, including the AIC and MDL criteria.
(3) Compare and benchmark various methods for selecting the number of PCs using data from a simulated batch reactor and two industrial processes data sets. The effectiveness of these methods will be tested and bench-marked on the basis of these data sets.

The organization of the paper is given as follows. Section 2 describes the selection criteria based on AIC, MDL, and IEF methods. Section 3 describes different selection criteria commonly used in the PCA literature. Section 4 presents the VRE method based on the correlation matrix and covariance matrix. Section 5 compares the methods using three examples: a simulated batch reactor, a boiler process, and an incineration process. Section 6 concludes the paper.

2. Criteria Based on AIC, MDL, and IEF

The AIC and MDL criteria are popular in signal processing, while the IEF approach is proposed in factor analysis.[16] Although these three methods are developed in different areas. There are common attributes about them: (1) they work only with covariance-based PCA; (2) the variances of measurement noise in each variable are assumed to be identical; and (3) they demonstrate a minimum over the number of PCs given the above assumptions.

In this section, we first present the notation and basic relations for PCA. Then we provide a linkage between the use of PCA for signal processing and the use of PCA for process modeling. Finally, the criteria for AIC, MDL, and IEF are presented.

2.1. PCA Notation. Let $\mathbf{x} \in R^m$ denote a sample vector of m sensors. Assuming that there are N samples for each sensor, a data matrix $\mathbf{X} \in R^{N\times m}$ is composed with each row representing a sample. The matrix $\mathbf{X}$ is scaled to zero-mean for covariance-based PCA and, in addition to unit variance for correlation-based PCA, by either the NIPALS[24] or a singular-value decomposition (SVD) algorithm. The matrix $\mathbf{X}$ can be decomposed into a score matrix $\mathbf{T}$ and a loading matrix $\mathbf{P}$ whose columns are the right singular vectors of $\mathbf{X}$,

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \tilde{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \tilde{\mathbf{T}}\tilde{\mathbf{P}}^{\mathrm{T}} = [\mathbf{T} \quad \tilde{\mathbf{T}}][\mathbf{P} \quad \tilde{\mathbf{P}}]^{\mathrm{T}} \quad (1)$$

where $\tilde{\mathbf{X}} = \tilde{\mathbf{T}}\tilde{\mathbf{P}}^{\mathrm{T}}$ is the residual matrix. Since the columns of $\mathbf{T}$ are orthogonal, the covariance matrix is

$$\mathbf{S} \approx \frac{1}{N-1}\mathbf{X}^{\mathrm{T}}\mathbf{X} = [\mathbf{P} \quad \tilde{\mathbf{P}}]\,\Lambda\,[\mathbf{P} \quad \tilde{\mathbf{P}}]^{\mathrm{T}} \quad (2)$$

where

$$\Lambda = \frac{1}{N-1}[\mathbf{T} \quad \tilde{\mathbf{T}}]^{\mathrm{T}}[\mathbf{T} \quad \tilde{\mathbf{T}}] = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_m\} \quad (3)$$

$$\lambda_i = \frac{1}{N-1}\mathbf{t}_i^{\mathrm{T}}\mathbf{t}_i = \mathrm{var}\{\mathbf{t}_i\} \quad (4)$$

and $\mathbf{t}_i$ is the ith column of $\mathbf{T}$ and λ_i are the eigenvalues of the covariance matrix. If N is very large, the approximate equality in eq 2 becomes an equality. We assume the equality holds in this paper unless otherwise specified. For variance scaled $\mathbf{X}$, eq 2 is the correlation matrix $\mathbf{R}$.

A sample vector $\mathbf{x}$ can be projected on the PCS (S_p), which is spanned by $\mathbf{P} \in R^{m\times l}$, and RS ($S_r$), respectively,

$$\hat{\mathbf{x}} = \mathbf{P}\mathbf{P}^{\mathrm{T}}\mathbf{x} \in S_p \quad (5)$$

$$\tilde{\mathbf{x}} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^{\mathrm{T}}\mathbf{x} = (\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})\,\mathbf{x} \in S_r \quad (6)$$

Since S_p and S_r are orthogonal,

$$\hat{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} = 0 \quad (7)$$

and

$$\hat{\mathbf{x}} + \tilde{\mathbf{x}} = \mathbf{x} \quad (8)$$

The task for determining the number of PCs is to choose l such that $\hat{\mathbf{x}}$ contains mostly information and $\hat{\mathbf{x}}$ contains noise.

2.2. PCA for Signal Processing and Process Modeling. In signal processing, PCA has been used to detect q independent signal sources from m observation variables, that is,

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{n}(k) \quad (9)$$

where $\mathbf{x}(k) \in R^m$ is a vector of observation variables, $\mathbf{s}(k) \in R^q$ is a vector of signal sources, $\mathbf{n}(k) \in R^m$ is a vector of observation noise, and $\mathbf{A} \in R^{m\times q}$ is a matrix with appropriate dimension. This model is used for radar signal detection,[1] for instance.

In chemical process modeling, the physical principles that govern the process behavior are mainly material and energy balances. Assuming there exists a linear (or linearized) relationship among the process variables, the following steady-state relation can be derived,

$$\mathbf{B}\bar{\mathbf{x}}(k) = \mathbf{0} \quad (10)$$

$$\mathbf{x}(k) = \bar{\mathbf{x}}(k) + \mathbf{n}(k) \quad (11)$$

where $\bar{\mathbf{x}}(k) \in R^m$ are the true process variables free of measurement noise, $\mathbf{B} \in R^{\tilde{q}\times m}$ is the model matrix, and $\mathbf{n}(k) \in R^m$ is the measurement noise.

While eqs 9 and 11 represent different physical phenomena, we have the following remarks under the condition that $q = m - \tilde{q}$:

(1) Equation 9 can be converted to eqs 10 and 11 by premultiplying

$$\mathbf{B} = [\mathbf{I} - \mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}] \quad (12)$$

on both sides of eq 9, where

$$\bar{\mathbf{x}}(k) = \mathbf{x}(k) - n(k) = \mathbf{A}\mathbf{s}(k)$$

(2) Equation 11 can be converted to eq 9 by selecting $\mathbf{A}$ as a basis in R($\mathbf{B}^{\mathrm{T}}$), where R($\mathbf{B}^{\mathrm{T}}$) is the range space of $\mathbf{B}^{\mathrm{T}}$.

On the basis of the equivalence of eqs 9 and 11, we can directly apply AIC and MDL criteria originally

proposed for the detection of signal sources in signal processing[1] to the determination of the number of PCs in process modeling.

2.3. AIC and MDL Criteria. AIC and MDL are respectively proposed by Akaike[14] and Rissanen[15] for model selection. Assuming that $\mathbf{n}(k) \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ is independently identically distributed, we can describe the covariance matrix of $\mathbf{x}(k)$ in the following form,

$$\mathbf{S} = E\{\mathbf{x}\mathbf{x}^T\} = \bar{\mathbf{S}} + \sigma^2\mathbf{I} \tag{13}$$

where $\bar{\mathbf{S}} = E\{\bar{\mathbf{x}}\bar{\mathbf{x}}^T\}$ has a rank $q \leq m$.

Denoting the eigenvalues of $\mathbf{S}$ by $\lambda_1 \geq \lambda_2, \cdots, \geq \lambda_m$, it is easy to verify that the smallest $m - q$ eigenvalues of $\mathbf{S}$ are equal to σ^2, that is,

$$\lambda_{q+1} = \lambda_{q+2} = \cdots = \lambda_m = \sigma^2$$

and

$$\mathbf{S} = \sum_{i=1}^{q}(\lambda_i - \sigma^2)\mathbf{v}_i\mathbf{v}_i^T + \sigma^2\mathbf{I}$$

where $\lambda_1, \cdots, \lambda_q$ and $\mathbf{v}_1, \cdots, \mathbf{v}_q$ are the q largest eigenvalues and the associated eigenvectors of $\mathbf{S}$. We denote

$$\mathbf{w} = [\lambda_1, \cdots, \lambda_q, \sigma^2, \mathbf{v}_1^T, \cdots, \mathbf{v}_q^T]^T \in R^{(m+1)q+1}$$

as the parameter vector for the PCA model. Wax and Kailath[1] propose to determine the number of PCs based on AIC and MDL, which can be summarized as follows. Given a set of N observations $\mathbf{x}(1), \cdots, \mathbf{x}(N)$, which are assumed to be a sequence of independent and identically distributed zero-mean Gaussian vectors, the maximum likelihood estimate of w[1,25] is given by

$$\hat{\mathbf{w}} = \left[d_1, \cdots, d_l, \frac{1}{m-l}\sum_{i=l+1}^{m} d_i, \mathbf{c}_1^T, \cdots, \mathbf{c}_l^T\right]^T$$

where $d_1, \cdots, d_m$ are the eigenvalues and $\mathbf{c}_1, \cdots, \mathbf{c}_l$ the eigenvectors of the sample covariance matrix, which is $\hat{\mathbf{S}} = 1/(N-1)\sum_{k=1}^{N}\mathbf{x}(k)\mathbf{x}^T(k)$. Wax and Kailath[1] have shown that the AIC and MDL functions will have the following forms,

$$\mathrm{AIC}(l) = -2\log\left(\frac{\prod_{s=l+1}^{m} d_s^{1/(m-l)}}{\frac{1}{m-l}\sum_{s=l+1}^{m} d_s}\right)^{(m-l)N} + 2M \tag{14}$$

$$\mathrm{MDL}(l) = -2\,1\log\left(\frac{\prod_{s=l+1}^{m} d_s^{1/(n-l)}}{\frac{1}{m-l}\sum_{s=l+1}^{m} d_s}\right)^{(m-l)N} + M\log N \tag{15}$$

where M is the number of independent parameters in the model vector $\hat{\mathbf{w}}$. For complex valued signals, the eigenvectors $\mathbf{c}_1, \mathbf{c}_2, \cdots, c_l$ are complex, and the number of independent parameters is[1]

$$M = l(2m - l)$$

For real valued signals such as those encountered in chemical processes, the eigenvectors $\mathbf{c}_1, \mathbf{c}_2, \cdots, c_l$ are real. Since these vectors are normalized and mutually orthogonal, the number of independent parameters in these eigenvectors is

$$M = \frac{l}{2}(m - l + 1)$$

This formula is used for the application examples in this paper.

The first term in AIC(l) or MDL(l) tends to decrease and the second term increases with l. Theoretically, a minimum results for $l \in [1, m]$. Wax and Kailath[1] show that, for a large data length N, MDL gives a consistent estimate, while AIC tends to overestimate the number of PCs.

2.4. Imbedded Error Function. The IEF method[16] is based on the idea that the measurement space can be divided in two subspaces: a principal component subspace which represents the important signals of the process and a residual subspace that has only noise in its measurements. The only requirement to calculate this function is to know the eigenvalues. The calculation of this function is done using only the residual eigenvalues:

$$\mathrm{IEF}(l) = \left[\frac{l\sum_{j=l+1}^{m}\lambda_j}{Nm(m-l)}\right]^{1/2} \tag{16}$$

where l represents the number of PCs used to represent the data. The IEF method assumes that the data contain signals plus measurement errors similar to those in eq 11. Each PC will contain a portion of the signals and a portion of imbedded errors. Before all signals are extracted, the IEF(l) contains a mixture of signal and imbedded errors, which tends to decrease with l. At the point all signals are extracted, the IEF(l) contains imbedded errors only and should increase with l. The minimum value of the IEF, if exists, corresponds to the number of PCs needed to describe the data by leaving out measurement errors only.

3. Selection Criteria in Chemometrics

3.1. Cumulative Percent Variance. The CPV[2] is a measure of the percent variance captured by the first l PCs:

$$\mathrm{CPV}(l) = 100\left[\frac{\sum_{j=1}^{l}\lambda_j}{\sum_{j=1}^{m}\lambda_j}\right]\% \tag{17}$$

With this criterion one selects a desired CPV, e.g., 90%, 95%, or 99%, which is very subjective. While we would like to account for as much of the variance as possible, we want to retain as few principal components as possible. The decision then becomes a balance between the amount of parsimony and comprehensiveness of the model. We must decide what balance will serve our purpose. The CPV method to select the number of PCs is often ambiguous because the CPV is monotonically increasing with the number of PCs.

3.2. Scree Test on Residual Percent Variance. The Scree test on RPV[17,18] is an empirical procedure to select the number of PCs using the RPV as a basis. The method looks for a "knee" point in the *residual percent variance* plotted against the number of principal components. The method is based on the idea that the residual variance should reach a steady state when the factors begin to account for random errors. When a break point is found or when the plot stabilizes, that point corresponds to the number of principal components to represent the process. The RPV uses only the residual eigenvalues to do the calculations:

$$\mathrm{RPV}(l) = 100\left[\frac{\sum_{j=l+1}^{m}\lambda_j}{\sum_{j=1}^{m}\lambda_j}\right]\% \qquad (18)$$

The implementation of this method is relatively easy, but in some cases it is difficult to find a knee if the curve decreases smoothly. This point is observed in section 5.3 of the paper using the incineration process data.

3.3. Average Eigenvalue. The average eigenvalue approach[19] is a somewhat popular criterion to choose the number of PCs. This criterion accepts all eigenvalues with values above the average eigenvalue and rejects those below the average. The reason is that a PC contributing less than an "average" variable is insignificant. For covariance-based PCA, the average eigenvalue is $^1/_m$ trace(**S**), and for correlation-based PCA the average eigenvalue $^1/_m$ trace(**R**), which is 1. Then all the eigenvalues above 1 will be selected as the principal eigenvalues to form the model. Therefore, this criterion is also known as the eigenvalue-one rule.

3.4. Paralle1 Analysis. The PA method basically builds PCA models for two matrices: one is the original data matrix and the other is an uncorrelated data matrix with the same size as the original matrix. This method was developed originally by Horn[26] to enhance the performance of the Scree test. When the eigenvalues for each matrix are plotted in the same figure, all the values above the intersection represent the process information and the values under the intersection are considered noise. Because of this intersection, the parallel analysis method is not ambiguous in the selection of the number of PCs.

For a large number of samples, the eigenvalues for a correlation matrix of uncorrelated variables are 1. In this case, the PA method is identical to the AE method. However, when the samples are generated with a finite number of samples, the initial eigenvalues exceed 1, while the final eigenvalues are under 1. That is why Horn[26] suggested comparing the correlation matrix eigenvalues for uncorrelated variables with those of a real data matrix based on the same sample size.

3.5. Autocorrelation. According to Shrager and Hendler,[21] it is possible to use an autocorrelation function to separate the noisy eigenvectors from the smooth ones using either the score matrix or the loading matrix. A simple statistic for detecting noise is the autocorrelation function of order 1 for the PCA scores,

$$\mathrm{AC}(l) = \sum_{i=1}^{N-1} t_{i,l}t_{i+1,l} \qquad (19)$$

where N represents the number of samples. A value greater than +0.5 indicates smoothness of the scores, while a value smaller than 0.5 indicates that the component contains mainly noise and should not be included in the model.[21] Two major drawbacks of this method are (1) this 0.5 threshold is somewhat arbitrary and (2) a large variance PC can have little autocorrelation. The second drawback is observed in the incineration data in section 5 of this paper.

3.6. Cross Validation Based on the *R* Ratio. Cross validation[3,27] is a popular statistical criterion to choose the number of factors in PCA. The basis of this method is to estimate the values of some deleted data from a model and then compare these estimates with the actual values. Wold[3] proposes a cross-validation method known as the *R ratio*. This method randomly divides the data set **X** into G groups. Then, for a starting value of $l = 1$, a reduced data set is formed by deleting each group in turn and the parameters in the model are calculated from the reduced data set. The predicted error sum of squares (PRESS) is calculated from the predicted and actual values of the deleted objects. In the meantime, the residual sum of squares (RSS) the next component is extracted is calculated. A ratio,

$$R(l) = \frac{\mathrm{PRESS}(l)}{\mathrm{RSS}(l-1)} \qquad (20)$$

is then calculated starting with $l = 1, \cdots, m - 1$.

A value of the *R ratio* less than 1 indicates that the newly added component in the model improved the prediction and hence the calculation proceeds. A value of this ratio larger than 1 indicates that the new component does not improve the prediction and hence should be deleted. Since the PCA model calculation involves randomly deleted data, the NIPALS algorithm is used which can handle missing values. The brief procedure for calculating the *R ratio* is summarized as follows. Refer to Wold[3] for the detailed procedure.

(1) Scale the data **X** to zero-mean (and unit variance for correlation-based PCA). Set $l = 1$ and $\mathbf{X}_0 = \mathbf{X}$.
(2) Calculate:

$$\mathrm{RSS}(l-1) = \mathrm{trace}(\mathbf{X}_{l-1}^{\mathrm{T}}\mathbf{X}_{l-1})$$

(3) Divide $\mathbf{X}_{l-1}$ randomly into G groups. For all but one group calculate the lth PCA factor using the NIPALS algorithm that handles missing values. After leaving out each group in turn. calculate PRESS(l) for all groups of data.
(4) Calculate the *R ratio* using eq 20. If $R < 1$, then calculate the lth PCA factor based on all data $\mathbf{X}_{l-1}$, set $\mathbf{X}_l = \mathbf{X}_{l-1} - \mathbf{t}_l\mathbf{p}_l^{\mathrm{T}}$, set $l := l + 1$, and return to step 2. If $R \geq 1$, terminate the program.

3.7. Cross Validation Based on PRESS. In the *R ratio* method, the PRESS is calculated in a manner that $\mathbf{X}_{l-1}$ may be rerandomized after each increment of l. If a new PC results in a PRESS that is larger than the RSS without this PC, it is discarded. An alternative approach is to use PRESS alone to determine the number of PCs. This procedure is related to Wold[3] and has been successfully used in practice.[28] The procedure to use PRESS to determine the number of PCs is summarized as follows:

(1) Scale the data **X** to zero-mean (and unit variance for correlation-based PCA). Set $l = 1$.

(2) Divide **X** randomly into G groups. For all but one group calculate all PCA factors using the NIPALS algorithm that handles missing values. After leaving out each group in turn, calculate PRESS(l) for all groups of data.
(3) A minimum in PRESS(l) corresponds to the best number of PCs to choose.

Note that the number of PCs calculated in this procedure can be different from that in the method of the *R ratio*.

4. Variance of the Reconstruction Error Criterion

The variance of the reconstruction error method was developed by Qin and Dunia[29] to select the number of principal components based on the best reconstruction of the variables. An important characteristic of this approach is that the VRE index has a minimum, corresponding to the best reconstruction. When the PCA model is used to reconstruct missing values or faulty sensors, the reconstruction error is a function of the number of PCs. The minimum found in the VRE calculation directly determines the number of PCs. This is because the VRE is decomposed in two subspaces: the principal components subspace and a residual subspace. The portion in PCS has a tendency to increase with the number of PCs, and that in the RS has a tendency to decrease, resulting in a minimum in VRE.

Qin and Dunia[29] consider that the sensor measurement is corrupted with a fault along a direction ξ_i,

$$\mathbf{x} = \mathbf{x}^* + f\xi_i$$

where $\mathbf{x}^*$ is the normal portion, f the fault magnitude, and $\|\xi_i\| = 1$. For sensor faults, ξ_i corresponds to the ith column of an identity matrix. For process faults, ξ_i can be an arbitrary vector with unit norm. The reconstruction of the fault is given by correction along the fault direction, that is,

$$\mathbf{x}_i = \mathbf{x} - f_i\xi_i$$

so that $\mathbf{x}_i$ is most consistent with the PCA model. The difference $\mathbf{x}^* - \mathbf{x}_i$ is known as the *reconstruction error*. Qin and Dunia[29] define the variance of the reconstruction error as

$$u_i \equiv \mathrm{var}\{\xi_i^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_i)\} = \frac{\tilde{\xi}_i^{\mathrm{T}}\mathbf{R}\tilde{\xi}_i}{(\tilde{\xi}_i^{\mathrm{T}}\tilde{\xi}_i)^2} \tag{21}$$

where

$$\tilde{\xi}_i = (\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})\xi_i = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^{\mathrm{T}}\xi_i \tag{22}$$

The problem to find the number of PCs is to minimize u_i with respect to the number of PCs. Considering all possible faults, the VRE to be minimized is defined as

$$\mathrm{VRE}(l) = \sum_{i=1}^{m}\frac{u_i}{\mathrm{var}\{\xi_i^{\mathrm{T}}\mathbf{x}\}} = \sum_{i=1}^{m}\frac{u_i}{\xi_i^{\mathrm{T}}\mathbf{R}\xi_i} \tag{23}$$

The variance-based weighting factors are necessary to equalize the importance of each variable. Note that the above derivation is for a correlation-based PCA sincethe correlation matrix **R** is used. For a covariance-based PCA, one can simply replace with **S** the covariance matrix.

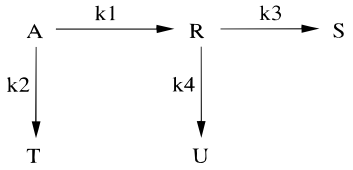

Figure 1. Reaction network for the batch reactor.

The VRE procedure for selecting the number of PCs can be summarized as follows:

(1) Build a PCA model using normal data from all variables.
(2) Reconstruct each variable using other variables and calculate the VRE, u_i.
(3) The total VRE is calculated using eq 23.
(4) The number of PCs that gives the minimum VRE is selected, which corresponds to the best reconstruction.

For a particular variable or fault direction ξ_i, it is possible that $u_i \geq \mathrm{var}\{\xi^{\mathrm{T}}\mathbf{x}\}$, which means the model gives a worse prediction than the mean of the data. In this case, the variable has little correlation with others and should be dropped from the model. Therefore, the VRE can be used to achieve the following objectives: (1) determine the number of PCs by leaving out variables; (2) select variables that can be reliably reconstructed from the model; (3) simultaneously leave out a group of variables; and (4) include a particular fault or disturbance direction in selecting the number of PCs for the purpose of fault or disturbance detection.

4.1. VRE versus PRESS. Like PRESS, the VRE method works for both a correlation-based PCA and covariance-based PCA. Another similarity between VRE and PRESS is that they do not require any fault to occur; they just need to leave out a portion of the data and reconstruct them from the models. Furthermore, the missing value treatment in the NIPALS algorithm is equivalent to the fault reconstruction as shown in Dunia et al.[11] However, the VRE and PRESS methods are different in the following ways:

(1) The VRE method can include process faults with arbitrary direction ξ_i to emphasize particular process faults or disturbances, while PRESS cannot.
(2) If each sensor direction (ξ_i) is used once and only once, the VRE method is similar to cross validation by leaving sensors out, while the PRESS leaves samples out randomly.
(3) In PRESS, as many as G models are built in the presence of "missing" values, while VRE builds one PCA model only based on all available data.
(4) In the VRE method, it is possible to check if a group of data can be reliably reconstructed, while the PRESS method fails to do this rigorously.
(5) The VRE method is also suitable for recursive online calculation, as shown in Li et al.[30]

4.2. Consistency of the VRE Method. Wax and Kailath[1] show that the MDL approach gives a consistent estimate of the number of PCs, while the AIC method

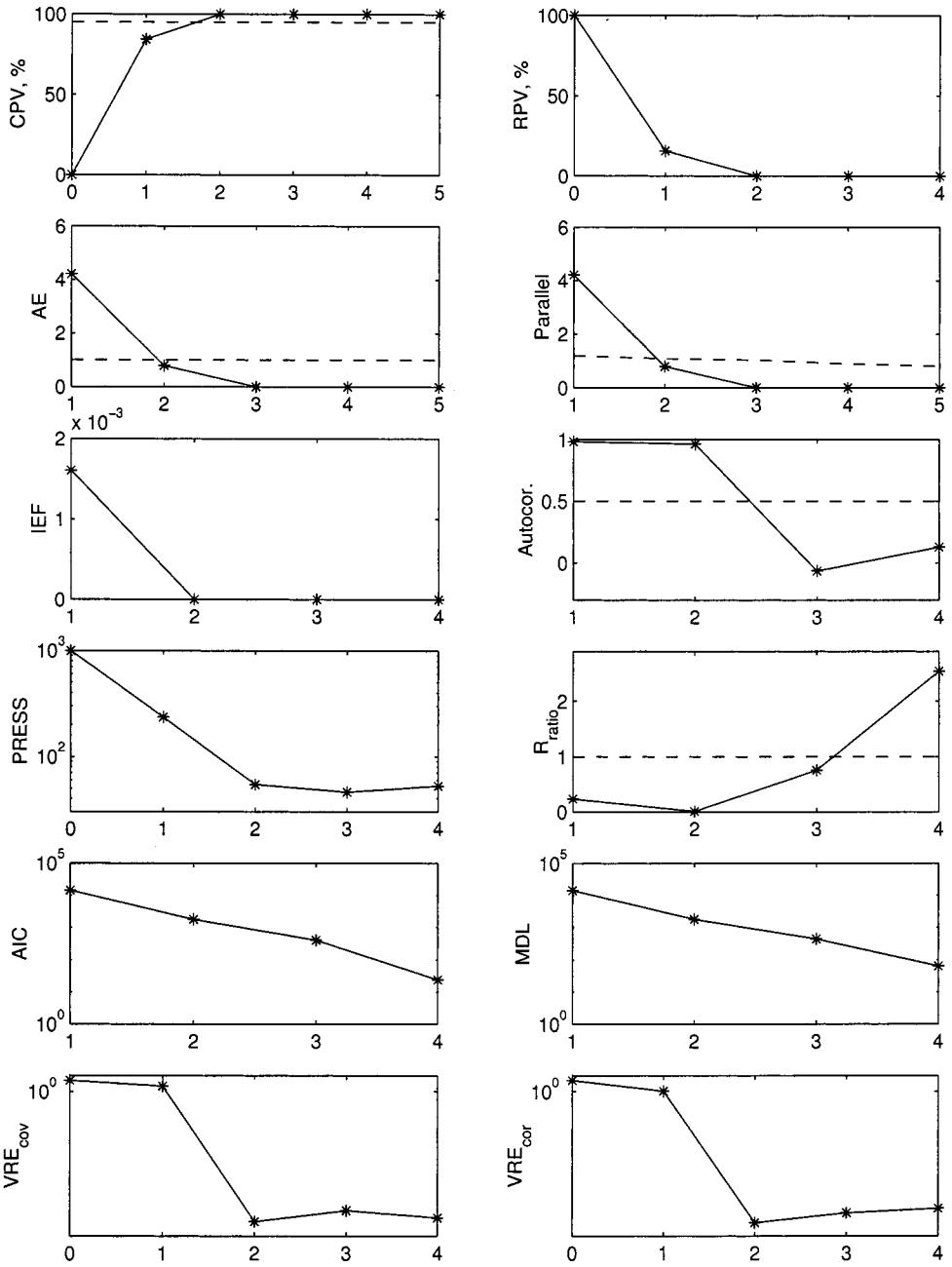

Figure 2. Comparison for the simulated batch reactor free of noise.

does not and tends to overestimate the number of PCs. The consistency of most other methods are not known. For the VRE method, Dunia and Qin[12] show that the VRE demonstrates a minimum, but it is not known whether the minimum corresponds to the true number of PCs. In this section, we establish the consistency conditions for the VRE method.

First, we examine the covariance-based PCA where the noise variances are assumed equal. From eq 13, the eigenvalues of covariance matrix **S** are

$$\lambda_i = \begin{cases} \bar{\lambda}_i + \sigma^2 & 1 \le i \le q \\ \sigma^2 & \mathrm{q} < i \le m \end{cases}$$

where $\bar{\lambda}_i (i = 1, 2, ..., q)$ are the non-zero eigenvalues of $\bar{\mathbf{S}}$, the covariance of $\bar{\mathbf{x}}$. For convenience we require $\bar{\lambda}_1 \ge \bar{\lambda}_2 \ge \cdots \ge \bar{\lambda}_q$. Assuming l PCs are included in the PCA model, the VRE can be written as follows using eq 22,

$$u_{il} = \frac{\tilde{\xi}_{il}^{\mathrm{T}}\tilde{\xi}_{il}}{(\tilde{\xi}_{il}^{\mathrm{T}}\tilde{\xi}_{il})^2} = \frac{\xi_i^{\mathrm{T}}\tilde{\mathbf{P}}_l(\tilde{\mathbf{P}}_l^{\mathrm{T}}\mathbf{S}\tilde{\mathbf{P}}_l)\tilde{\mathbf{P}}_l^{\mathrm{T}}\xi_i}{(\xi_i^{\mathrm{T}}\tilde{\mathbf{P}}_l\tilde{\mathbf{P}}_l^{\mathrm{T}}\xi_i)^2} \equiv \frac{\mathbf{e}_{il}^{\mathrm{T}}\tilde{\mathbf{P}}_l^{\mathrm{T}}\mathbf{S}\tilde{\mathbf{P}}_l\mathbf{e}_{il}}{(\mathbf{e}_{il}^{\mathrm{T}}\mathbf{e}_{il})^2} = \frac{\mathbf{e}_{il}^{\mathrm{T}}\Lambda_l\mathbf{e}_{il}}{(\mathbf{e}_{il}^{\mathrm{T}}\mathbf{e}_{il})^2} \quad (24)$$

where $\mathbf{e}_{il} \equiv \tilde{\mathbf{P}}_l^{\mathrm{T}}\xi_i \in R^l$ and $\Lambda_l = \mathrm{diag}\{\lambda_{l+1}, \lambda_{l+2}, \cdots, \lambda_m\}$.

Given the above analysis, we have the following theorem about the consistency of the covariance-based VRE method.

Theorem 1. Assuming the covariance matrix of the data is given in eq 13, then (1) the VRE achieves a minimum at $l \le q$ and (2) the VRE achieves a global minimum at $l = q$ if $\bar{\lambda}_q \ge \|e_{il}\|^2/\|e_{iq}\|^2\sigma^2$ for $l < q$.

The proof of this theorem is given in Appendix A. We make the following remarks about Theorem 1:

(1) The VRE method will never over estimate the number of PCs.

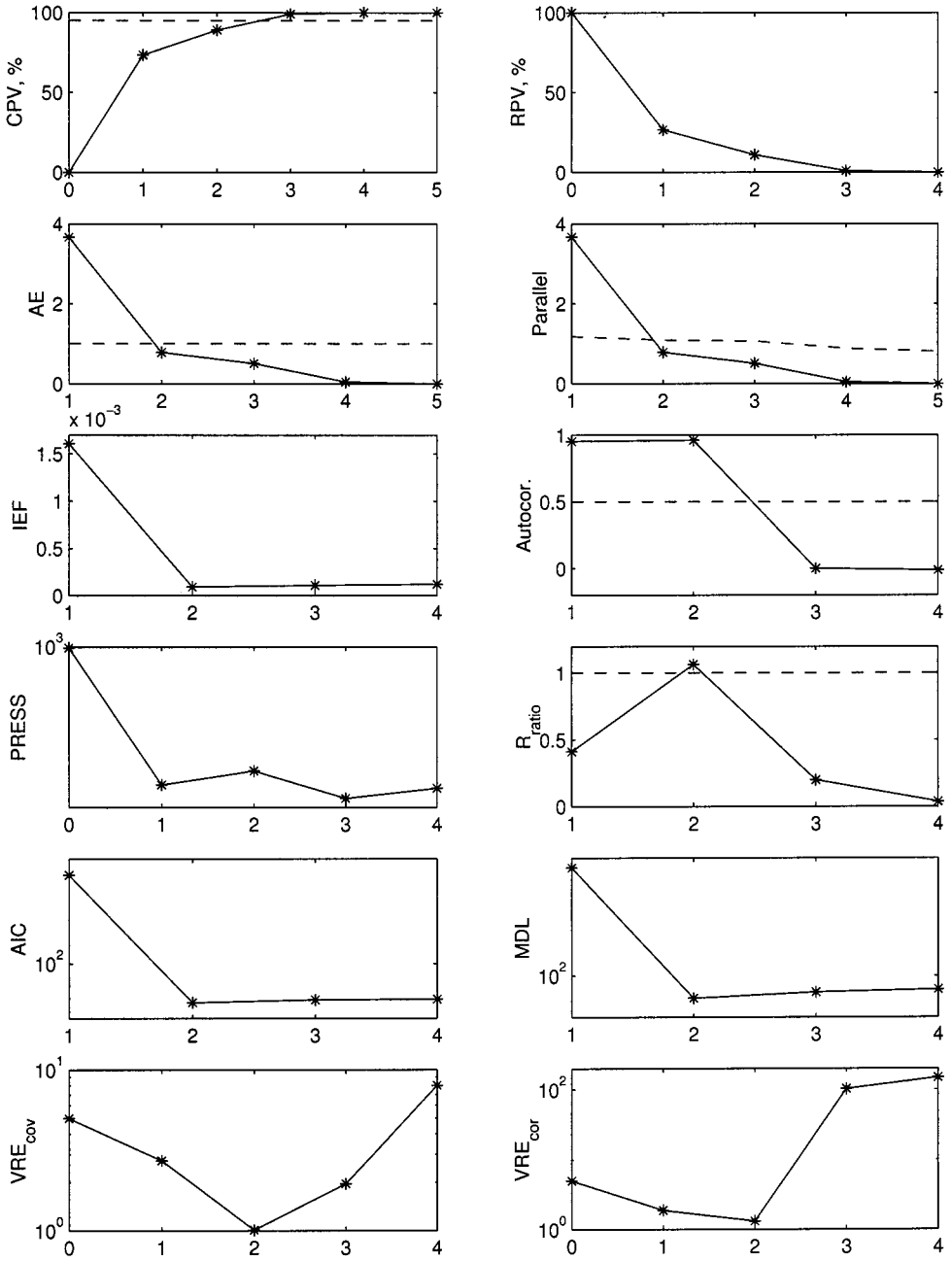

Figure 3. Comparison for the simulated batch reactor with stationary additive Gaussian noise.

(2) The condition for a consistent estimate requires that the variance of the signal is slightly larger than the variance of the noise, which is quite reasonable.

For a correlation-based PCA or the general case of a covariance-based PCA, the measurement errors do not necessarily have the same variance. In this case, we denote the eigenvalues of the correlation (or covariance) matrix as $\lambda_1, \lambda_2, \cdots, \lambda_q, \sigma^2_{q+1}, \sigma^2_{q+2}, \cdots, \sigma^2_m$, where σ^2_i are variance quantities of the noise terms. To guarantee that VRE gives a consistent estimate in this case, we have the following theorem.

Theorem 2. Assuming the eigenvalues of the correlation or covariance matrix with descending order are $\lambda_1, \lambda_2, \cdots, \lambda_q, \sigma^2_{q+1}, \cdots, \sigma^2_m$, the VRE achieves a minimum at $l = q$ if (1) $\sigma^2_m \geq \sigma^2_{q+1}\|e_{il}\|^2/\|e_{iq}\|^2$ for $l \geq q$, and (2) $\lambda_l \geq \sigma^2_{q+1}(1 + \|e_{il}\|^2/\|e_{iq}\|^2)$ for $l < q$.

The proof of this theorem is given in Appendix B. The first condition requires that the noise variances, although different, should not be too different. The second condition requires that the variance of the signal be larger than the variance of the noise.

5. Comparisons on Three Case Studies

5.1. Simulated Batch Reactor Data. An isothermal batch reactor is simulated with four first-order reactions taking place.[31] Figure 1 depicts the reaction network for the batch reactor. The material balance equation for the batch reactor is

$$\frac{dC_i}{dt} = r_i \tag{25}$$

where the reaction rates for the five species are

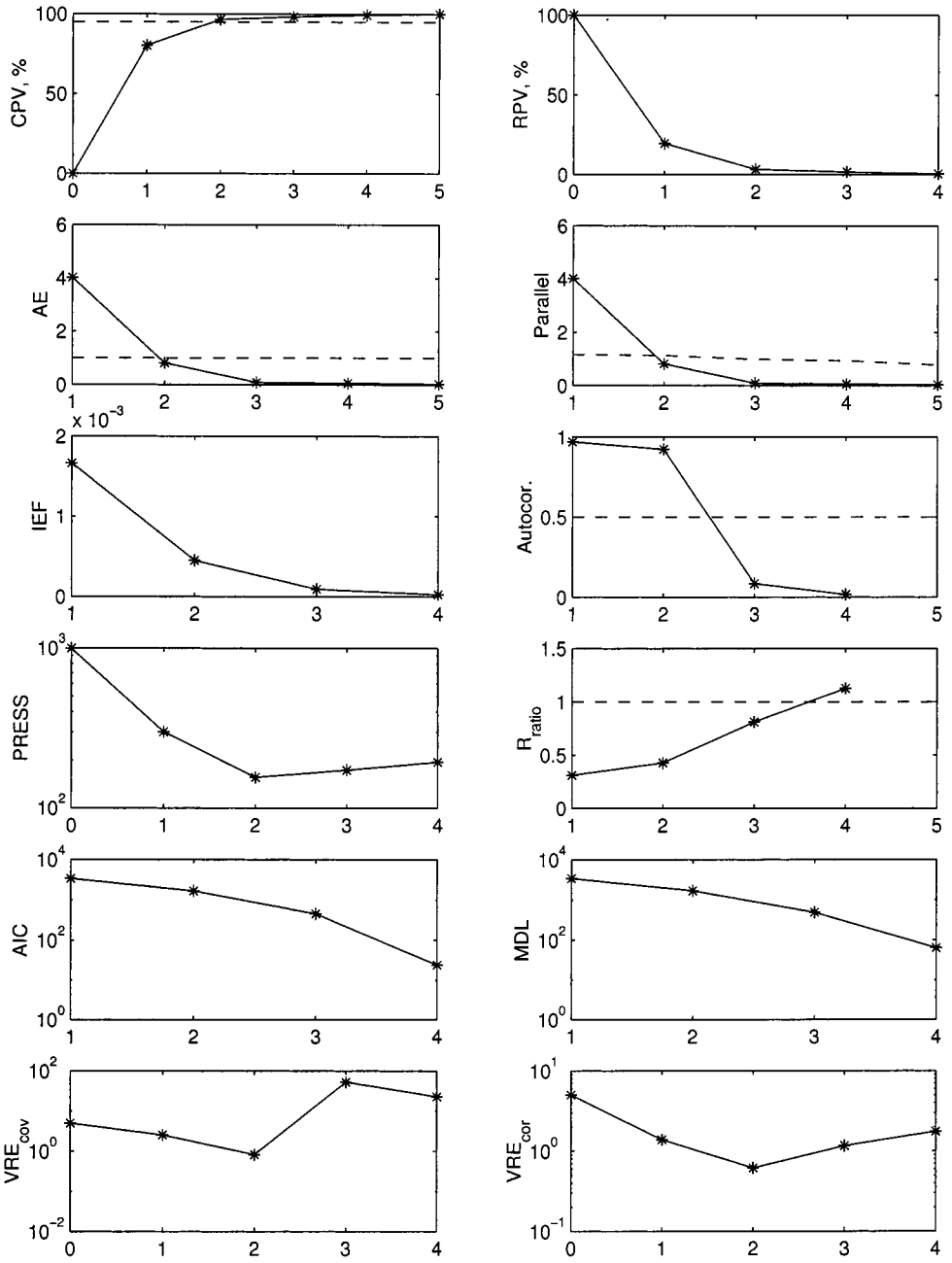

Figure 4. Comparison for the simulated batch reactor with nonstationary multiplicative Gaussian noise.

$$r_A = -k_1 C_A - k_2 C_A = -(k_1 + k_2) C_A$$

$$r_R = k_1 C_A - k_3 C_R - k_4 C_R = k_1 C_A - (k_3 + k_4) C_R$$

$$r_T = k_2 C_A$$

$$r_U = k_4 C_R$$

$$r_S = k_3 C_R$$

The reaction rate constants are

$$k_1 = 10^7 \exp\left\{-\frac{5000}{T}\right\}[=]\mathrm{h}^{-1}$$

$$k_2 = 1.246 \times 10^5 \exp\left\{\frac{4000}{T}\right\}[=]\mathrm{h}^{-1}$$

$$k_3 = 10^{10} \exp\left\{-\frac{9000}{T}\right\}[=]\mathrm{h}^{-1}$$

$$k_4 = 1.246 \times 10^8 \exp\left\{-\frac{7000}{T}\right\}[=]\mathrm{h}^{-1}$$

The temperature was kept constant at 334.6 °C. It was assumed that initially only the raw material A was present and the vessel was well-mixed. Three different cases of measurement noise were simulated for the batch reactor: (1) a noise-free case, (2) stationary additive Gaussian noise with a standard deviation of 0.002, and (3) nonstationary multiplicative Gaussian noise with a standard deviation of 10% of the variables amplitude.

The simulation generated a total of 200 samples for the 5 variables using this kinetic model. The data were scaled to zero-mean and unit variance, and then analyzed with the 11 different methods to determine the number of PCs to adequately represent the model.

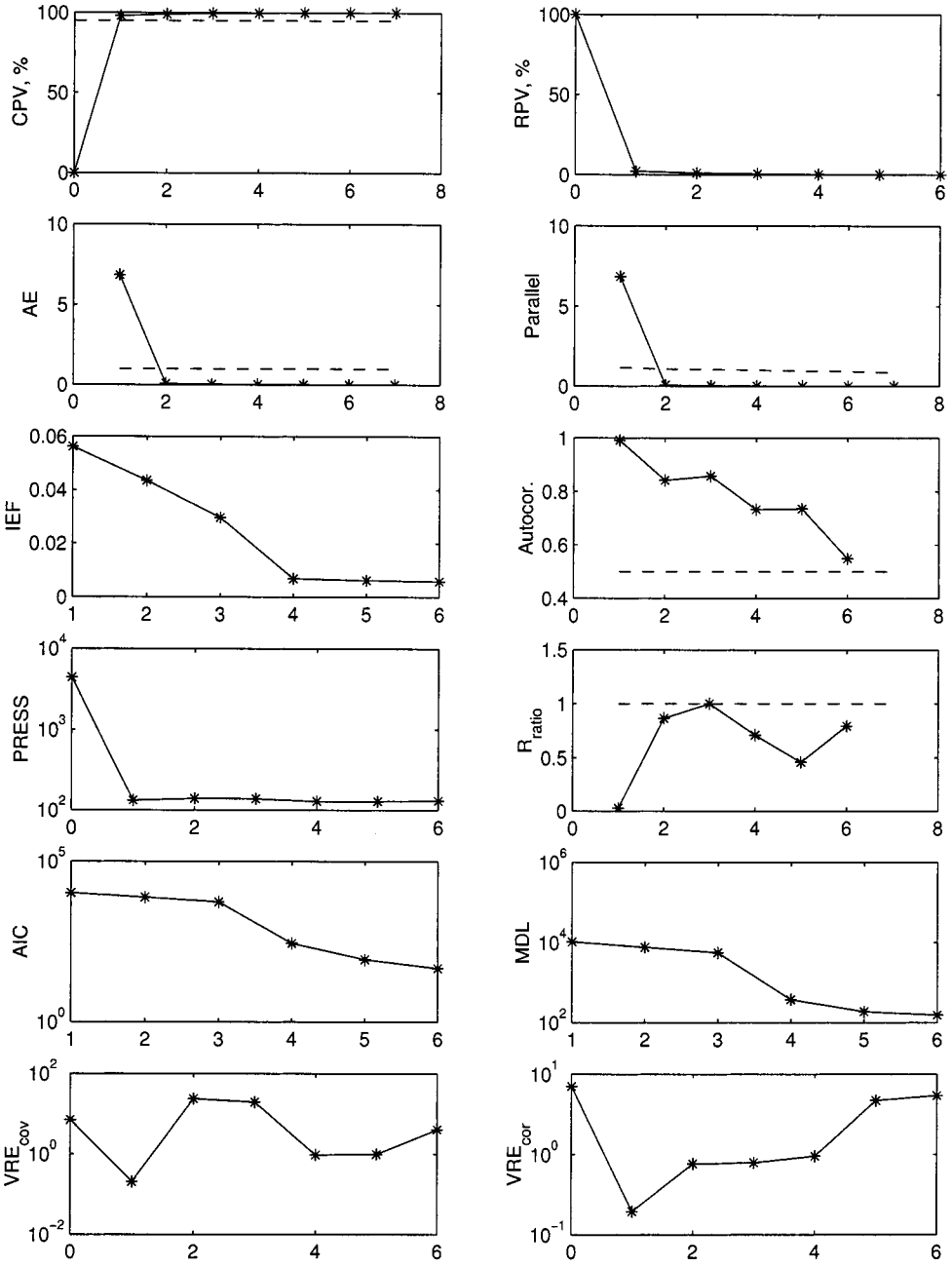

Figure 5. Comparison for a real boiler process.

Table 1. Number of PCs Determined from Each Analyzed Method

	reactor	boiler	incinerator
CPV	3	1	15
RPV	ambiguous	1	ambiguous
AE	1	1	7
PA	1	1	7
IEF	2	no solution	no solution
AC	2	no solution	2
PRESS	3	1	15
R_{ratio}	2	2	5
AIC	2	no solution	no solution
MDL	2	no solution	no solution
VRE_{cov}	2	1	2
VRE_{cor}	2	1	5

The analysis results are shown in Figure 2 for the noise-free case, Figure 3 for the additive noise, and Figure 4 for the case of multiplicative noise. Table 1 summarizes the results for the case of additive noise and shows if the method has found a solution, an ambiguous solution or no solution at all.

Observe that, for the additive noise case, the RPV method shows a smooth curve, making it difficult to decide on the number of PCs. The CPV and PRESS methods choose three PCs. AE and PA determine one PC, and the rest of the seven methods agree in two PCs. We conclude that to have an adequate representation of the process, it is necessary to use two PCs. Also observe in Figure 2 (noise-free) and Figure 4 (multiplicative noise) that the AIC and MDL methods were unable to find a solution.

5.2. Boiler Process Data. The data from an industrial boiler consist of 7 variables and 630 samples. The variables represent temperatures, pressures, flows, and concentrations. Before the analysis was done the data were centered and scaled to zero-mean and unit vari-

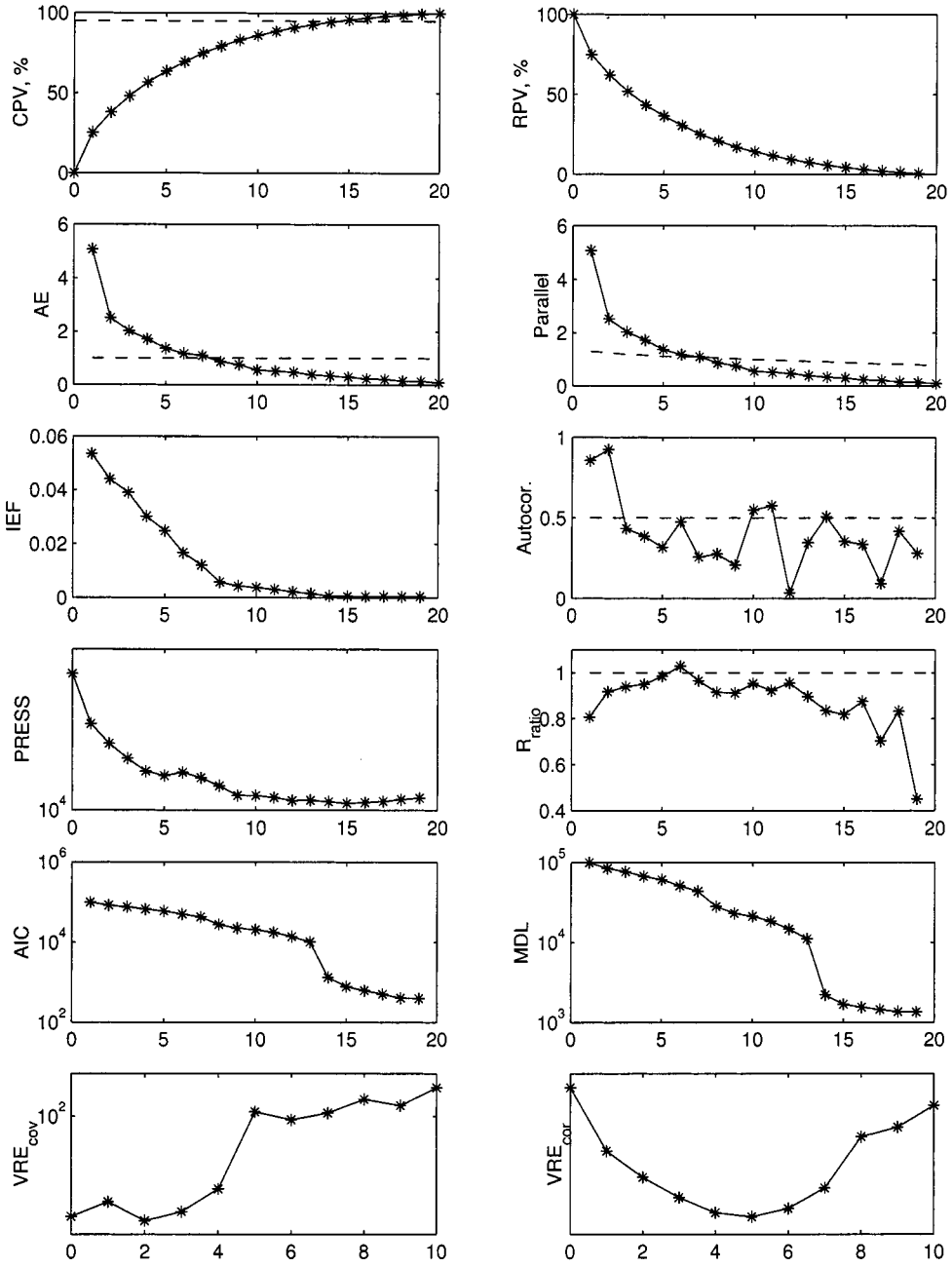

Figure 6. Comparison for a real incineration process.

Table 2. Characteristics for the 11 Methods (× = no; √ = yes; L = Low; M = Medium; H = High)

	CPV	RPV	AE	PA	AC	IEF	R	PRESS	AIC	MDL	VRE
objectiveness	×	×	√	√	×	√	√	√	√	√	√
uniqueness	√	×	√	√	×	√	√	√	√	√	√
covariance-based	√	√	√	√	√	√	√	√	√	√	√
correlation-based	√	√	√	√	√	×	√	√	×	×	√
reliability	√	×	√	√	×	×	×	√	×	×	√
computation	L	L	L	L	L	L	H	M	L	L	L

ance. The results from the 11 methods are shown in Figure 5, with the selected number of PCs shown in Table 1.

It is seen from Figure 5 that, for the CPV method, this process is highly correlated because only with one PC that one is able to capture more than 95% of the variance. Basically, all the methods that found a solution agree with one PC with the exception of the *R ratio* method. The IEF, AIC, and MDL methods are unable to find a minimum for this real data set. The autocorrelation method is also ineffective.

5.3. Incineration Process Data. The incineration data are collected from an industrial waste treatment incinerator, where waste material from different reactions in a process are burned before they are released to the atmosphere. The total analyzed data are 900 samples and 20 variables. The variables included are temperatures, pressures, flows, and concentrations. Figure 6 depicts the results from the 11 methods, and Tables 1 and 2 summarize the selected number of PCs for each method and the characteristics for the 11 methods respectively.

This process is very noisy and, as expected, the number of PCs to represent the process is large. The IEF, AIC, and MDL methods did not find a solution at all. The autocorrelation method is somewhat ambiguous as the autocorrelation of the 10th and 11th PCs become larger than 0.5 after several PCs being smaller than 0.5. The RPV index is monotonically decreasing but fails to show a knee location, making the solution ambiguous. The rest of the methods found a solution. However, the PRESS solution is too high and VRE_{cov} too low for this process. The *R ratio* and VRE_{cor} agree in the same solution.

6. Conclusions

Among the 11 methods presented in the paper, most of them have monotonically decreasing or increasing indices. The IEF, AC, AIC, and MDL worked well with the simulated example, but they failed when used with real data. The most reliable methods are CPV, AE, PA, PRESS, and VRE.

A comparison of the 11 tested methods for the selection of the number of PCs is summarized in Table 2. Considering the effectiveness, reliability, and objectiveness, the PRESS- and correlation-based VRE methods are superior to the others. The VRE method is preferred to the PRESS method in the consistency of the estimate, computational cost, and ability to include a particular disturbance or fault direction in selecting the number of PCs.

Acknowledgment

This work is supported by the National Science Foundation (under Grant CTS-9814340), AMD, Air Products, ALCOA Foundation, DuPont, Union Carbide, Consejo Nacional de Ciencia y Tecnología (CONACyT), and Instituto Tecnológico de Durango (ITD).

Nomenclature

AC = autocorrelation index
$\mathbf{B}$ = model matrix
$\mathbf{C}$ = model projection matrix
C_i = component i concentration
CPV = cumulative percent variance index
f = fault magnitude
$\mathbf{I}$ = identity matrix
IEF = imbedded error function index
k_i = reaction rate constant for component i
l = number of principal components
m = number of variables
M = number of independent parameters
$\mathbf{n}$ = observation noise vector
N = number of samples
$\mathbf{P}$ = loadings matrix
PRESS = predicted error sum of squares index
r_i = reaction rate for component i
R = R ratio index
$\mathbf{R}$ = correlation matrix
RPV = residual percent variance index
RSS = residual sum of squares
E = expectation
R = real
$\mathbf{s}$ = signal sources vector
$\mathbf{S}$ = covariance matrix
S_p = principal components subspace
S_r = residual subspace
$\mathbf{t}$ = score vector
T = temperature
$\mathbf{T}$ = score matrix
u = variance of the reconstruction error
$\mathbf{x}$ = sample vector
$\mathbf{X}$ = data matrix

Greek Letters

λ = eigenvalue
ξ = fault direction
Λ = eigenvalues diagonal matrix
$\|\cdot\|$ = Euclidean norm

Superscripts and Subscripts

$\tilde{\cdot}$ = referred to the residual subspace
$\hat{\cdot}$ = referred to the principal components subspace
$*$ = uncorrupted portion
T = transpose

Abbreviations

AC = autocorrelation
CPV = cumulative percent variance
diag = diagonal
IEF = imbedded error function
PCA = principal components analysis
PCS = principal components subspace
PRESS = predicted sum of squares
RS = residual subspace
RSS = residual sum of squares
PCs = principal components
RPV = residual percent variance
VRE = variance of the reconstruction error

Appendix A

Proof of Theorem 1.

(1) Denoting $\mathbf{e}_{il} = [\epsilon_{il+1}, \epsilon_{i,l+2}, \cdots, \epsilon_{i,m}]^T$, it is clear that $\mathbf{e}_{iq} = [\mathbf{e}_{i,q+1}, \cdots, \mathbf{e}_{i,m}]^T$ for $q > l$. If $l \geq q$, the VRE can be written as

$$u_{il} = \frac{\mathbf{e}_{il}^T \sigma^2 \mathbf{I}_l \mathbf{e}_{il}}{(\mathbf{e}_{il}^T \mathbf{e}_{il})^2} = \frac{\sigma^2}{(\mathbf{e}_{il}^T \mathbf{e}_{il})} = \frac{\sigma^2}{\sum_{j=l+1}^{m} \epsilon_{i,j}^2}$$

which is monotonically increasing with l. Therefore, it is impossible for VRE to have a minimum at $l > q$.

(2) To prove that VRE achieves a global minimum at $l = q$, we just need to show $u_{il} \geq u_{iq}$ for $l < q$. Denoting

$$\Lambda_l = \bar{\Lambda}_l + \sigma^2 \mathbf{I}_l$$

where $\bar{\Lambda}_l = \mathrm{diag}\{\bar{\lambda}_{l+1}, \cdots, \bar{\lambda}_q, 0, \cdots, 0\}$, the VRE in eq 24 can be written as follows:

$$\begin{aligned} u_{il} &= \frac{\mathbf{e}_{il}^T (\bar{\Lambda}_l + \sigma^2 \mathbf{I}_l) \mathbf{e}_{il}}{(\mathbf{e}_{il}^T \mathbf{e}_{il})^2} \\ &= \frac{\mathbf{e}_{il}^T \bar{\Lambda}_l \mathbf{e}_{il}}{\|\mathbf{e}_{il}\|^4} + \frac{\sigma^2}{\|\mathbf{e}_{il}\|^2} \\ &= \frac{\sum_{j=l+1}^{q} \lambda_j \epsilon_{i,j}^2}{\|\mathbf{e}_{il}\|^4} + \frac{\sigma^2}{\|\mathbf{e}_{il}\|^2} \geq \frac{\bar{\lambda}_q \sum_{j=l+1}^{q} \epsilon_{i,j}^2}{\|\mathbf{e}_{il}\|^4} + \frac{\sigma^2}{\|\mathbf{e}_{il}\|^2} \end{aligned} \quad (26)$$

and

$$u_{iq} = \frac{\sigma^2}{||\mathbf{e}_{iq}||^2}$$

To require $u_{il} \geq u_{iq}$, we need to have

$$\bar{\lambda}_q \frac{\sum_{j=l+1}^{q} \epsilon_{i,j}^2}{||\mathbf{e}_{il}||^4} \geq \frac{\sigma^2}{||\mathbf{e}_{iq}||^2} - \frac{\sigma^2}{||\mathbf{e}_{il}||^2} = \frac{\sigma^2 \sum_{j=l+1}^{q} \epsilon_{i,j}^2}{||\mathbf{e}_{iq}||^2 ||\mathbf{e}_{il}||^2}$$

or

$$\bar{\lambda}_q \geq \frac{||\mathbf{e}_{il}||^2}{||\mathbf{e}_{iq}||^2}\sigma^2 \quad \text{QED}$$

Appendix B

Proof of Theorem 2. In the case of unequal noise variances eq 24 still holds. If $l \geq q$,

$$u_{il} = \frac{\mathbf{e}_{il}^{\mathrm{T}} \Lambda_l \mathbf{e}_{il}}{(\mathbf{e}_{il}^{\mathrm{T}} \mathbf{e}_{il})^2} \geq \frac{\mathbf{e}_{il}^{\mathrm{T}} \sigma_m^2 \mathbf{I}_l \mathbf{e}_{il}}{||\mathbf{e}_{il}||^4} = \frac{\sigma_m^2}{||\mathbf{e}_{il}||^2}$$

and

$$u_{iq} = \frac{\mathbf{e}_{iq}^{\mathrm{T}} \Lambda_q \mathbf{e}_{iq}}{(\mathbf{e}_{iq}^{\mathrm{T}} \mathbf{e}_{iq})^2} \leq \frac{\sigma_{q+1}^2}{||\epsilon_{iq}||^2}$$

To require $u_{il} \geq u_{iq}$, we need

$$\frac{\sigma_m^2}{||\mathbf{e}_{il}||^2} \geq \frac{\sigma_{q+1}^2}{||\mathbf{e}_{iq}||^2}$$

or

$$\sigma_m^2 \geq \sigma_{q+1}^2 \frac{||\mathbf{e}_{il}||^2}{||\mathbf{e}_{iq}||^2}$$

If $I < q$,

$$u_{il} = \frac{\sum_{j=l+1}^{q} \lambda_j \epsilon_{i,j}^2}{(\mathbf{e}_{il}^{\mathrm{T}} \mathbf{e}_{il})^2} + \frac{\sum_{k=q+1}^{m} \sigma_k^2 \epsilon_{i,k}^2}{(\mathbf{e}_{il}^{\mathrm{T}} \mathbf{e}_{il})^2}$$

$$u_{iq} = \frac{\sum_{k=q+1}^{m} \sigma_k^2 \epsilon_{i,k}^2}{(\mathbf{e}_{iq}^{\mathrm{T}} \mathbf{e}_{iq})^2}$$

$$u_{il} - u_{iq} = \frac{\sum_{j=l+1}^{q} \lambda_j \epsilon_{i,j}^2}{||\mathbf{e}_{il}||^4} + \sum_{k=q+1}^{m} \sigma_k^2 \epsilon_{i,k}^2 \left[\frac{1}{||\mathbf{e}_{il}||^4} - \frac{1}{||\mathbf{e}_{iq}||^4}\right]$$

$$\geq \frac{\lambda_q \sum_{j=l+1}^{q} \epsilon_{i,j}^2}{||\mathbf{e}_{il}||^4} - \sum_{k=q+1}^{m} \sigma_k^2 \epsilon_{i,k}^2 \frac{||\mathbf{e}_{il}||^4 - ||\mathbf{e}_{iq}||^4}{||\mathbf{e}_{il}||^4 ||\mathbf{e}_{iq}||^4}$$

$$= \frac{\lambda_q \sum_{j=l+1}^{q} \lambda_j \epsilon_{i,j}^2}{||\mathbf{e}_{il}||^4} - \sum_{k=q+1}^{m} \sigma_k^2 \epsilon_{i,k}^2 \frac{\sum_{j=l+1}^{q} \epsilon_{i,j}^2 (||\mathbf{e}_{il}||^2 + ||\mathbf{e}_{iq}||^2)}{||\mathbf{e}_{il}||^4 ||\mathbf{e}_{iq}||^4}$$

$$= \frac{\sum_{j=l+1}^{q} \epsilon_{i,j}^2}{||\mathbf{e}_{il}||^4} \left[\lambda_q - \frac{\sum_{k=q+1}^{m} \sigma_k^2 \epsilon_{i,k}^2 (||\mathbf{e}_{il}||^2 + ||\mathbf{e}_{iq}||^2)}{||\mathbf{e}_{iq}||^4}\right]$$

$$\geq \frac{\sum_{j=l+1}^{q} \epsilon_{i,j}^2}{||\mathbf{e}_{il}||^4} \left[\lambda_q - \frac{\sigma_{q+1}^2 \sum_{k=q+1}^{m} \epsilon_{i,k}^2}{||\mathbf{e}_{iq}||^4}\right]$$

$$= \frac{\sum_{j=l+1}^{q} \epsilon_{i,j}^2}{||\mathbf{e}_{il}||^4} \left[\lambda_q - \sigma_{q+1}^2 \frac{||\mathbf{e}_{il}||^2 + ||\mathbf{e}_{iq}||^2}{||\mathbf{e}_{iq}||^2}\right]$$

To require $u_{il} \geq u_{iq}$ for $l < q$, we need

$$\lambda_q \geq \sigma_{q+1}^2 \left(1 + \frac{||\mathbf{e}_{il}||^2}{||\mathbf{e}_{iq}||^2}\right) \quad \text{QED}$$

Literature Cited

(1) Wax, M.; Kailath, T. Detection of signals by information criteria. *IEEE Trans. Acoust. Speech Signal Process. ASSP-33*, **1985**, 387−392.

(2) Malinowski, E. R. *Factor Analysis in Chemistry*; Wiley-Interscience: New York, 1991.

(3) Wold, S. Cross validatory estimation of the number of components in factor and principal components analysis. *Technometrics* **1978**, *20*, 397−406.

(4) Jackson, J. E. *A User's Guide to Principal Components*; Wiley-Interscience: New York, 1991.

(5) MacGregor, J. F. Statistical process control of multivariate processes. In *IFAC ADCHEM* Proceedings, Kyoto, Japan, 1994.

(6) Nomikos, P.; MacGregor, J. F. Multivariate SPC charts for monitoring batch process. *Technometrics* **1995**, *37* (3), 403−414.

(7) MacGregor, J. F.; Kourti, T. Statistical process control of multivariate processes. *Control Eng. Practice* **1995**, *3* (3), 403−414.

(8) Harmon, J. L.; Duboc, Ph.; Bonvin, D. Factor analytical modeling of biochemical data. *Comput. Chem*. **1995**, *19*, 1287−1300.

(9) Piovoso, M. J.; Kosanovich, K. A.; Pearson, R. K. Monitoring process performance in real-time. In *Proceedings of ACC*, Chicago, 1992, 2359−2363.

(10) Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley and Sons: New York, 1989.

(11) Dunia, R.; Qin, J.; Edgar, T. F.; McAvoy, T. J. Sensor fault identification and reconstruction using principal component analysis. In *13th IFAC World Congress*, San Francisco, 1996; N:259−264.

(12) Dunia, R.; Qin, S. J. A unified geometric approach to process and sensor fault identification. *Comput. Chem*. **1998**, *22*, 927−943.

(13) Ku, W.; Storer, R. H.; Georgakis, C. Disturbance detection and isolation by dynamic principal component analysis. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 179−196.

(14) Akaike, H. Information theory and an extension of the maximum likelihood principle. In *Proceedings 2nd International Symposium on Information Theory*; Petrov and Caski, Eds., 1974; pp 267−281.

(15) Rissanen, J. Modeling by shortest data description. *Automatica* **1978**, *14*, 465−471.

(16) Malinowski, E. R. Determination of the number of factors and the experimental error in a data matrix. *Anal. Chem.* **1977**, *49* (4), 612−617.

(17) Cattell, R. B. The scree test for the number of factors. *Multivariate Behav. Res.* **1966**, April, 245−276.

(18) Rozett, R. W.; Petersen, E. M. Methods of factor analysis of mass spectra. *Anal. Chem.* **1975**, *47* (8), 1301−1308.

(19) Kaiser, H. F. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **1960**, *20* (1), 141−151.

(20) Zwick, W. R.; Velicer, W. F. Comparison of five rules for determining the number of components to retain. *Psychol. Bull.* **1986**, *99* (3), 432−442.

(21) Shrager, R. I.; Hendler, R. W. Titration of individual components in a mixture with resolution of difference spectra, pKs, and redox transitions. *Anal. Chem.* **1982**, *54* (7), 1147−1152.

(22) Carey, R. N.; Wold, N. S.; Westgard, J. O. Principal component analysis: an alternative to referee methods in method comparison studies. *Anal. Chem.* **1975**, *47* (11), 1824−1829.

(23) Osten, D. W. Selection of optimal regression models via cross-validation. *J. Chemom.* **1988**, *2*, 39−48.

(24) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37−52.

(25) Anderson, T. W. Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **1963**, *34*, 122−148.

(26) Horn, J. L. A rationale and test for the number of factors in factor analysis. *Psychometrica* **1965**, *30* (2), 73−77.

(27) Eastment, H. T.; Krzanowski, W. Cross-validatory choice of the number of components from a principal component analysis. *Technometrics* **1982**, *24* (1), 73−77.

(28) MacGregor, J. F. Personal communication, 1998.

(29) Qin, S. J.; Dunia, R. Determining the number of principal components for best reconstruction. In *IFAC DYCOPS'98*, Greece, June 1998.

(30) Li, W.; Yue, H.; Valle, S.; Qin, J. Recursive PCA for adaptive process monitoring. Submited to *J. Process Control* **1999**.

(31) Kramer, M. A. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* **1991**, *37*, 233−243.

(32) Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24* (6), 417−441.

(33) Pearson, K. On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **1901**, series 6, *2* (11), 559−572.

(34) Himes, D. M.; Storer, R. H.; Georgakis, C. Determination of the number of principal components for disturbance detection and isolation. *Proceedings of the American Control Conference*, Baltimore, MD, June 1994; pp 1279−1283.

Received for review February 15, 1999
Revised manuscript received July 28, 1999
Accepted August 2, 1999

IE990110I