Statistical Monitoring of Multivariable Dynamic Processes with State-Space Models

Antoine Negiz and Ali Çinar
Dept. of Chemical and Environmental Engineering, Illinois Institute of Technology, Chicago, IL 60616

Industrial continuous processes may have a large number of process variables and are usually operated for extended periods at fixed operating points under closed-loop control, yielding process measurements that are autocorrelated, cross-correlated, and collinear. A statistical process monitoring (SPM) method based on multivariate statistics and system theory is introduced to monitor the variability of such processes. The statistical model that describes the in-control variability is based on a canonical-variate (CV) state-space model that is an equivalent representation of a vector autoregressive moving-average time-series model. The CV state variables obtained from the state-space model are linear combinations of the past process measurements that explain the variability of the future measurements the most. Because of this distinctive feature, the CV state variables are regarded as the principal dynamic directions A T^2 statistic based on the CV state variables is used for developing an SPM procedure. Simple examples based on simulated data and an experimental application based on a high-temperature short-time milk pasteurization process illustrate advantages of the proposed SPM method.

Introduction

The goal of statistical process monitoring (SPM) is to detect the existence and the time of occurrence of changes that cause a manufacturing environment to deviate from its desired operation. The methodology for detecting changes is based on probability theory and statistical methods that deal with the collection, classification, analysis and interpretation of data. Current SPM methods have limitations for monitoring multivariable continuous dynamic processes. A new approach is proposed for SPM of such processes, and the methodology is illustrated by implementing it to various case studies.

The design of an SPM procedure involves the translation of the desired process operating region into a region of acceptable variation of measurements. This region is defined by using statistical measures that are based on process measurements. If the measurements are within the acceptable region of variation, the process is declared to be in-control; otherwise, it is declared out-of-control. Most SPM methods assume that process measurements obtained under the in-control variation are serially independent and follow an identical parametric distribution (*iid*), usually Gaussian (Montgomery, 1991; Wetherill and Brown, 1991). Several researchers have pointed out the adverse effects of the *iid* assumption on SPM procedures when used with autocorrelated measurements (Montgomery and Mastrangelo, 1991; Harris and Ross, 1991; Negiz and Çinar, 1995). Instead, monitoring of residuals between measurements and their predictions have been proposed. In this approach, the variables are predicted by using dynamic process models. Often these models are empirical time-series models developed from data collected when the process was in-control. The residuals are considered as *iid* if the time-series model captures the autocorrelation structure of the serially dependent process variables. Most residuals-based methods have focused on univariate problems, and therefore obtaining an accurate time-series model for one process variable is thought to be well developed.

It is difficult to apply the residuals-based SPM schemes directly for monitoring continuous chemical processes with large numbers of process variables and processes controlled with feedback control systems. The modeling effort increases significantly with the number of process variables, even if only autoregressive (AR) time-series models are considered for each autocorrelated process variable. Strong autocorrelation

Correspondence concerning this article should be addressed to A. Çinar.
Present address of A. Negiz: UOP Inc., 25 East Algonquin Road., P.O. Box 5017, Des Plaines, IL 60017.

is widely encountered in processes under feedback automatic control. Negiz et al. (1994) applied a residuals-based multivariate SPM scheme to a spray dryer that had only four process variables, among which two were severely autocorrelated due to feedback control. The residuals-based SPM schemes were practically insensitive to even very large level (mean) changes in the process variables. This observation led to the development of an alternative SPM scheme that monitors the changes in models parameters as they are updated by an adaptive recursive estimator. This *parameter change detection* (PCD) method performs significantly better in monitoring level changes of highly positively autocorrelated process variables (Negiz and Çinar, 1994, 1995).

Monitoring the residuals based on time-series models by using Shewhart or cumulative sum (CUSUM) charts and PCD methods are suitable when the process of interest has a small number of variables and moderate levels of cross-correlations. In cases where there are large numbers of process variables, the traditional approach has been to use tools based on multivariate statistical analysis (Alt, 1985). A usual assumption in multivariate statistical analysis is that the vector of observations obtained at each sampling instant are independent observations drawn from an identical multivariate Gaussian distribution. Principal-component analysis (PCA) (Jackson and Mudholkar, 1979; Kresta et al., 1991; Raich and Çinar, 1996), multivariate linear-regression analysis (Hawkins, 1991), and partial least-squares (PLS) analysis (MacGregor et al., 1994; Negiz and Çinar, 1992) are among the popular multivariate techniques used for developing SPM tools for processes with large numbers of variables. Researchers have recently realized the important restrictions imposed by the *iid* assumption when time-dependent observations are monitored as if they are *iid*. Within the *iid* framework, the entire variability of the observation vector can be explained by estimating a zero lag covariance. Naturally the zero lag covariance matrix is not fully descriptive of the entire variation if the vectors of observations are serially correlated.

Extensions of the *iid* multivariate SPM procedures have been proposed for handling serially correlated multivariate observations. PCA (Broomhead and King, 1986; Ku et al., 1995; Nomikos and MacGregor, 1994, 1995a) and PLS techniques (Nomikos and MacGregor, 1995b) have been applied based on a covariance matrix constructed by including lagged values of process variables. Nomikos and MacGregor (1994, 1995a,b) apply lagged versions of PCA and PLS techniques to monitor batch processes. They have built lagged PC and PLS models for describing the variability around an average trajectory that is obtained from combining several successful batch runs together and using all the data available from the beginning to the end of a batch run. Continuous processes provide one-at-a-time observations, which makes the problem of statistical modeling different from the context of batch-process monitoring. PCA on lagged variables have been suggested for developing dynamic models and SPM tools for dynamic continuous processes (Ku et al., 1995; Kourti and MacGregor, 1995; Wold et al., 1984b). Ku et al. (1995) proposed a method to extract the time-series model from the eigenvectors of the covariance matrix that correspond to zero eigenvalues. However, building a lagged version of a covariance matrix in order to extract the PCs requires an extensive search, especially for processes with large numbers of variables (Broomhead and King, 1986). More importantly, the PCs extracted in this fashion are not guaranteed to yield accurate and minimal dynamic representations, as illustrated in Examples 1 and 2.

Multivariate monitoring techniques for serially correlated observations have been used in applications with various degrees of success. However, they do not provide a complete solution to the problem of statistical monitoring of continuous processes with many variables that are serially correlated and simultaneously cross-correlated to each other. A unified methodology is needed to address the issues in multivariable dynamic process modeling, system theory, assumption of normal marginal distributions, and monitoring of multivariate processes.

The objective of this article is to introduce a class of statistical tools that can accurately describe the in-control variability of continuous processes and compute a single statistic to monitor effectively the variability of a process. The statistical model of the in-control variation is in the form of a *state-space model* where statistical assumptions are consistent with expected statistical properties of measurements from continuous processes. The model is developed such that the state variables are statistically independent (orthogonal) at zero lag. One novel feature of the proposed SPM method is the use of state variables for computing the monitoring statistic. Other important features include the option to base the SPM on a significant subset of state variables, and the use of *lambda* distributions to identify the marginal distributions of the residuals and check if they are normal.

The marginal lambda-distribution information based on the one-step-ahead model residuals is used in the model-development stage for detecting outliers in data and correcting the inflated variances of the residuals due to outliers (Negiz, 1995). Since the monitoring procedure is based solely on the CV state variables, the percentile information based on the residual lambda distribution is not used directly in the proposed SPM procedure. However, as will be discussed in detail, the residual marginal distribution information is utilized to provide a basis for setting the control limits on the CV state-space SPM procedure.

The state-space models and the statistical characterization of the random components are obtained directly from process data collected when the process was operating in-control. This approach is known as *stochastic realization*, and it solves the modeling problem. The realization algorithm is suitable for handling a large number of variables that are autocorrelated, cross-correlated, and collinear. The state-space modeling paradigm is well known in the control community, facilitating the appreciation of the proposed SPM methodology. It also provides the opportunity to utilize well-accepted fault-diagnosis paradigms such as parity methods. Furthermore, the state-variable trajectories follow different patterns for different faults, enabling fault diagnosis by direct pattern recognition for a number of critical faults (Negiz, 1995). The fault diagnosis issue is not discussed in this article.

Once an accurate statistical description of the in-control variability of a continuous process is available, the next step is the design and implementation of an on-line SPM procedure that includes information from all measured process variables. The on-line SPM procedure is implemented with a single statistic, the Hotelling T^2, which is computed by using

process data and information from the state-space model of the in-control variation. Since the *state variables* are statistically independent at zero lag because of the modeling procedure utilized, they decompose the T^2 statistic as summations of scalars in the same way that the classic (static) PCA decomposes the T^2 statistic for *iid* observations. The use of CV state variables for SPM also eliminates the limitations of using residuals for SPM (Harris and Ross, 1991; Negiz and Çinar, 1994, 1995).

The organization of this article is as follows. The motivation for the state-space model and its performance relative to the lagged PCA analysis are established first with simple examples. Then, the theoretical background is introduced for the stochastic realization method (including the versatile lambda distributions) that identifies canonical variate state-space models in terms of vector autoregressive moving-average (VARMA) processes. Finally, an application of the proposed stochastic realization algorithm to monitor the in-control variability of a high-temperature short-time (HTST) milk-pasteurization pilot plant with six process variables is described.

Motivating Examples

Two examples are used to emphasize the limitations of PCA for dynamic model development and SPM of processes with autocorrelated data. Example 1 illustrates that an SPM method based on "static" PCA would be indifferent to changes in the dynamics of the system if the zero lag covariance structure is not affected by those changes. Example 2 illustrates the effects of noise in data to time-series model development by using dynamic PCA. The systems in these examples are used later to illustrate the implementation of CV state-space methods and to compare the results.

Example 1. Consider an autocorrelated process with two variables y_1 and y_2 described as

$$y_1(k) = \phi y_1(k-1) + \epsilon_1(k)$$
$$y_2(k) = \sqrt{2}\, y_1(k) + \epsilon_2(k), \qquad (1)$$

where integer k denotes the current time index; and $\epsilon_1(k)$ and $\epsilon_2(k)$ are serially and jointly independent zero mean Gaussian random variables with variances $1-\phi^2$ and 0.5, respectively. The asymptotic variance of $y_1(k)$ is unity based on $1-\phi^2$ variance for $\epsilon_1(k)$. The observations $y_1(k)$ and $y_2(k)$ are autocorrelated and cross-correlated with each other, with the degree of autocorrelation depending on the value of the parameter ϕ. In this illustration, the values 0 and 0.95 are assigned to ϕ. For each value of ϕ, 1,000 pairs of $y_1(k)$ and $y_2(k)$ observations were generated according to Eq. 1 by using two distinct normal random-number-generating functions in MATLAB.

Consider the system with $\phi = 0$. Process measurements are not autocorrelated; however, there is a positive crosscorrelation between $y_1(k)$ and $y_2(k)$ at zero lag. Under these circumstances the observations $y_1(k)$ and $y_2(k)$ are considered as *iid*, and their principal-component structure computed based on the zero lag covariance matrix is shown in Figure 1. PCA analysis is illustrated in two-dimensional spaces. The circle points indicate a representative subset of individual data points out of 1,000 observations; the elliptic regions denote the 95 and 99% confidence region; and PC1 and PC2 are the principal components in the order of decreasing variance. Figure 1a indicates that the zero lag covariance between the two variables can be described with one PC in the direction of PC1. PC2 has very little variability compared to PC1. The lagged covariance matrices (Figure 1b and 1c) does not contribute to the knowledge of the variation between the two variables. This is consistent with the fact that the observations were *iid* by the choice of $\phi = 0$. The principal dimensions (denoted by PC1 and PC2) coincide with the original coordinates.

Consider the system with $\phi = 0.95$. $y_1(k)$ and $y_2(k)$ are both autocorrelated and cross-correlated. The principal-component structure shows that although ϕ is changed from 0 to 0.95, the zero lag covariance structure (Figure 2a) remained the same as in the *iid* case. This implies that a SPM procedure based on the zero lag covariance would be indifferent to changes in the dynamics of the system as long as those changes do not drastically affect the zero lag covariance structure. This illustrates an important shortcoming of classic multivariate SPM tools when applied to dynamic processes. The drastic changes in the lagged covariance matrices (Figure 2b and 2c) cannot be observed from Figure 2a.

A possible remedy to the shortcomings illustrated in Example 1 is to include lagged observations into the covariance matrix and then apply the classic PCA method (Broomhead

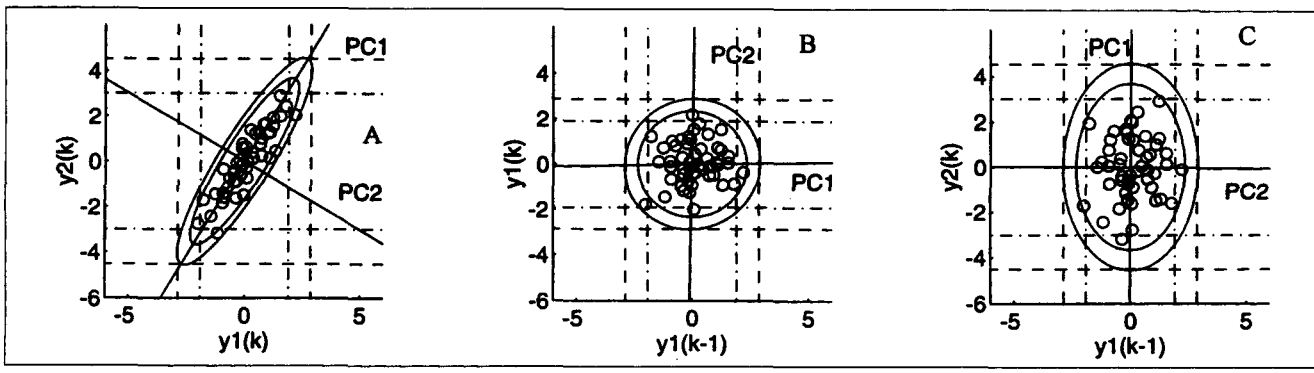

Figure 1. Correlation and principal components loadings for the system described by Eq. 1 for $\phi = 0.0$

(a) Cross-correlation between $y_1(k)$ and $y_2(k)$; (b) autocorrelation between $y_1(k)$ and $y_1(k-1)$; (c) cross-correlation between $y_2(k)$ and $y_1(k-1)$.

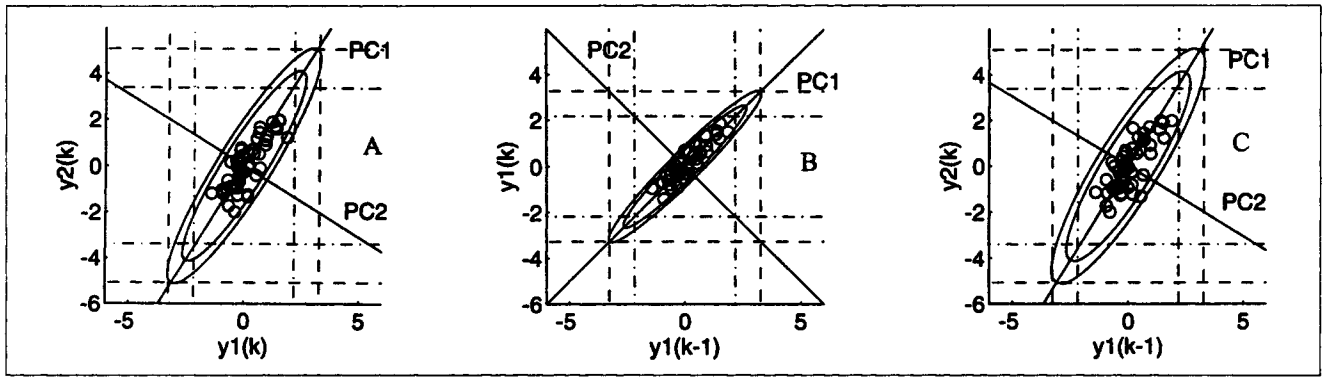

Figure 2. Correlation and principal components loadings for the system described by Eq. 1 for $\phi = 0.95$.

(a) Cross-correlation between $y_1(k)$ and $y_2(k)$; (b) autocorrelation between $y_1(k)$ and $y_1(k-1)$; (c) cross-correlation between $y_2(k)$ and $y_1(k-1)$.

and King, 1986; Ku et al., 1995; Wold et al., 1984b). In theory, if a linear relationship(s) exists between the present observations (at time k) and the lagged (past) observations (at times $k-1$, $k-2$, ...), then the linear combination(s) (the linear dynamic model) can be extracted from the eigenvector(s) with zero eigenvalue(s) (Ku et al., 1995). The eigenvectors corresponding to nonzero eigenvalues are regarded as the principal dimensions of the dynamic variation. Consequently, the magnitudes of the eigenvalues can be used to decide the number of eigenvectors to be retained with significant variation (i.e., nonzero eigenvalues) and to extract the linear dynamic model. Unfortunately, process noise can significantly affect the eigenvectors with small eigenvalues and influence the model development effort, as illustrated in Example 2.

Example 2. A dynamic system with an exogenous input is used to investigate the effect of process noise levels on model development. Consider a two-dimensional linear dynamic system with an exogenous input $u(k)$

$$y_1(k) = 0.95y_1(k-1) - 0.35y_1(k-2) + 0.8u(k-1) + a_1\epsilon_1(k)$$
$$y_2(k) = 0.35y_2(k-1) + u(k-1) + a_2\epsilon_2(k), \qquad (2)$$

where $\epsilon_1(k)$ and $\epsilon_2(k)$ are serially and jointly independent zero-mean unit variance Gaussian random variables that represent additive process noise. Parameters a_1 and a_2 are used to adjust the standard deviations of $\epsilon_1(k)$ and $\epsilon_2(k)$, respectively. The exogenous input $u(k)$ is independent of $\epsilon_1(k)$ and $\epsilon_2(k)$ and it is varied as a serially independent zero-mean unit variance Gaussian process to provide persistent excitation. Equation 2 represents a third-order dynamic system, since its minimal state-space representation requires three state variables. The system has three poles (in z-domain) inside the unit circle at $z = 0.4750 \pm 0.3527i$, 0.3500.

One thousand data points were generated at five different process noise levels. For the case of zero process noise (i.e., a_1 and a_2 are set to zero), the standard deviation of $y_1(k)$ and $y_2(k)$ are 1.2 and 1.1, respectively. A percent noise-to-signal (%NTS) parameter is defined as the percent of a_1 and a_2 relative to the standard deviations at zero noise. Data are used to compute a covariance matrix of the variables, which include the present observations ($y_1(k)$ and $y_2(k)$), and the past lagged outputs and inputs of order 2 [$y_1(k-1)$, $y_2(k-1)$, $u(k-1)$, $y_1(k-2)$, $y_2(k-2)$, $u(k-2)$]. The modeling effort involves the search for the maximum lag to be included in the model. But since Eq. 2 is given in this example, an extensive search is avoided by assuming the correct maximum lag.

The first six eigenvalues of the 8×8 covariance matrix are listed in the order of decreasing magnitude in the first row of Table 1 for 0 %NTS level. The remaining two eigenvalues are practically zero (2.085×10^{-15}, -7.4153×10^{-16}). By using the time-series model-development method of Ku et al. (1995), Eq. 2 is extracted from the eigenvectors of the covariance matrix corresponding to zero eigenvalues. Note that although the system can be represented using three dimensions (minimal state-space representation of order 3), the eigenvalues do not drop to zero after the first three dynamic PCA directions. In contrast, the canonical-variate (CV) state-space realization method indicates the correct number of minimal state variables, as discussed in detail later. The singular values drop to zero right after the first three CV state variables for the zero-noise case (Table 1). As the %NTS increases, the magnitudes of the eigenvalues increase as well. This is expected, since the eigenvalues can also be interpreted as the variances of the PC directions.

Table 2 gives the model coefficients obtained by using the two eigenvector-eigenvalue pairs with the minimum eigenvalues as a function of %NTS. The level of process noise adversely affects the estimates of the model parameters. Deterioration in parameter accuracy is significantly higher for PCA than CV analysis. As shown later, the accuracy of model parameters obtained from PCA at 0.05% noise level is worse than the CV model accuracy at 10% noise level (Table 3).

A comparison of PLS and CV state space realization for constructing dynamic process models would be informative. Table 4 lists the percentage of variance explained in **X** and **Y**

Table 1. Eigenvalues of Covariance Matrix or Singular Values of Hankel Matrix for 0% NTS

Eigenvalues of covariance matrix for DPCA	5.068	2.396	1.892	0.1682	0.014	6.391×10^{-5}
Singular values of scaled Hankel matrix for CV state-space modeling	1.000	1.000	1.000	0	0	0

Table 2. Parameters of Models Based on Dynamic PCA Methodology for Various Percent Noise to Signal (%NTS) Levels*

%NTS	Variable	$y_1(k-1)$	$y_2(k-1)$	$u(k-1)$	$y_1(k-2)$	$y_2(k-2)$	$u(k-2)$
0	$y_1(k)$	0.95	0	0.8	−0.35	0	0
0	$y_2(k)$	0	0.35	1	0	0	0
0.05	$y_1(k)$	0.9678	−0.081	0.8007	−0.3549	0.0194	0.067
0.05	$y_2(k)$	−0.0039	0.4416	0.9996	0.0031	−0.0328	−0.088
0.5	$y_1(k)$	0.8553	−2.11	0.7957	−0.3295	0.8012	2.1866
0.5	$y_2(k)$	0.1921	2.7318	1.0081	−0.0437	−0.9534	−2.534
5	$y_1(k)$	19.9366	66.879	1.328	−7.8583	−32.067	−83.41
5	$y_2(k)$	−21.0083	−74.31	44.216	8.3054	35.722	92.979
10	$y_1(k)$	4.4231	9.4328	0.9134	−2.009	−4.6165	−12.63
10	$y_2(k)$	−3.6331	−10.08	0.938	1.7313	5.045	13.823

*Entries in the second column are the variables estimated by the model; entries in all succeeding columns are the parameters of terms with regressor specified in the first rows, respectively.

by each PLS component, and the cumulative variances explained. Six PLS components which are linear combinations of the past information that explain the present values are necessary for describing all the variation in the data. Although the last few components have marginal contributions, they are not equal to zero. The PLS approach is doing well in finding a one-step-ahead prediction model, but 6 PLS components are needed even though the minimal order of the system is 3. The one-step-ahead PLS model requires more linear combinations of the past than needed (three) because the regression is performed by using only information from the present and the past, information in the future has not been included. In the CV state space modeling approach if the future observation window (J defined in Eq. 7) was set equal to 1, then the rank of the Hankel matrix is limited to 2. Therefore, it is necessary to make $J > 1$ for identifying the minimal order of the system correctly. This comparison does not mean that the PLS approach can not be used for identifying the minimal states as the CV analysis does. Negiz and Çinar (1996) show that it is possible to obtain the minimal states by PLS by adding a nonsingular transformation step to the classical PLS algorithm.

Example 1 illustrated the shortcomings of classical *iid* (zero lag) PCA applied to autocorrelated process measurements. Example 2 shows that modeling and parameter estimation by dynamic PCA is severely affected by process noise. Furthermore, the number of dynamic principal dimensions which have nonzero eigenvalues are not the minimum dimensions of a linear dynamic system. In the following, an alternative model development paradigm is proposed by using the CV state space identification method for constructing SPM tools for continuous processes with autocorrelated data. The n dimensional CV state obtained by canonical correlation analysis between the future and the past observation sequences gives the n linear combinations of the past which are maximally correlated to the future observations. At any given time the CV state variables are the minimal memory of the past that can describe the future of the process not just for one step ahead, but for the whole length of the future observations window (denoted by J in the derivations that follow). Furthermore, the CV state variables are orthogonal at zero lag. Consequently, their $n \times n$ zero lag covariance matrix (denoted as Σ) is diagonal with positive (nonzero) elements and the state covariance matrix is nonsingular.

Canonical Variate Stochastic Realization Procedure

The first step in building the SPM system for multivariable continuous processes is to develop an accurate process model. Often, detailed models based on first principles are not available and empirical dynamic models must be developed by using input-output data. Several model identification methods which generate polynomial models such as VARMA models are discussed in system identification literature (Wei, 1990; Lütkepohl, 1991). Such models need *a priori* knowledge of model order and of observability or controllability indices to start up the identification effort (Ljung, 1987). Generating this *a priori* knowledge is not trivial for high-order, multivariable processes, and also numerically ill-conditioned mathematical problems can be formulated (Van Overschee and de Moor, 1994). *Subspace algorithms* offer a viable alternative (Larimore, 1983; Swindlehurst et al., 1995; Van Overschee and de Moor, 1994; Verhaegen and Dewilde, 1992). They

Table 3. Parameters of Models Based on CV State-Space Methodology for Various Percent Noise to Signal (%NTS) Levels*

%NTS	Variable	$y_1(k-1)$	$y_2(k-1)$	$u(k-1)$	$y_1(k-2)$	$y_2(k-2)$	$u(k-2)$
0	$y_1(k)$	0.95	0	0.8	−0.35	0	0
0	$y_2(k)$	0	0.35	1	0	0	0
0.05	$y_1(k)$	0.951	−0.005	0.8008	−0.3497	0.0015	0.0045
0.05	$y_2(k)$	0.0042	0.3228	0.9992	0.0005	0.0048	0.0242
0.5	$y_1(k)$	0.9527	0.0234	0.8026	−0.3489	−0.0086	−0.024
0.5	$y_2(k)$	0.0062	0.3178	0.9974	0.004	0	0.0284
5	$y_1(k)$	0.9553	0.0397	0.8085	−0.3461	−0.0127	−0.039
5	$y_2(k)$	−0.0079	0.3219	0.9918	0.0203	−0.0106	0.0381
10	$y_1(k)$	0.9556	0.0421	0.8122	−0.3446	−0.0111	−0.04
10	$y_2(k)$	−0.0146	0.3222	0.9884	0.0276	−0.0145	0.0449

*Entries in the second column are the variable estimated by the model; entries in all succeeding columns are the parameters of terms with regressor specified in the first rows, respectively.

Table 4. Percent Variance Explained in X and Y, and Cumulative Variances Explained

PLS Comp.	% Variance Explained		Cumulative Variance Explained	
	X	Y	X	Y
1	50.184	61.012	50.184	61.012
2	20.418	37.439	70.602	98.451
3	26.991	0.818	97.593	99.269
4	2.271	0.491	99.864	99.760
5	0.135	0.240	99.999	100.000
6	0.0009	0.0001	100.000	100.000

generate the process model by successive approximation of the memory or the state variables of the process by determining successively functions of the past that have the most information for predicting the future (Larimore, 1990). Most of the *a-priori* parametrization need in traditional system identification methods is avoided. Only the system order is needed and it is determined by inspecting the dominant singular values of a covariance matrix generated by singular value decomposition. Since the model-development algorithm is not iterative, it provides a significant computational advantage over VARMA modeling methods that require structure selection and use maximum-likelihood parameter estimations. In this work one class of suboptimal identification algorithms, called the canonical-variates analysis (CVA), is used. The philosophy of CVA has many common features to PCA and PLS, in that they use covariance and cross-covariance information to determine the number of variables for building the model (Schaper et al., 1994). Classic PLS maximizes cross-covariance and uses a sequential procedure for selecting the important latent variables (Hoskuldsson, 1988; Wold et al., 1984a). More recent versions of PLS use singular-value decomposition (SVD) to determine the appropriate number of latent variables (Kaspar and Ray, 1993a,b; Negiz and Çinar, 1996). CVA maximizes cross-correlation and uses SVD (Larimore, 1990). Canonical-correlation analysis has been used in the statistics community for identifying time-series models in the discrete-time domain. The canonical-correlation analysis performed maximizes the predictability of the *present* process output(s) from the filtered past output(s), but the use of future windows of outputs has not been considered (Box and Tiao, 1977; Tsay and Tiao, 1985). Consequently, this canonical-correlation approach is different from the canonical-variate analysis method used in our study. The canonical-variate analysis proposed in our study focuses on canonical correlations between the past process outputs and future process outputs.

Stochastic realization

Let $y_k \in \Re^p$ represent the values of p process outputs (measurement variables) at sampling instant k. The process measurements are obtained under conditions when the process is assumed to be in-control, operating with acceptable levels of variation around process output targets. Under these circumstances, it can be assumed that the AR part of a vector time series is time invariant, weakly stationary (may have poles close to or on the unit circle), and its moving average (MA) part is of minimum phase (invertible). These conditions are referred to as *regularity conditions*. Based on regularity conditions, a general state-space representation of a VARMA process can be given as (Aoki, 1990)

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{w}_k \quad \text{with} \quad \boldsymbol{w}_k = \boldsymbol{B}\boldsymbol{\epsilon}_k \\ \boldsymbol{y}_k &= \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{\epsilon}_k, \end{aligned} \tag{3}$$

where $\boldsymbol{x}_k \in \Re^n$ is the state vector; $\boldsymbol{w}_k = \boldsymbol{B}\boldsymbol{\epsilon}_k$ is a stochastic input vector referred to as a state disturbance; $\boldsymbol{\epsilon}_k$ is another stochastic input vector referred to as the measurement noise; $\boldsymbol{A}$ is an $n \times n$ time-invariant (constant) matrix; and the output matrix $\boldsymbol{C}$ is $p \times n$ and is also known as the observation matrix. Consistent with the assumptions of time-series modeling (Box and Jenkins, 1976), it is assumed that $E(\boldsymbol{\epsilon}_k) = \boldsymbol{0}$, $E(\boldsymbol{w}_k) = \boldsymbol{0}$, and

$$E(\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_{k+l}^T) = \begin{cases} \boldsymbol{\Delta} & \text{if } l = 0 \\ \boldsymbol{0} & \text{otherwise} \end{cases} \tag{4}$$

$$E(\boldsymbol{w}_k \boldsymbol{w}_{k+l}^T) = \begin{cases} \boldsymbol{W} & \text{if } l = 0 \\ \boldsymbol{0} & \text{otherwise}, \end{cases} \tag{5}$$

where $E(\cdot)$ denotes the expectation operator. Equations 4 and 5 indicate that the random vectors $\boldsymbol{w}_k$ and $\boldsymbol{\epsilon}_k$ are serially uncorrelated except at the zero lag ($l = 0$) with $\boldsymbol{W}$ and $\boldsymbol{\Delta}$ as their covariance matrices. The covariance of the random vectors is

$$E\left\{\begin{bmatrix} \boldsymbol{w}_k \\ \boldsymbol{\epsilon}_k \end{bmatrix} [\boldsymbol{w}_k \quad \boldsymbol{\epsilon}_k]^T\right\} = \begin{bmatrix} \boldsymbol{B}\boldsymbol{\Delta}\boldsymbol{B}^T & \boldsymbol{B}\boldsymbol{\Delta} \\ \boldsymbol{\Delta}\boldsymbol{B}^T & \boldsymbol{\Delta} \end{bmatrix}. \tag{6}$$

Equations 3–6 are referred to as the *data-generating process* from which process measurements $\boldsymbol{y}_k$ are observed at each instant k.

The Stochastic Realization Problem. Construct a state-space model by estimating the system matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{\Delta}$, and identify the marginal distributions of the elements of $\boldsymbol{\epsilon}_k$, given a set of measurements of $\boldsymbol{y}_k$.

Assume that the data-generating process follows the general state-space form given in Eqs. 3–6, and that the regularity conditions hold. Define a truncated form of the infinite Hankel matrix from the future ($\mathcal{Y}_{k_J}^+$) and past ($\mathcal{Y}_{k-1_K}^-$) stacked measurement vectors as

$$\begin{aligned} \boldsymbol{H}_{JK} &= E\left(\mathcal{Y}_{k_J}^+ \mathcal{Y}_{k-1_K}^{-T}\right) \\ &= E\left\{\begin{bmatrix} \boldsymbol{y}_k \\ \vdots \\ \boldsymbol{y}_{k+J-1} \end{bmatrix} [\boldsymbol{y}_{k-1}^T \quad \cdots \quad \boldsymbol{y}_{k-K}^T]\right\}. \end{aligned} \tag{7}$$

The superscripts + and − distinguish between the future and past stacked measurement vectors, and the superscript T denotes the transpose operation. Each block entry of the Hankel matrix describes the autocovariance structure of the process measurements ($\boldsymbol{y}_k$) at various lags. The length of the future and past observation windows are represented by J and K, respectively. These integers must be chosen sufficiently large ($J > n$) so that the truncated Hankel matrix be-

comes a good approximation to the infinite Hankel matrix, which has all zero elements after a certain J and K due to regularity conditions. The value of K is selected based on univariate autoregressive model analysis for each of the variables. K is chosen to be the maximum significant lag, after which the autocorrelation of autoregressive model residuals of all the variables becomes insignificant. In the information theoretic approach (Candy et al., 1979; Larimore, 1993) there is no penalty to the minimization criterion in terms of J or K. The parametrization penalty term includes the number of inputs, outputs, and the order of the system (number of state variables n). Since the number of state variables is obtained from the rank of the truncated Hankel matrix, $J \geq K$ in order to avoid a rank-restricted Hankel matrix. Otherwise, if the true order of the system (number of state variables) is greater than Jp, despite a sufficiently large K, the order of the system can be misidentified as Jp.

Define the state z_k and the random error sequences (innovation vector) e_k as

$$z_k = E(x_k|\mathcal{Y}^-_{k-1_K}), \quad e_k = y_k - E(y_k|\mathcal{Y}^-_{k-1_K}), \tag{8}$$

where z_k is the best linear estimate of x_k given the past measurements $\mathcal{Y}^-_{k-1_K}$, and $E(y_k|\mathcal{Y}^-_{k-1_K})$ is the best linear estimator of y_k based on $\mathcal{Y}^-_{k-1_K}$. The innovation process (z_k, e_k) is a realization of the dynamic system given in Eq. 3 (Anderson and Moore, 1979; Aoki, 1990; Negiz, 1995). The innovation model is

$$\begin{aligned} z_{k+1} &= Az_k + Be_k \\ y_k &= Cz_k + e_k. \end{aligned} \tag{9}$$

The cross-covariance matrix M between the conditional state z_k and the measurement vector y_{k-1} can be expressed in terms of the system matrices in Eq. 9 as

$$M = E(z_k y^T_{k-1}) = AE(z_{k-1}z^T_{k-1})C^T + B\Delta. \tag{10}$$

The reachability pair in this case is M and A. Based on regularity conditions, the Hankel matrix can also be expressed as (Aoki, 1990; Negiz, 1995)

$$H_{JK} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{J-1} \end{bmatrix} [M \; AM \; \cdots \; A^{K-1}M] = \mathcal{O}_J \Omega_K, \tag{11}$$

where $\mathcal{O}_J$ and Ω_K are the truncated observability and reachability matrices, and (A, M) is the reachability pair. Equation 11 is a direct consequence of the fact that the innovation model is an equivalent realization of the data-generating model (Eqs. 3–6) provided that the regularity conditions are satisfied. The state-space model (Eq. 9) is minimal when both the truncated form of the reachability and observability matrices (J and K replaced with n in Eq. 11) have rank n (Åstrom and Wittenmark, 1990; Maciejowski, 1989).

Combining the main result (Eq. 9) with the preceding definitions for the Hankel matrix, the conditional state is expressed as

$$z_k = \Omega_K (R^-_K)^{-1} \mathcal{Y}^-_{k-1_K}, \tag{12}$$

where $R^-_K = E(\mathcal{Y}^-_{k-1_K} \mathcal{Y}^{-T}_{k-1_K})$. The derivation of Eq. 12 is given in Appendix A. Equation 12 indicates that the definition of the state vector is coordinate system dependent through Ω_K. The decomposition of the Hankel matrix into the observability and reachability matrices allows one to compute the state vectors through Eq. 12. The coordinate system is based on the type of decomposition performed on the Hankel matrix. Note that the innovation model is constructed by using only the observation sequence y_k. Several realization techniques, balanced realization (Aoki, 1990), PLS realization (Negiz, 1995), and the CV realization, which is described in the following section, are available. These are members of a class of identification procedures known as subspace algorithms (Larimore, 1990; Swindlehurst et al., 1995; Van Overschee and de Moor, 1994; Verhaegen and Dewilde, 1992).

Estimation of A, C, B, and Δ by canonical variates

Canonical variate analysis was introduced by Hotelling (1936), and it is discussed in various texts on multivariate statistical analysis (Johnson and Wichern, 1992). The application of CV to dynamic system modeling was proposed by Akaike (1974a,b). An analysis of the CV realization when there are no exogenous inputs present in the model and the theoretical derivations of the CV realization by a *direct substitution* are given by Desai et al. (1985). Schaper et al. (1994) applied the CV analysis for modeling chemical processes with exogenous inputs, based on the works of Larimore (1983, 1993). The *least-squares solution* interpretation given in this section is due to Larimore (1983).

The CV realization requires that the covariances of the future and past stacked observations be conditioned against any singularities by taking their square roots. The Hankel matrix is scaled by using $R^-_K = E(\mathcal{Y}^-_{k-1_K} \mathcal{Y}^{-T}_{k-1_K})$ and $R^+_J = E(\mathcal{Y}^+_{k_J} \mathcal{Y}^{+T}_{k_J})$. The CV decomposition of the scaled Hankel matrix $\bar{H}_{JK}$ yields

$$(R^+_J)^{-1/2} H_{JK} (R^-_K)^{-1/2} = \bar{H}_{JK} = U\Sigma V^T, \tag{13}$$

where the $pJ \times n$ matrix U contains the n left eigenvectors of the decomposition; the $n \times n$ diagonal matrix Σ contains the singular values of the decomposition; and the $Kp \times n$ matrix V contains the n right eigenvectors of the decomposition. The SVD matrices in Eq. 13 include only the singular values and eigenvectors corresponding to the n state variables retained in the model. The full singular-value matrix Σ is $Jp \times Kp$, and it contains the singular values of the decomposition in a descending order. If process noise is small, all singular values smaller than the nth singular value are effectively zero, and therefore the corresponding state variables are excluded from the model. Several methods have been proposed for selecting the value of n. The ratio of the specific singular value to the sum of all the singular values provides a popular selection criterion (Aoki, 1990). An information theoretic approach, such as the Akaike information criterion (AIC), is also used frequently (Candy et al., 1979; Larimore, 1993). Both the inspection of the relative magnitude of singular values and the

AIC criteria have been used in our studies for selecting n. While the *ad hoc* nature of the former approach may cause some arbitrariness in order selection, both approaches have given similar results in our studies.

The state vectors based on the CV decomposition are constructed as

$$z_k = \Sigma^{1/2} V^T (R_K^-)^{-1/2} \mathcal{Y}_{k-1_K}^- \tag{14}$$

by substituting $\Omega_K = \Sigma^{1/2} V^T (R_K^-)^{1/2}$ in Eq. 12, based on Eqs. 11 and 13. Once the state-vector sequence is known, a multiple least-squares solution for A and C in the form of Eq. 9 will yield

$$\begin{bmatrix} \hat{A} \\ \hat{C} \end{bmatrix} = E\left\{ \begin{bmatrix} z_{k+1} \\ y_k \end{bmatrix} [z_k]^T \right\} E\{[z_k \;\; z_k^T]\}^{-1}. \tag{15}$$

The prediction errors for the state and measurements are expressed as

$$\begin{aligned} \tilde{z}_{k+1} &= z_{k+1} - \hat{A} z_k \\ e_k = \tilde{y}_k &= y_k - \hat{C} z_k. \end{aligned} \tag{16}$$

By comparing to Eq. 6 the covariance of the prediction errors reveals that

$$E\left(\begin{bmatrix} \tilde{z}_{k+1} \\ \tilde{y}_k \end{bmatrix} [\tilde{z}_{k+1}^T \;\; \tilde{y}_k^T] \right) = \begin{bmatrix} \mathcal{E}_{11} & \mathcal{E}_{12} \\ \mathcal{E}_{21} & \mathcal{E}_{22} \end{bmatrix} = \begin{bmatrix} \hat{B}\hat{\Delta}\hat{B}^T & \hat{B}\hat{\Delta} \\ \hat{\Delta}\hat{B}^T & \hat{\Delta} \end{bmatrix}. \tag{17}$$

It follows from Eq. 17 that $\hat{\Delta} = \mathcal{E}_{22}$ and $\hat{B} = \mathcal{E}_{12}\mathcal{E}_{22}^{-1}$.

The covariance matrix of the conditional state vector based on CV decomposition is

$$E(z_k z_k^T) = \Sigma, \tag{18}$$

which is also equal to the covariance matrix of the future and past canonical variates (Negiz, 1995). Since the canonical variates have identity covariances, the singular values in the diagonal matrix Σ are also referred to as the correlation coefficients of the future and past canonical variates. Equation 18 reveals that the conditional state variables in the CV realization are independent at zero lag. This property also guarantees that the least-squares solution (Eq. 15) to the system matrices exists and that resulting estimates are stable. The inverse of the state covariance matrix is diagonal and is made numerically stable by selecting only the n state variables corresponding to nonzero singular values. Furthermore, applying CV analysis between the future and the past information gives the optimal predictors (Yohai and Garcia Ben, 1980). This property is lacking in models based on PCA.

Lambda distributions

Fitting marginal distributions to the elements of the measurement-error vector will identify the statistical distributions of the random components in the model. The marginal distributions provide accurate statistical descriptions for hypothesis-testing procedures. In systems-theory applications the errors are assumed to follow a multivariate Gaussian distribution. This implies that the marginal distributions are Gaussian (Johnson and Wichern, 1992). If the marginal distributions of measurement errors are not Gaussian, statistical inference from models residuals may deteriorate. By identifying the proper marginal distributions for the random errors e_k, the stochastic realization for a VARMA model is determined such that accurate hypothesis-testing procedures can be used.

The marginal distributions of the measurement errors are identified by using a family of distributions known as the *lambda* distributions first introduced by Tukey (1962). The lambda random variable Z is defined through a transformation

$$Z = \frac{[U^{\lambda_D} - (1-U)^{\lambda_D}]}{\lambda_D} = G(U), \tag{19}$$

where U is a uniformly distributed random variable on [0, 1], and λ_D is the lambda parameter. A useful property of random variable Z is that for a specific value of λ_D, its quantile of order u, where u is on [0, 1], is computed from Eq. 19 by inserting u for U (Filliben, 1969; Joiner and Rosenblatt, 1971). A detailed analysis of the statistical properties of lambda distributions is also given in Negiz (1995). The random variable Z is distributed approximately as normal for $\lambda_D = 0.135$. For $-1 < \lambda_D < 1$, Z has a unimodal distribution with various tail lengths that can also accommodate the t distributions with various degrees of freedom (Filliben, 1969). For $1 < \lambda_D < 2$, the distribution of Z is U-shaped, while for $\lambda_D > 2$, they are peaked and truncated. For $\lambda_D = 0$ it is logistic, and for $\lambda_D = 1$ or $\lambda_D = 2$ it is rectangular (uniform).

The task is to fit a lambda distribution to a set of univariate prediction errors obtained from a state-space model. The length of the data set is N. The expected value of the errors is zero if an appropriate model is used to generate the residuals. The first step is to order the residuals. The next step is to generate the N uniform (0,1)-order statistic medians. This is done by the formula given in Filliben (1975). The formula gives the N percentile points for which the corresponding quantiles are computed using Eq. 19 for a specific value of λ_D. A correlation coefficient is computed between the λ_D quantiles and the ordered residuals. This procedure is repeated for various λ_D values that cover a reasonable range. The λ_D value that corresponds to the correlation coefficient closest to 1 is selected. The λ_D range used in this study is between -2 and 6.

Application of CV to Examples 1 and 2

Example 3. Consider the system given in Example 1 by Eq. 1 with $\phi = 0.95$. The data used in Example 1 are used to identify the system given in Eq. 1.

First, develop a state-space model by analytical derivation by defining the state as $x_k = y_1(k) - \epsilon_1(k)$, which is equivalent to $x_k = \phi y_1(k-1)$. Rewriting Eq. 1 in state-space form as

$$x_{k+1} = \phi x_k + \phi \epsilon_1(k)$$

$$\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix} x_k + \begin{bmatrix} 1 & 0 \\ \sqrt{2} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_1(k) \\ \epsilon_2(k) \end{bmatrix} \tag{20}$$

reveals that Eq. 20 is the same as Eq. 1. The order of the system is 1, since one state variable is sufficient to describe the system. Note that the state is a function of the past $[y_1(k-1)]$.

Now apply the CV state-space identification method based only on the observation sequence to illustrate how this approach can identify the single minimal state and eventually the model. To obtain a CV state-space realization of the sequence $y_1(k)$ and $y_2(k)$, J and K have to be specified; they are both set at 1. The Hankel matrix becomes 2×2, therefore we shall see how the CV state space will eliminate one dimension to obtain the true order of the system, which is 1. The 2×2 Hankel matrix is constructed according to Eq. 7. The CV decomposition is performed on the scaled Hankel matrix as described in Eq. 13 and the conditional state variables are computed via Eq. 14.

The singular values of the scaled Hankel matrix are 0.9620 and 0.0019, which suggests that the system is effectively described with a single state variable. Therefore setting $n = 1$ for the dimensions of the SVD in Eq. 13, the single conditional state variable is

$$z_k = [0.8929 \;\; 0.0048] \begin{bmatrix} y_1(k-1) \\ y_2(k-1) \end{bmatrix} \tag{21}$$

based on Eq. 14. The variance of the single state is also 0.9620 because of Eq. 18. The corresponding future canonical variate direction is

$$[0.8870 \;\; 0.0091] \begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix}, \tag{22}$$

which is computed by using an equation similar to Eq. 14 by replacing V with U and matrices related to the past with their future counterparts (Negiz, 1995). The conditional state and the corresponding future canonical variate indicate that $y_1(k)$ is the principal component that correlates the past and future observations the most. This is also evident from the structure of Eq. 1. The conditional state pinpoints $y_1(k-1)$ as the only direction in the past that is maximally related to the future. The conditional state identified is consistent with the exact state-space model (Eq. 20).

Since the state variables are known, the least-squares solution provides the estimates for the system matrices. Using Eqs. 9 and 21, the time-series model is obtained by multiplying the 2×1 $\hat{C}$ with Eq. 21. The result is

$$\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \begin{bmatrix} 0.9544 & 0.0051 \\ 1.3596 & 0.0073 \end{bmatrix} \begin{bmatrix} y_1(k-1) \\ y_2(k-1) \end{bmatrix}, \tag{23}$$

which is very close to the true dynamic system (Eq. 1). The slight discrepancies in the model parameters are mostly due to random noise. This example provide the basis for the definition of the principal directions of a *dynamic* system. They are the number of linear combinations of the past information, which describes the future information maximally.

Example 4. Recall that the system described by Eq. 2 in Example 2 requires only three state variables for its minimal state-space representation. The dynamic PCA analysis presented in Example 2 indicated that even at 0 %NTS, the eigenvalues of the PCA do not show distinctively 0 eigenvalues after the third principal component.

To apply the CV analysis let $K = 2$ and $J = 3$. This yields a 6×6 Hankel matrix. Although $K = 2$, the size of the past information vector is 6, since two additional terms are added to include the exogenous input $u(k-1)$ and $u(k-2)$. Only process outputs are used for constructing the future vector. The CV state-space realization as outlined here for VARMA processes needs some modifications when exogenous inputs are present. The details of the algorithm in the presence of the exogenous inputs is described by Larimore (1983, 1990).

As before, Σ is the diagonal matrix that contains the variances of the orthogonal state variables computed via SVD of the Hankel matrix. The second row of Table 1 lists the 6 singular values of the Hankel matrix at 0 %NTS. The first three singular values are equal to 1, and all the remaining singular values are zero. The system order is identified directly by looking at the cutoff between nonzero and zero singular values. This cutoff still prevails as the %NTS increases, but the magnitudes of the last three singular values increase, and the distinction between the third and fourth (from the left) singular value tends to blur. But the dynamically significant directions are not confounded with noise as in the case of PCA. The AIC is consistent with this cutoff point. For example, at 10 %NTS the AIC values as a function of the number of state variables in the model are -3213.6, -4333.2, -4372.9, -4365.0, and -4357.0, respectively. The minimum at -4372.9 indicates three state variables.

Table 3 provides the model equations identified by the CV realization for $n = 3$. The CV model parameters obtained at 10 %NTS are much better than the PCA model parameters identified at 0.05 %NTS given in Table 2. The coefficients of the terms that should be identified as insignificant are on the order of 10^{-2} for the CV analysis at all noise levels. Table 5 lists the poles of the system that are computed as the eigenvalues of the CV state-transition matrix A at various noise levels. Note that this information cannot be extracted using the PCA approach. The poles are estimated very consistently at all noise levels.

The CV approach with $J = 1$, which gives a 2×6 Hankel matrix, cannot identify the third minimal state. This information is only possible when $J = 2$ or higher; in our example we select $J = 3$, to provide the Hankel matrix a row dimensionality that is at least equal to the column dimensionality (dimension of the past).

Table 5. Poles of Identified CV State-Space Model at Various Noise Levels

NTS %	Poles of Identified CV State-Space Model		
0	$0.4750+0.3527i$	$0.4750-0.3527i$	0.35
0.05	$0.4759+0.3512i$	$0.4759-0.3512i$	0.3528
0.5	$0.4784+0.3471i$	$0.4784-0.3471i$	0.3607
5	$0.4812+0.3389i$	$0.4812-0.3389i$	0.3704
10	$0.4805+0.3366i$	$0.4805-0.3366i$	0.3642

The traditional way of model building by regressing the *present* on the past information does not have strong theoretical grounds. This example illustrates how the future information is essential in identifying the overall dynamics of the process.

A traditional regression approach (standard linear regression, ridge regression, PLS regression), which attempts to model the future outputs with a window length of $J = 3$ by regressing it on the past outputs with a window length of $K = 2$, will result in three different sets of model parameters in order to map the past to the three future outputs. This is not desirable, and furthermore does not have a theoretical justification since the number of linear relations to describe a linear dynamic system is only a function of the number of outputs present in the system rather than the future or past window lengths.

Statistical Process Monitoring via CV State Variables

The CV state-space realization provides the principal directions of variability of a linear dynamic system through the conditional state variables z_k. The conditional state variables are orthogonal vectors of the space of the past measurements that are highly correlated to the space of the future measurements. The CV realization is symmetric with respect to the past and future observations. This is consistent with the fundamental symmetry property of the Markov chains (Desai et al., 1985).

Assume that the autocorrelated in-control variability satisfies time invariance and regularity conditions. Based on the CV analysis, the mean of the conditional state vector is zero and its covariance matrix is a constant diagonal matrix. It is therefore possible to define a scalar statistic T^2, also known as the Mahalanobis distance (Johnson and Wichern, 1992),

$$T_k^2 = z_k^T \Sigma^{-1} z_k \sim \frac{n(N^2-1)}{N(N-n)} F_{n,N-n}, \qquad (24)$$

where subscript k indicates time. Equation 24 is obtained by assuming that the state variables follow a Gaussian distribution while they are orthogonal at zero lag (Tracy et al., 1992). The T^2 statistic is a metric that includes information for both level (mean) and covariance structure of the state variables. A change in the parameters of the in-control VARMA process can cause a bias on the in-control zero-mean state variables and/or the loss of orthogonality between the in-control state variables, resulting in T^2 values that are outside the in-control range. Therefore, monitoring the T^2 statistic will indicate whether a change has occurred in the overall in-control variability of a plant by including information from *all* process measurements.

The F distribution is obtained by assuming that the future observations of the state variables to be monitored are independent to those obtained from the in-control data. Nomikos and MacGregor (1995) used a T^2 statistic for monitoring batch processes, where the conditional state variables are replaced with scores (PLS or PCA), which are linear combinations of lagged observations. They justify the use of F distribution by arguing that the linear combination provides access to the central limit theorem. In this article, the F statistic will be used based on a justification using lambda distribution analysis of the random errors. If it is established that the random errors e_k have normal distributions by fitting lambda distributions, since orthogonal state variables are added to a linear combination of approximately Gaussian measurement errors, they can also be considered as multivariate Gaussian. Under these conditions the T^2 statistic given in Eq. 24 can follow an F distribution where n depends on the number of state variables retained for the purposes of monitoring. If the marginal distributions of e_k do not have normal distributions, since the state variables will again be linear combinations of the non-Gaussian measurement errors, it is still possible to argue normality of the state variables due to the central-limit theorem. However, a more robust and reliable way will be to fit a multivariate lambda distribution directly on the in-control state variables, but this requires some additional research.

Equation 24 is written by including all n state variables in the T^2 statistic. As is shown later, a subset of the n state variables can be used in computing T^2. The number of conditional state variables to include in the computation of T^2 depends on the fraction of total state variability that will be considered in implementing process monitoring.

Statistical Monitoring of a HTST Pasteurization Process

Milk pasteurization is the process of heating every particle of milk product in properly designed and operated equipment, to one of the seven temperature and residence (holding) time combinations specified in the Grade A: Pasteurized Milk Ordinance (U.S. Public Health Service, 1993). To maintain process safety and product quality standards, the flow and temperature of the milk is to be controlled as tightly as possible around preselected target values of process variables. The HTST milk pasteurization unit used in this study is located at the National Center for Food Safety and Technology (NCFST), Summit-Argo, IL. It is a pilot-scale continuous milk pasteurizer with a product flow-rate capacity of 400 gph. The process description is given in Appendix B and Figure A1. Detailed instrumentation and control strategies are given in Negiz et al. (1996) and Negiz (1995).

Cinar et al. (1995) introduced the prototype of the HTST multivariable control system where the product temperature and flow measurement control were incorporated into a single parameter known as total (or equivalent) process lethality. It is the product of current residence time in the holding tube and a lethality that is unity at 161°F (72°C). At 161°F the regulation requires a 15-s residence time, which corresponds to a total lethality level of "15 s." There are several temperature residence time combinations that can achieve the 15-s total lethality level. The control system developed uses this flexibility.

In-control variation

The state of the process is determined from five process measurements taken at 1-s intervals. The measured process variables are the holding tube exit temperature (product temperature), the hot water outlet temperature, the hot water inlet temperature, the product flow rate, and the differential pressure in the regenerator. The total lethality is computed

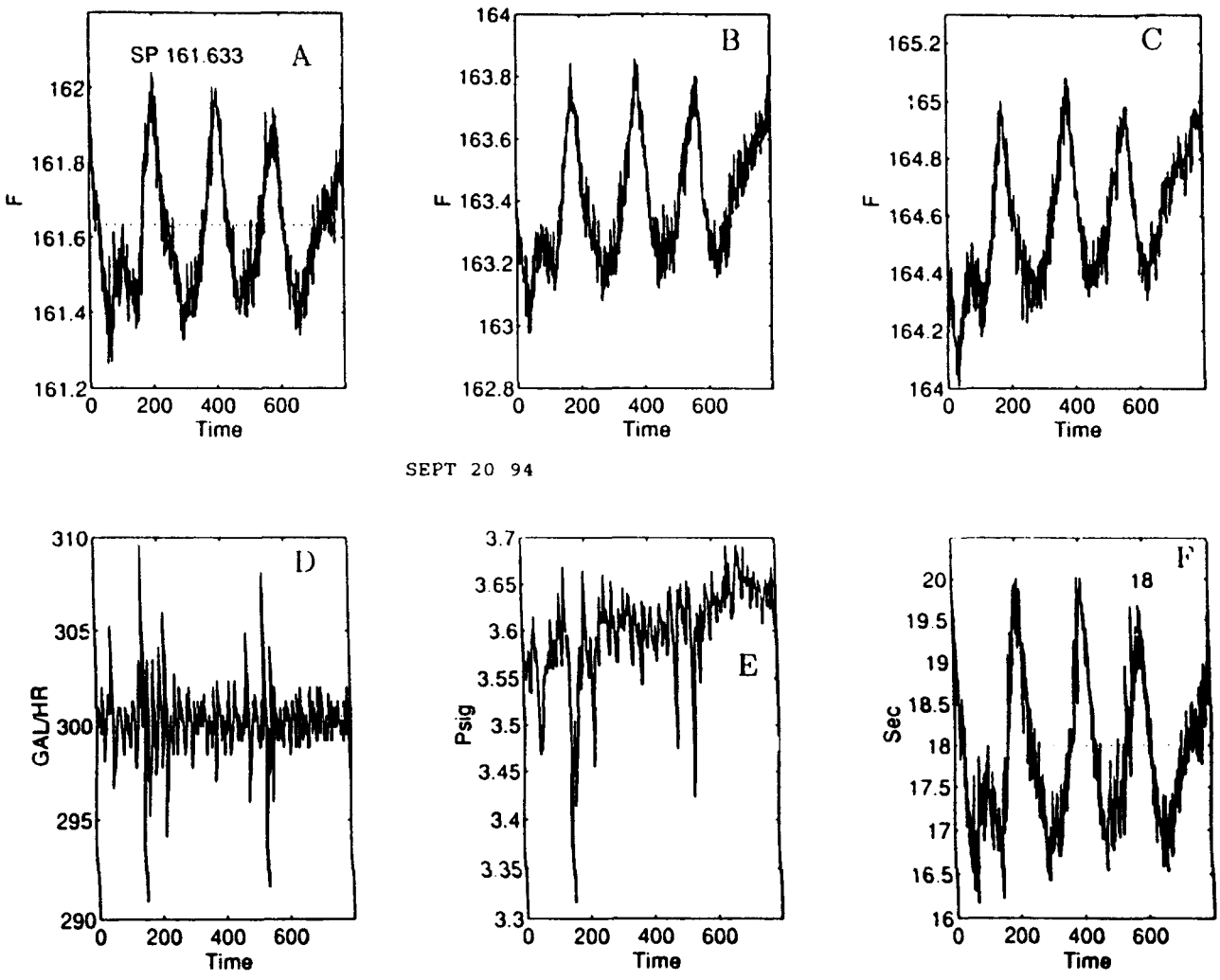

Figure 3. In-control variation of the HTST process.

Measurements of (a) product temperature; (b) hot water outlet temperature; (c) hot water inlet temperature; (d) product flow rate; (e) pressure differential; and (f) total lethality.

based on the product temperature and product flow-rate measurements. Since it is a critical process variable, it is included in the list of variables monitored. The process is perturbed by the variations and disturbances normally available in the inputs.

Figure 3 depicts results from an experiment where the product temperature has a setpoint value of 161.633°F (72.018°C) (this value is computed automatically by the control algorithm based on total lethality and it cannot be rounded off manually to a realistic temperature reading), and the setpoint for the product flow rate is at 300 gal/h (1,137 L/h). Based on the setpoints for flow rate and product temperature, the total lethality is set to 18 s, which provides a safety margin over the 15-s lethality level imposed and corresponds to 20% overprocessing. The standard deviation of product temperature and product flow rate around their setpoint values are 0.17°F (−17.68°C) and 1.7 gal/h (6.4 L/h), respectively. The minimum and maximum values observed are 161.26°F (71.81°C) and 162.04°F (72.24°C) for product temperature, and 291.0 gal/h (1,102.9 L/h) and 310.0 gal/h (1,174.9 L/h) for product flow rate, respectively. Although the variation of the process variables around their targets are acceptable as far as public safety is concerned, the variation is not totally random. Figure 4 supports this claim statistically, where the autocorrelations of the observations given in Figure 3 are shown at various lags. The solid and dashed lines around the zero autocorrelation are the 95 and 99% statistical thresholds of the null hypothesis that the observations are uncorrelated (*iid*). The main disturbances affecting the temperatures are at low frequencies, while the disturbances affecting the flow rate show a wider range of frequencies. The low frequency disturbance acting on the temperatures was due to a worn-out regulator in the steam line, which reduces the steam-line pressure from 200 psig (1.4 MPa) to 40 psig (413 kPa). This persistent disturbance generates control moves from the feedback controller. As a result of feedback control actions, the temperatures are varying in a highly positively autocorrelated manner around their setpoints (Figures 3 and 4). The large surges in product flow rate are due to the frequency changes in the electricity line, which might have been avoided with a buffer.

It is desired to keep the process within this variability re-

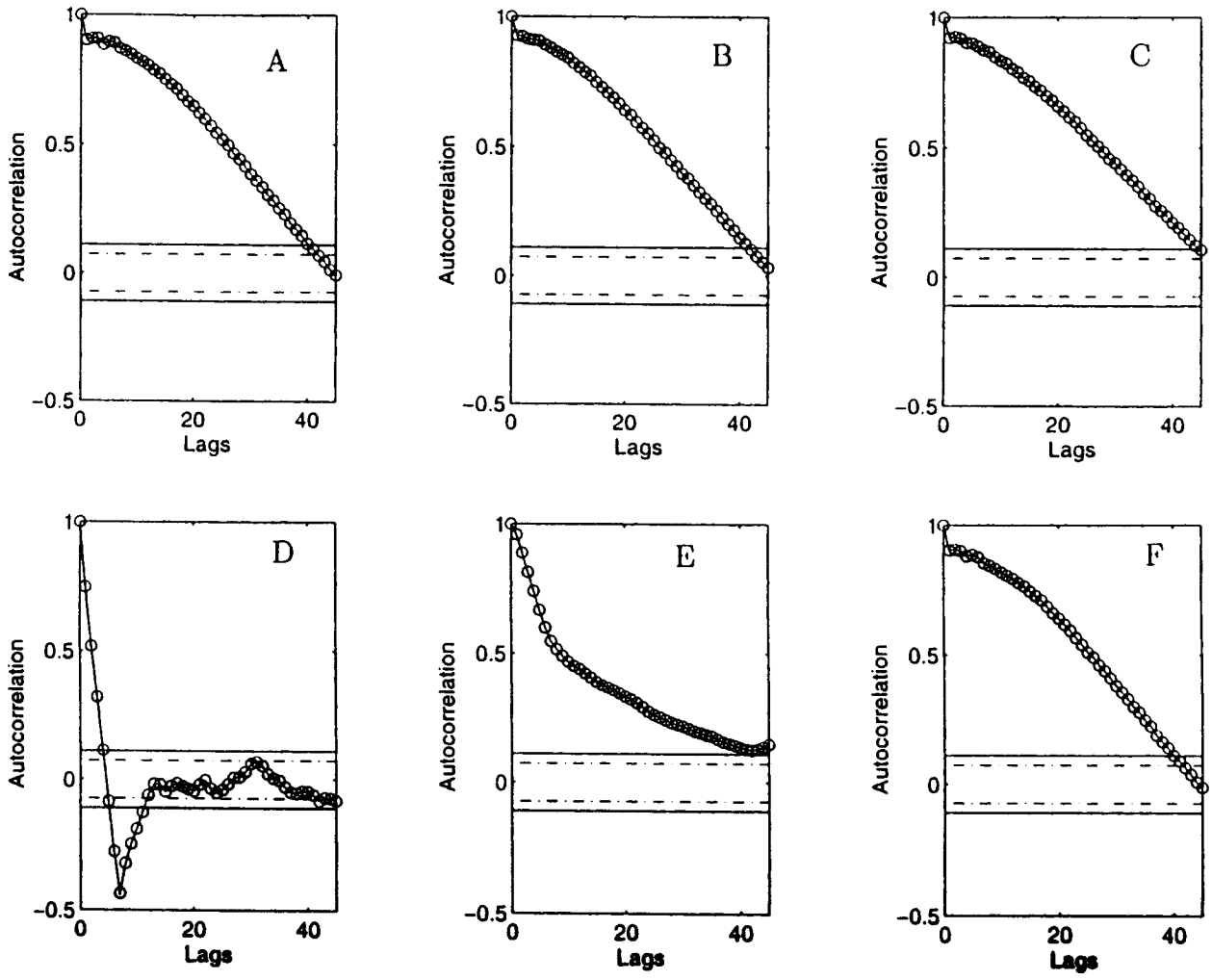

Figure 4. Autocorrelation coefficients of the in-control variation.

Measurements of (a) product temperature; (b) hot water outlet temperature; (c) hot water inlet temperature; (d) product flow rate; (e) pressure differential; and (f) total lethality.

gion. SPM will provide early warnings in case the process drifts away from its acceptable region of variability. To develop an effective SPM procedure, the structure of the allowed variability has to be statistically modeled. Due to autocorrelations, and possible cross-correlations, this task requires a dynamic model that can describe the behavior in such a way that the unexplained part of the variation is random noise. The statistical characterization of the in-control process variability is performed with CV stochastic realization. First a CV state-space model is extracted from the data, which yields residuals with insignificant auto- and cross-correlations. Then the marginal distributions of these residuals are characterized by using the lambda family of distributions.

There are six process measurements ($p = 6$), and the length of the calibration data is $N = 798$. Preliminary modeling activity for building autoregressive models for each variable indicates that 20 is the maximum significant lag, after which autocorrelations become insignificant for all the process variables. A robustness margin is added, and the past (K) and the future (J) window lengths for the CV state-space model are set at 24 and 30, respectively.

The Hankel matrix is constructed by computing the covariance between the 180×1 future stacked vector and the 144×1 past stacked vector according to Eq. 7. The CV decomposition is performed according to the scaled Hankel matrix (Eq. 13). Figure 5a shows the singular values of the Hankel matrix obtained from Eq. 7. These values tend to zero exponentially. The 20 singular values that are retained for the state-space model are shown in circles. After the first 20, the remaining singular values are essentially zero and they are not included in the model. Figure 5b shows the singular values of the normalized Hankel matrix (obtained by Eq. 13), which are the elements of the diagonal covariance matrix of the conditional state variables (Eq. 18). They also represent the correlation coefficients between the 20 pairs of future and past canonical variates. Figure 5b shows that the canonical correlations get smaller until a plateau is reached around the tenth state, and from that point on the canonical correlations fall off slowly. This is typical of sample canonical correlations of real systems that can have an infinite order. Twenty state variables were selected to retain in the model in order to generate the residuals and identify their distributions. We wanted to make sure that all dynamic information is retained for generating

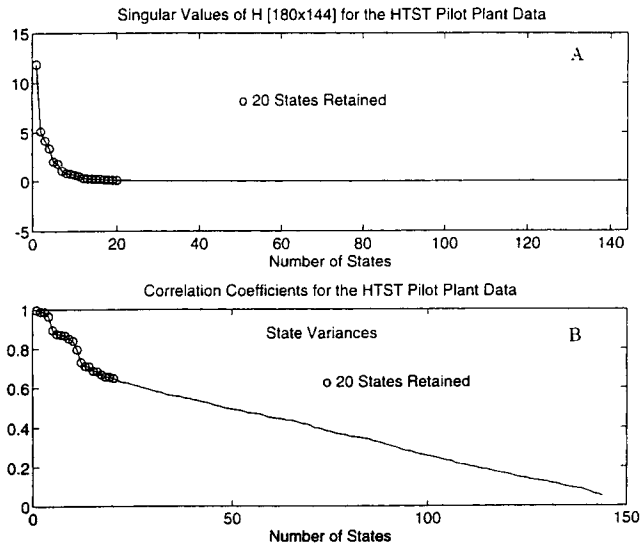

Figure 5. Singular values for the canonical variate model: (a) Hankel matrix; (b) normalized Hankel matrix.

the residuals and assessing the performance of the model. However, as discussed in the "Constructing the T^2 Chart" subsection, only the first 4 state variables out of the 20 are used to monitor the dynamic variation of the entire process.

The AIC criterion suggested a 10-state model. Indeed, the first 11 singular values and correlation coefficients in Figure 5 are larger than the next nine that are clustered together. Hence, one could have constructed a 10- or 11-variable state-space model. Due to the orthogonality properties (with respect to the zero lag covariance) of the state variables obtained through the CV realization, reduced-order models are easily derived from large-order models by just eliminating rows and columns of the system matrices that belong to the undesired state variables (Desai et al., 1985). For fault diagnosis, the 20 state-variable model may be preferred since it describes the details in process behavior more accurately. Often these details play an important role in discriminating variation. A 20 state-variable model was used first for analyzing the predictive capability of the model by analyzing the residuals. As illustrated below the first four variables carry enough information to monitor the process. Consequently, T^2

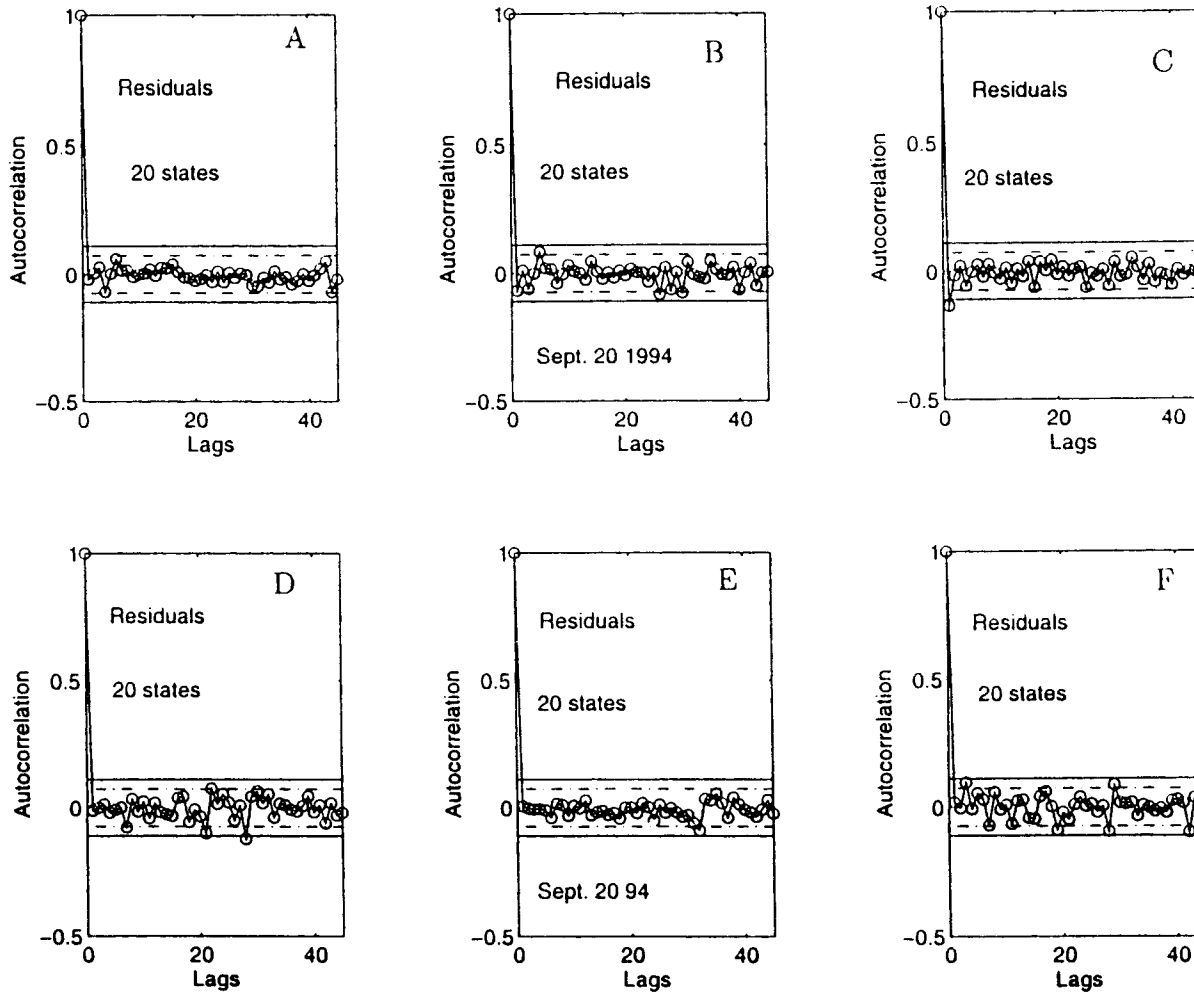

Figure 6. Residual autocorrelation coefficients based on the 20 state-variable model for the in-control variation.

(a) Product temperature; (b) hot water outlet temperature; (c) hot water inlet temperature; (d) product flow rate; (e) pressure differential; and (f) total lethality.

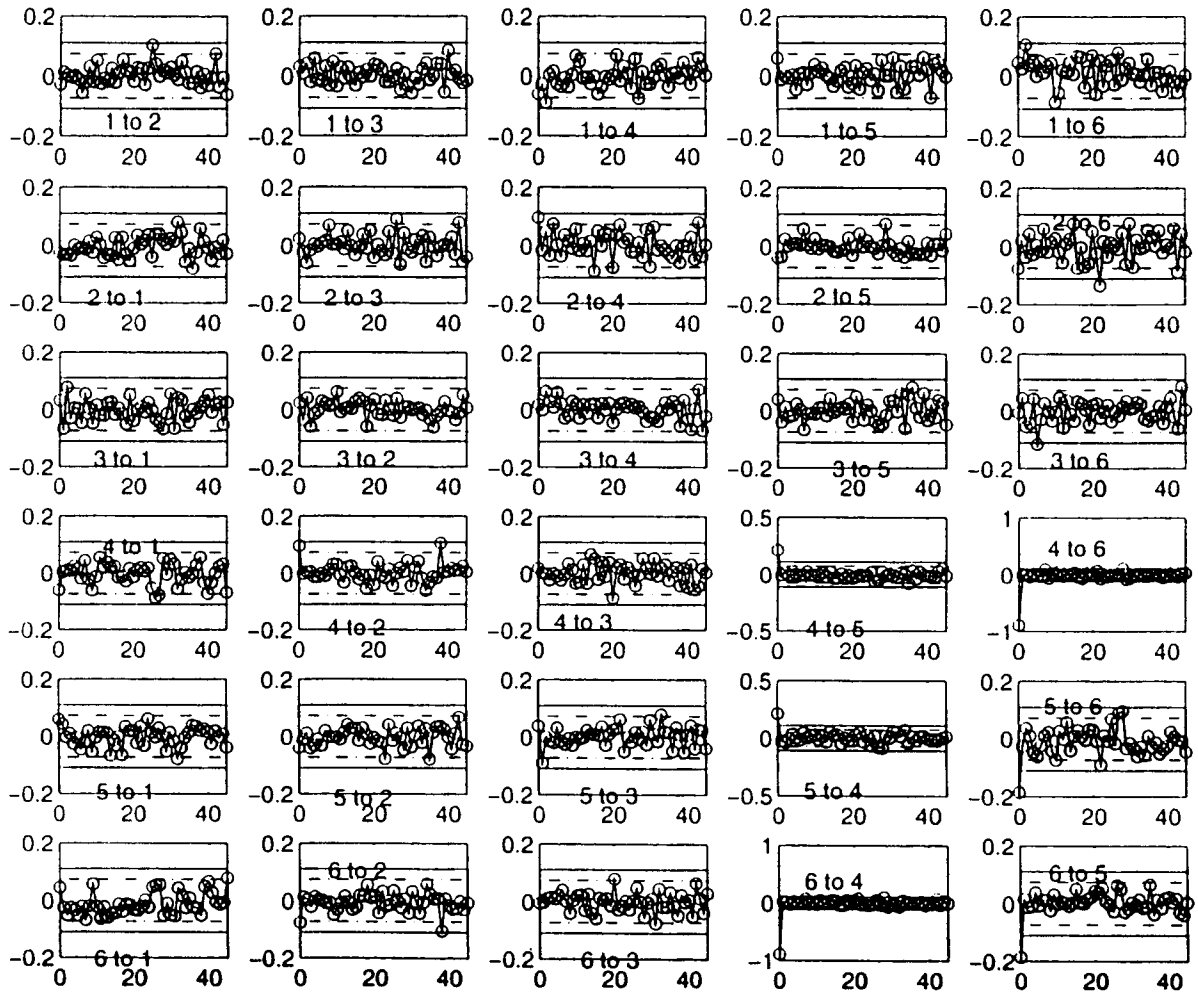

Figure 7. Residual cross-correlation structure based on the 20 state-variable model for the in-control variation.

Residuals include: (1) product temperature; (2) hot water outlet temperature; (3) hot water inlet temperature; (4) product flow rate; (5) pressure differential; and (6) total lethality.

HTST pasteurization system are based on the first four state variables.

Once the state variables are obtained by CV decomposition and Eq. 14, the least squares solution provides the system matrices A, C, B. The 20 eigenvalues of the transition matrix are all within the unit circle (z-domain); therefore, the model is stable and the statistical properties of the model are considered to be stationary. The small-amplitude oscillations in temperature measurements around their target values (Figure 3) are due to a complex pair of eigenvalues close to the unit circle ($0.9897 \pm 0.0356i$). This may be attributed to feedback-control action.

The one-step-ahead innovations (residuals) are computed by

$$e_k = y_k - Cz_k, \tag{25}$$

which must be serially uncorrelated except at the zero lag according to the innovation model. Indeed, all autocorrelation coefficients of the residuals are within the region where the variation is described as random (serially independent) (Figure 6). The systematic autocorrelation structure of the original measurements depicted is described by the CV state-space model. Figure 7 shows the cross-correlations between the six elements of the residual vector. The variables are ordered from one to six as the product temperature, hot water inlet and outlet temperatures, flow rate, differential pressure, and total lethality. Contrary to autocorrelations, the cross-correlations are not symmetric across the zero lag. Therefore it is important to identify which residual element is leading or lagging behind in the computations. For example, the crossplot at the upper-left corner of Figure 7 denotes the cross-correlation coefficients of the first residual element (product temperature residual) with respect to the leading values of the second residual element (hot water inlet temperature). Similarly, the crossplot located at the second row of the first column shows the cross-correlations between the same residual elements where product temperature residuals are leading hot water inlet temperature residuals. All residual cross-correlation plots of Figure 7 indicate that the residuals have no serial cross-correlation except at the zero lag, which satisfies the requirement of the innovation model. The zero lag structure of the residual autocorrelations can be extracted from the estimate of the residual covariance matrix ($\hat{\Delta}$) according to Eq. 17.

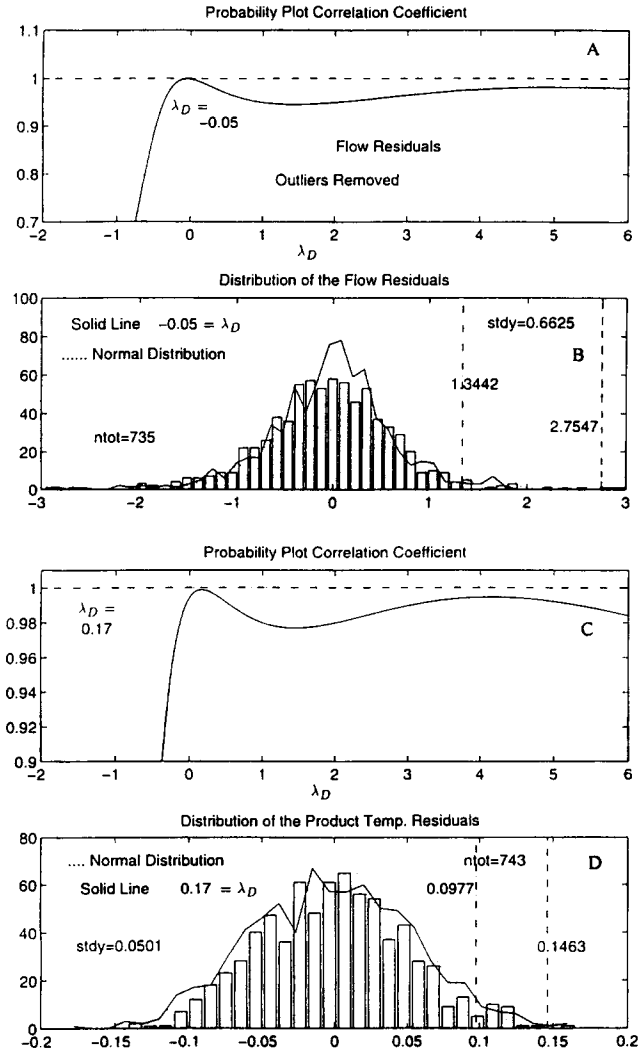

Figure 8. Fitting lambda distribution to flow rate and product temperature residuals.

Flow-rate residuals: (a) correlation coefficient vs. λ_D; (b) histograms. Product temperature residuals: (c) correlation coefficient vs. λ_D; (d) histograms. Residuals (bars), lambda random variates (solid line); Gaussian random variates (dotted line). The outliers are excluded.

Fitting lambda distributions to residuals

Figure 8a shows the correlation coefficients as a function of the λ_D parameter for the flow residuals. The coefficient is computed by excluding eight flow-rate residuals that are declared as outliers from a previous λ_D distribution that included all of the flow residuals (Negiz, 1995). For $\lambda_D = -0.05$, the correlation coefficient is closest to unity. This suggests that the lambda distribution with its parameter set at $\lambda_D = -0.05$ is a fit.

Figure 8b assesses this proposition by depicting several histograms. The histogram shown with the bars is the distribution of the 735 residual data set. The standard deviation of the residuals is 0.6625. The histogram with the dotted line is the histogram of the 735 randomly generated data points that follow a Gaussian distribution with the same variance as the residuals. The histogram with the solid line is the histogram of the 735 randomly generated data points that follow a $\lambda_D = -0.05$ distribution with the same variance as the residuals. The dash-dotted vertical line shows the 95th, and the dashed vertical line shows the 99th percentile point of the lambda distribution with $\lambda_D = -0.05$. The 99% point (± 2.7547) is larger than the normal percentile point that is approximately ± 1.9875.

Figures 8c and 8d depict the lambda distribution fits for the product temperature. The 99% point is almost at the three standard deviations away ($3 \times 0.0501 \simeq 0.1463$) from zero, indicating a Gaussian approximation. For product temperature, hot water inlet and outlet temperatures, and differential pressure residuals no data points were declared as outlier, since there were no specific source causes that are as plausible as in the case of the flow-rate residuals. The temperature and pressure residuals have lambda distributions that are very close to normal. For hot water inlet and outlet temperature residuals the parameters are 0.20 and 0.17, respectively. For pressure differential residuals λ_D is 0.12 and the residuals can be assumed to have normal distribution, since $\lambda_D = 0.135$ is normal.

Since the total lethality value is derived from the flow-rate measurement, the outliers for the flow-rate residuals were also taken out from the total lethality residuals. The lambda parameter for lethality residuals is found to be 0.030 and the 99% point is almost at three standard deviation units away from zero, indicating approximately a Gaussian distribution.

Constructing the T^2 chart

Measurement errors can be assumed to follow approximately a multivariate Gaussian distribution, since their marginals are approximately normal with the exception of product flow-rate residuals distribution. Since the orthogonal state variables are added to a linear combination of approximately Gaussian measurement errors, they can also be considered as multivariate Gaussian. Hence, the T^2 statistic given in Eq. 24 follows an F distribution, where n is the number of state variables retained for monitoring.

Figure 9a depicts the ratio of the sums of standard deviations of one-step-ahead CV state-space model predictions to those of the present outputs. The ratio is computed as a function of the state-space model order n. Figure 9b shows the cumulative ratio of the singular values as a function of model order. For the 20 CV state-space model 92% of the variation in the present outputs is explained (Figure 9a) by using 95% of the state-space variation (Figure 9b). The residual autocorrelation and cross-correlation plots indicate that the remaining 8% of the variance is random. The first four of the CV state variables explain almost 90% of the variation in the present outputs (Figure 9a) by using 70% of the overall state variability (shown by the dotted vertical line in Figure 9b). Hence, constructing a T^2 chart that includes only the first four state variables is considered as sufficient to determine if process variation is in-control, and the first four state variables are used to compute the T^2 statistic in this study.

Monitoring of the HTST pasteurization with a T^2 statistic

The T^2 statistic based on the first four state variables is computed by using Eq. 24 with $n = 4$. Figure 10a depicts the flow measurements of the *calibration* data. Figure 10b shows

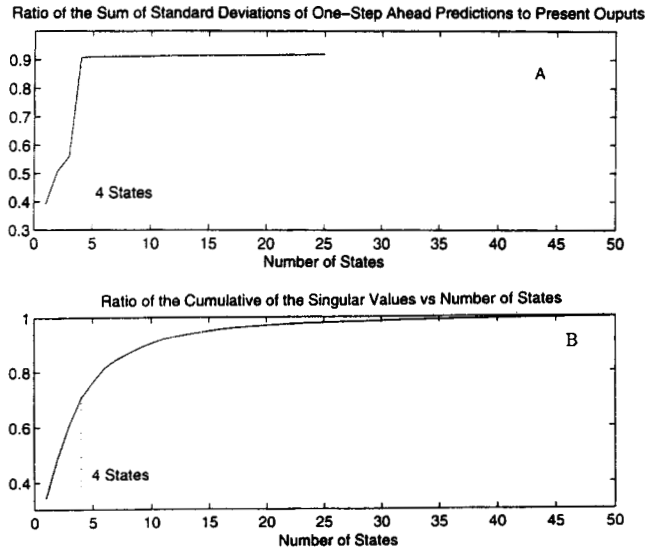

Figure 9. Variability of the CV state variables.

(a) Ratio of sum of standard deviations of one-step ahead predictions to present outputs; (b) cumulative ratio of the singular values vs. number of state variables.

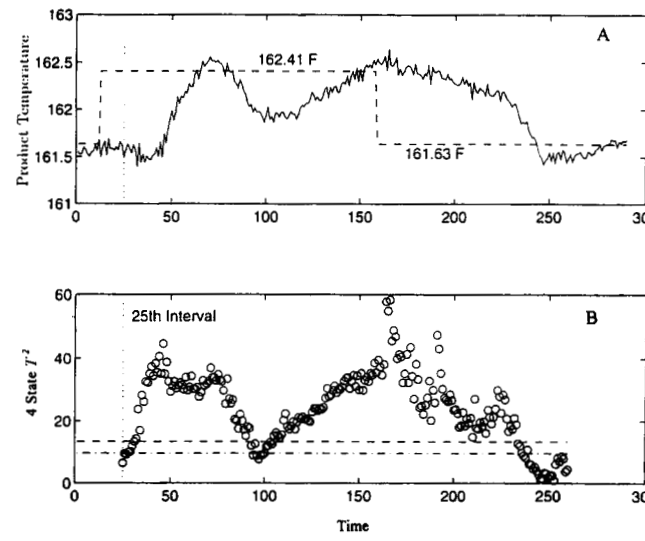

Figure 11. Monitoring of HTST pasteurization plant with T^2 statistic.

(a) A change in setpoint of product temperature; (b) response of T^2 statistic.

the T^2 statistics computed based on Eq. 24 with 95% (dashed-dotted line) and 99% (dashed line) confidence thresholds. Since $K = 24$, the computation of the values for the conditional state variables starts at $k = 25$. The T^2 statistics signals out-of-control situations synchronously with the flow disturbances. The two-state version of the T^2 statistic was also used to verify the outliers for the flow residuals (Negiz, 1995). If the T^2 statistic at times k are considered to be independent of the T^2 statistic at times $k-1$, then the average run length (ARL) of the T^2 chart will be 100, based on a geometric distribution. However, this assumption is not true in this case, where the T^2 statistic is constructed from conditional state variables that are autocorrelated. The T^2 statistic is used in order to detect a significant change either in the zero lag state covariance matrix or in the levels of the state variables. Figure 10b shows that the ARL between the flow disturbances is around 300, which is higher than the *iid* case (100).

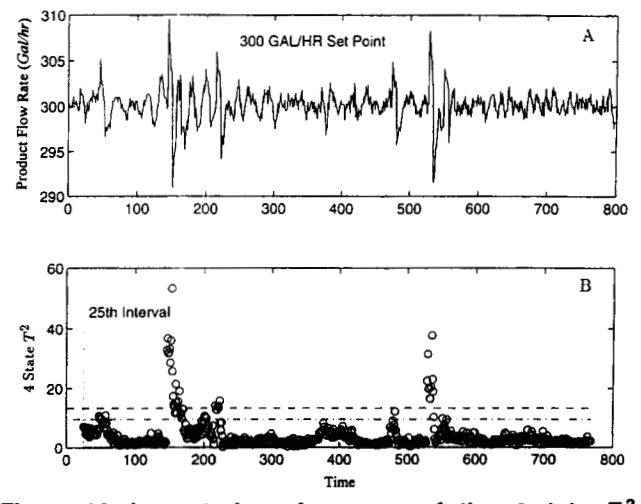

Figure 10. In-control performance of the 4-state T^2 chart for monitoring the HTST pasteurization process: (a) flow measurements; (b) T^2 chart.

Figure 11 shows the response of the T^2 statistic to a setpoint change in the product temperature (Figure 11a) while keeping the flow-rate setpoint at 300 gal/h (2.1 m^3/h). The setpoint change for the product temperature was conducted such that the process is returned to the in-control operating region after it is disturbed (Figure 11a). The T^2 statistic goes out of control before the product temperature starts to increase. Since the T^2 statistic covers the variability of a process by utilizing all of the process measurements, even if one process variable has a delay in its response to a disturbance or a change, the T^2 statistic (Figure 11b) still becomes sensitive to the change through the responses of the remaining process variables. In this case, the delay in the response of product temperature is due to the effect of the holding tube, but the hot water outlet temperature responds faster to the setpoint change because of the cascade controller. This illustrates one of the advantages of using multivariate statistical process monitoring techniques as opposed to univariate SPM techniques. Near 240 s, T^2 returns to its in-control region, as the product temperature setpoint is changed back to its average in-control value at 161.63°F (72.02°C).

Figure 12 depicts another case where the valve that regulates the flow rate of water circulating inside the heater is opened slightly to stimulate an increased heat load to the system. Figure 12a shows the response of the total lethality to the disturbance while the in-control set point is at 18 s. The T^2 statistic reflects this variability with a small time delay (Figure 12b). The time delay should be expected since the conditional state at time k is computed by using the past measurements up to time $k-1$, as part of the computation of the one-step-ahead predictions at time k.

Figure 13 shows a case where the setpoint for the flow rate was deliberately changed by a step change from the in-con-

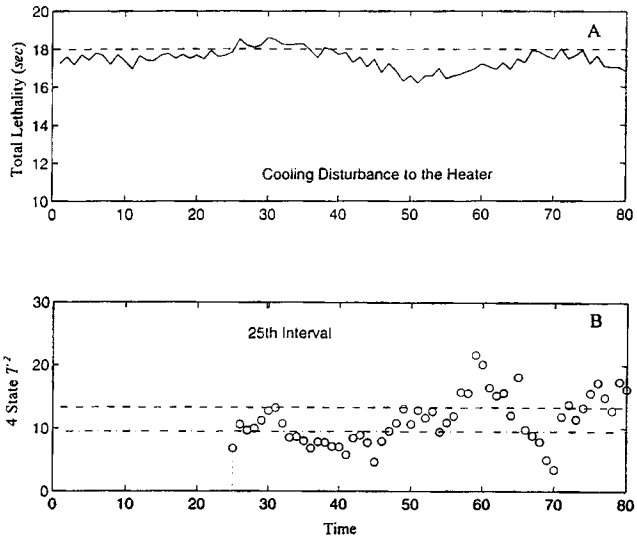

Figure 12. Monitoring the HTST pasteurization plant with T^2 statistic.

(a) Response of total lethality to a heating disturbance; (b) response of T^2.

trol flow rate setpoint, which is 300 gal/h (2.1 m^3/h). Figure 13a depicts the setpoint trajectories (dashed line), and the corresponding flow-rate measurement (solid line). Figure 13b shows the response of the T^2 statistic, starting from the 25th time interval due to initialization based on $K = 24$. The T^2 statistic shows good sensitivity in detecting changes even for the cases when the imposed setpoint change is short-lived.

Conclusions

Experience with experimental data indicates that measurements of dynamic processes do violate conditions of statistical independence. An SPM method is proposed to integrate canonical-variate state-space realization, determination of the marginal distributions of residuals, and monitoring the T^2 of canonical-variate state-space variables. The CV state-space realization provides a powerful dynamic model development environment for processes that have a large number of process variables yielding highly autocorrelated and cross-correlated measurements. The statistical properties of the CV state variables provide the principal dimensions that can be lumped into a scalar statistic, the T^2, which provides the single variable to monitor the variability of an entire process.

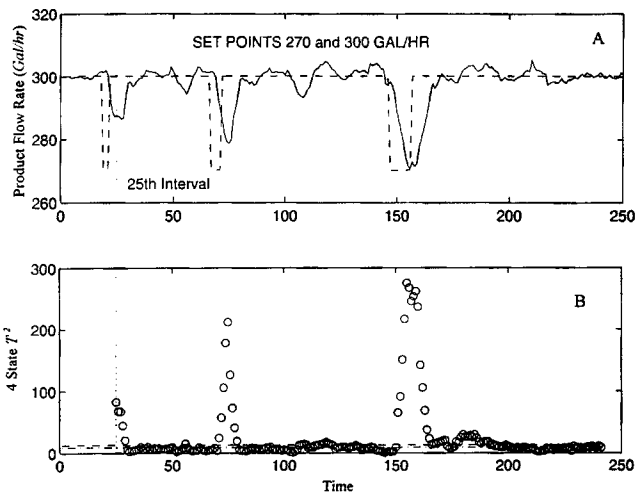

Figure 13. Monitoring the HTST pasteurization plant with T^2 statistic.

(a) Response of product flow rate to setpoint variations in flow rate; (b) response of T^2.

Acknowledgment

The authors gratefully acknowledge the ASA/NSF/NIST Research Associate appointment of A. Negiz in 1992–1993, and the research assistantship provided to A. Negiz by the National Center for Food Safety and Technology.

Literature Cited

Alt, F. B., "Multivariate Quality Control," *Encyclopedia of Statistical Sciences*, S. Case and N. L. Johnson, eds., Wiley, New York (1985).

Akaike, H., "Stochastic Theory of Minimal Realization," *IEEE Trans. Automat. Contr.*, **AC-19**, 667 (1974a).

Akaike, H., "A New Look at the Statistical Model Identification," *IEEE Trans. Automat. Contr.*, **AC-19**, 716 (1974b).

Anderson, B. D. O., and J. B. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ (1979).

Aoki, M., *State Space Modeling of Time Series*, 2nd ed., Springer-Verlag, New York (1990).

Åstrom, K. J., *Introduction to Stochastic Control Theory*, Academic Press, New York (1970).

Åstrom, K. J., and B. Wittenmark, *Computer-Controlled Systems: Theory and Design*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ (1990).

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

Box, G. E. P., and G. C. Tiao, "A Canonical Analysis of Multiple Time Series," *Biometrika*, **64**(2), 355 (1977).

Broomhead, D. S., and G. P. King, "Extracting Qualitative Dynamics from Experimental Data," *Physica D*, **20**, 217 (1986).

Candy, J. V., T. E. Bullock, and M. E. Warren, "Invariant System Description of the Stochastic Realization," *Automatica*, **15**, 493 (1979).

Cinar, A., J. E. Schlesser, A. Negiz, P. Ramanauskas, D. J. Armstrong, and W. Stroup, "Automated Control of High Temperature Short Time Pasteurization Systems Based on Lethality Limits," *Proc. Food Processing Automation Conf. IV*, ASAE, St. Joseph, MI p. 394 (1995).

Desai, U. B., D. Pal, and R. D. Kirkpatrick, "A Realization Approach to Stochastic Model Reduction," *Int. J. Contr.*, **42**, 821 (1985).

Filliben, J. J., "The Probability Plot Correlation Coefficient Test for Normality," *Technometrics*, **17**, 111 (1975).

Filliben, J. J., "Simple and Robust Linear Estimation of the Location Parameter of a Symmetric Distribution," PhD Diss., Princeton Univ., Princeton, NJ (1969).

Harris, T. J., and W. H. Ross, "Statistical Process Control Procedures for Correlated Observations," *Can. J. Chem. Eng.*, **69**, 48 (1991).

Hawkins, D. M., "Multivariate Quality Control Based on Regression Adjusted Variables," *Technometrics*, **33**, 61 (1991).

Hoskuldsson, A., "PLS Regression Methods," *J. Chemometrics*, **2**, 211 (1988).

Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components," *J. Educ. Psychol.*, **24**, 417, 498 (1933).

Jackson, J. E., and G. S. Mudholkar, "Control Procedures for Residuals Associated with Principal Components Analysis," *Technometrics*, **21**, 341 (1979).

Johnson, R. A., and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 3rd ed., Prentice Hall, Englewood Cliffs, NJ (1992).

Joiner, B. L., and J. R. Rosenblatt, "Some Properties of the Range in Samples from Tukey's Symmetric Lambda Distributions," *J. Amer. Stat. Assoc.*, **66**, 394 (1971).

Kaspar, M. H., and W. H. Ray, "Dynamic PLS Modeling for Process Control," *Chem. Eng. Sci.*, **48**, 3447 (1993a).
Kaspar, M. H., and W. H. Ray, "PLS Modeling as Successive Singular Value Decompositions," *Comput. Chem. Eng.*, **17**, 985 (1993b).
Kourti, T., and J. F. MacGregor, "Process Analysis, Monitoring and Diagnosis Using Multivariate Projection Methods," *Chemometrics Intelligent Lab. Syst.*, **28**, 3 (1995).
Kresta, J., J. F. MacGregor, and T. E. Marlin, "Multivariable Statistical Monitoring of Process Operating Performance," *Can. J. Chem. Eng.*, **69**, 35 (1991).
Ku, W., R. H. Storer, and C. Georgakis, "Disturbance Detection and Isolation by Dynamic Principal Components Analysis," *Chemometrics Intelligent Lab. Syst.*, **30**, 179 (1995).
Larimore, W. E., "Accuracy Confidence Bands Including the Bias of Model Under-Fitting," *Proc. Amer. Control Conf.*, p. 1995 (1993).
Larimore, W. E., "Canonical Variate Analysis in Identification, Filtering, and Adaptive Control," *Proc. IEEE Conf. on Decision and Control*, IEEE, Piscataway, NJ, p. 596 (1990).
Larimore, W. E., "System Identification, Reduced-Order Filtering and Modeling via Canonical Variate Analysis," *Proc. Automatic Control Conf.*, p. 445 (1983).
Lütkepohl, H., *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Heidelberg (1991).
Ljung, L., *System Identification*, Prentice Hall, Englewood Cliffs, NJ (1987).
MacGregor, J. F., C. Jaeckle, C. Kipparissides, and M. Koutoudi, "Process Monitoring and Diagnosis by Multiblock PLS Methods," *AIChE J.*, **40**, 826 (1994).
Maciejowski, J. M., *Multivariable Feedback Design*, Addison-Wesley, New York (1989).
Montgomery, D. C., *Introduction to Statistical Quality Control*, Wiley, New York (1991).
Montgomery, D. C., and C. M. Mastrangelo, "Some Statistical Process Control Methods for Autocorrelated Data," *J. Qual. Technol.*, **23**, 179 (1991).
Negiz, A., A. Çinar, J. E. Schlesser, P. Ramanauskas, D. J. Armstrong, and W. Stroup, "Automated Control of High Temperature Short Time Pasteurication," *Food Control*, **7**, 309 (1996).
Negiz, A., and A. Çinar, "PLS as a Technique in Identifying Vector Autoregressive Moving Average Models in State Space," *Chemoemtrics Intelligent Lab. Syst.*, in press (1996).
Negiz, A., "Statistical Dynamic Modeling and Monitoring Methods for Multivariable Continuous Processes," PhD Diss., Illinois Institute of Technology, Chicago (1995).
Negiz, A., and A. Çinar, "A Parametric Approach to Statistical Monitoring of Processes with Autocorrelated Observations," AIChE Meeting, Miami, FL (1995).
Negiz, A., and A. Çinar, "Statistical Process Control of Autocorrelated Observations by Monitoring Model Parameters," *Technometrics* (1994).
Negiz, A., and A. Çinar, "On the Detection of Multiple Sensor Abnormalities in Multivariable Processes," *Proc. Amer. Control Conf.*, p. 2364 (1992).
Negiz, A., E. Lagergren, and A. Cinar, "Statistical Quality Control of Multivariable Continuous Processes," *Proc. Amer. Control Conf.*, p. 1289 (1994).
Nomikos, P., and J. F. MacGregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, **37**, 41 (1995a).
Nomikos, P., and J. F. MacGregor, "Multiway Partial Least Squares in Monitoring Batch Processes," *Chemometrics Intelligent Lab. Syst.*, **30**, 97 (1995b).
Nomikos, P., and J. F. MacGregor, "Monitoring Batch Processes Using Multiway Principal Component Analysis," *AIChE J.*, **40**, 1361 (1994).
Raich, A., and A. Çinar, "Statistical Process Monitoring and Disturbance Diagnosis in Multivariable Continuous Processes," *AIChE J.*, **42**, 995 (1996).
Schaper, C. D., W. E. Larimore, D. E. Seborg, and D. A. Mellichamp, "Identification of Chemical Processes Using Canonical Variate Analysis," *Comput. Chem. Eng.*, **18**, 55 (1994).
Swindlehurst, A., R. Roy, B. Ottersten, and T. Kailath, "A Subspace Fitting Method for Identification of Linear State-Space Models," *IEEE Trans. Automat. Contr.*, **AC-40**, 311 (1995).
Tracy, N. D., J. C. Young, and R. Mason, "Multivariate Control Charts for Individual Observations," *J. Qual. Technol.*, **24**, 88 (1992).
Tsay, R. S., and G. C. Tiao, "Use of Canonical Analysis in Time Series Model Identification," *Biometrika*, **71**(2), 299 (1985).
Tukey, J. W., "The Future of Data Analysis," *Ann. Math. Stat.*, **33**, 1 (1962).
U.S. Public Health Service, "Grade A: Pasteurized Milk Ordinance —Revision," U.S. Dept. of Health and Human Services, U.S. Government Printing Office, Washington, DC (1993).
Van Overschee, P., and B. de Moor, "N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems," *Automatica*, **30**, 75 (1994).
Verhaegen, M., and P. Dewilde, "Subspace Model Identification. Part I: The Output Error State Space Model Identification Class of Algorithms," *Int. J. Contr.*, **56**, 1187 (1992).
Wei, W. W. S., *Time Series Analysis*, Addison-Wesley, Redwood City, CA (1990).
Wetherill, G. B., and D. W. Brown, *Statistical Process Control Theory and Practice*, Chapman & Hall, New York (1991).
Wold, S., A. Ruhe, H. Wold, and W. J. Dunn, "The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses," *SIAM J. Sci. Stat. Comput.*, **5**, 735 (1984a).
Wold, S., C. Albano, W. J. Dunn, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, and M. Sjöström, "Multivariate Data Analysis in Chemistry," *Chemometrics: Mathematics and Statistics in Chemistry*, B. R. Kowalski, ed., Reidel, Dordrecht, The Netherlands (1984b).
Yohai, V. J., and M. S. Garcia Ben, "Canonical Variables as Optimal Predictors," *Ann. Stat.*, **8**, 865 (1980).

Appendix A: Derivation of Eq. 12

Use Eqs. 7 and 10 to obtain Eq. 11 as

$$\boldsymbol{H}_{JK} = E\left(\mathcal{Y}_{k_J}^{+}\,\mathcal{Y}_{k-1_K}^{-T}\right) = \mathcal{O}_J\,\Omega_K. \tag{A1}$$

The projection theorem (Åstrom, 1970), or equivalently the multiple least-squares linear regression (Johnson and Wichern, 1992) gives

$$E(y_k|\mathcal{Y}_{k-1_K}^{-}) = E\left(y_k\,\mathcal{Y}_{k-1_K}^{-T}\right)E\left(\mathcal{Y}_{k-1_K}^{-}\,\mathcal{Y}_{k-1_K}^{-T}\right)^{-1}\mathcal{Y}_{k-1_K}^{-}. \tag{A2}$$

Based on the projection theorem Eq. A2 and the definitions of the covariances of the future and the past $\boldsymbol{R}_K^{-} = E(\mathcal{Y}_{k-1_K}^{-}\,\mathcal{Y}_{k-1_K}^{-T})$ and $\boldsymbol{R}_J^{+} = E(\mathcal{Y}_{k_J}^{+}\,\mathcal{Y}_{k_J}^{+T})$:

$$E(\mathcal{Y}_{k_J}^{+}|\mathcal{Y}_{k-1_K}^{-}) = \mathcal{O}_J\,\Omega_K(\boldsymbol{R}_K^{-})^{-1}\mathcal{Y}_{k-1_K}^{-}. \tag{A3}$$

Use repeated substitutions to Eq. 3 to arrive at

$$\mathcal{Y}_{k_J}^{+} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \boldsymbol{CA}^2 \\ \vdots \\ \boldsymbol{CA}^{J-1} \end{bmatrix} \boldsymbol{x}_k + \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{CB} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{CAB} & \boldsymbol{CB} & \boldsymbol{I} & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{CA}^{J-2}\boldsymbol{B} & \cdots & \boldsymbol{CB} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \epsilon_k \\ \epsilon_{k+1} \\ \epsilon_{k+2} \\ \vdots \\ \epsilon_{k+J-1} \end{bmatrix}. \tag{A4}$$

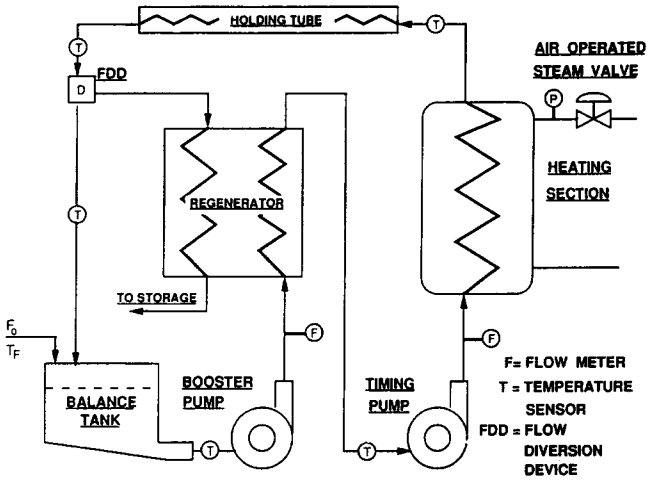

Figure A1. HTST pasteurization process.

Equation A4 can also be expressed by using the finite-dimensional observability matrix and defining $\mathcal{G}$ and $\mathcal{E}_k^+$ accordingly as

$$\mathcal{Y}_{k_j}^+ = \mathcal{O}_J \boldsymbol{x}_k + \mathcal{G}\mathcal{E}_k^+ . \tag{A5}$$

Taking the conditional expectation of Eq. A5 with respect to $\mathcal{Y}_{k-1_K}^-$, and noting that $E(\mathcal{E}_k^+ | \mathcal{Y}_{k-1_K}^-) = \mathbf{0}$,

$$E(\mathcal{Y}_{k_j}^+ | \mathcal{Y}_{k-1_K}^-) = \mathcal{O}_J E(\boldsymbol{x}_k | \mathcal{Y}_{k-1}^-) = \mathcal{O}_J z_k . \tag{A6}$$

Equation 12 is obtained by comparing Eqs. A3 and A6.

Appendix B: High-Temperature Short-Time Pasteurization Process

The HTST pasteurization system consists of a balance tank, a regeneration unit, a booster pump, a timing pump, a heating unit, a holding tube, and a flow-diversion device (Figure A1). Unpasteurized (raw) milk is transferred from the balance tank to the regeneration unit with the help of the booster pump. The regeneration unit consists of a plate heat exchanger in which the raw milk product passes on one side and the hot pasteurized product coming from the holding tube passes on the other. In this way, the raw milk is initially heated by recovering some of the heat that is available from the pasteurized product. From the regenerator, the raw milk is sent to the heating unit through a timing pump that assures the fixed flow rate required by the pasteurized milk ordinance. The heating unit, which is another plate heat exchanger, has raw milk passed through on one side and hot water on the other. As the milk passes through this unit, its temperature is raised well above the required pasteurization temperature. The heated milk then goes through a holding tube where it is held for a specific length of time according to the milk ordinance. The holding time is a function of the flow rate and the length of the holding tube. A flow-diversion device is located at the end of the holding tube. If the temperature of the heated milk exiting the holding tube is lower than the temperature according to the legal pasteurization temperature, the flow diversion-device directs the milk back to the balance tank for reprocessing.

If the milk temperature at the exit of the holding tube is above the legal limit, then it is deemed to have been properly pasteurized. The hot milk product travels through the heat-regeneration unit where it is partially cooled by raw milk flowing on the other side. The partially cooled pasteurized milk may then be sent for further cooling and/or storage.

Manuscript received Aug. 26, 1996, and revision received Feb. 27, 1997.