Statistical Process Monitoring Based on Dissimilarity of Process Data

Manabu Kano, Shinji Hasebe, and Iori Hashimoto
Dept. of Chemical Engineering, Kyoto University, Kyoto 606-8501, Japan

Hiromu Ohno
Dept. of Chemical Science and Engineering, Kobe University, Kobe 657-0013, Japan

Multivariate statistical process control (MSPC) has been widely used for monitoring chemical processes with highly correlated variables. In this work, a novel statistical process monitoring method is proposed based on the idea that a change of operating condition can be detected by monitoring a distribution of process data, which reflects the corresponding operating conditions. To quantitatively evaluate the difference between two data sets, a dissimilarity index is introduced. The monitoring performance of the proposed method, referred to as DISSIM, and that of the conventional MSPC method are compared with their applications to simulated data collected from a simple 2×2 process and the Tennessee Eastman process. The results clearly show that the monitoring performance of DISSIM, especially dynamic DISSIM, is considerably better than that of the conventional MSPC method when a time-window size is appropriately selected.

Introduction

Fault detection is one of the most important tasks for the successful operation of any process. On-line process monitoring plays an important role in detecting process upsets, equipment malfunctions, or other special events as early as possible. On-line process monitoring systems are required not only to detect changes in a process as early as possible but also to reduce the number of false alarms. Many approaches have been developed in order to fulfill those two conflicting requirements at the same time.

In chemical processes, statistical process control (SPC), which is a data-based approach for process monitoring, has been used widely and successfully. For example, Shewhart control charts, cumulative sum (CUSUM) control charts, and exponentially weighted moving average (EWMA) control charts are well-known SPC techniques. Such SPC charts are well established for monitoring univariate processes, but univariate SPC does not function well for multivariable processes. For example, when univariate SPC is applied to a multivariable process, the number of false alarms will increase. If there are P independent variables in a process and if an SPC chart with $P\{\text{type I error}\} = \alpha$ is maintained on each variable, then the probability of type I error for the joint control procedure is given by

$$\alpha' = 1 - (1 - \alpha)^P \tag{1}$$

Therefore, the true probability of type I error increases as the number of variables increases. In addition, if variables are not independent, it is difficult to measure the distortion in the joint SPC charts. To decrease the number of false alarms, the control limits should be relaxed. However, this modification reduces the detection speed of SPC charts. Furthermore, chemical processes are becoming more heavily instrumented, and process data are more frequently recorded. Therefore, in order to extract useful information from process data and utilize it for process monitoring, multivariate statistical process control (MSPC) has been developed.

The original MSPC method is based on the Hotelling T^2 statistic. The Hotelling T^2 control chart is considered as a multivariate version of the Shewhart control chart. For improving the monitoring performance, especially for detecting small shifts in the process, multivariate CUSUM control

Correspondence concerning this article should be addressed to M. Kano.

charts and multivariate EWMA control charts are available; or, a set of decision rules can be used together with the Shewhart control charts. For example, the Western Electric Handbook suggests the use of the following rules:

(1) One or more points outside the control limits

(2) Two of three consecutive points outside the 2-sigma limits

(3) Four of five consecutive points plot outside the 1-sigma limits

(4) A run of eight consecutive points on one side of the center line.

The use of these sensitizing rules improves the ability of control charts, but at the same time it causes an excessive number of false alarms. In addition, the use of supplemental rules makes monitoring systems more complicated, and the simplicity of the Shewhart control chart is lost.

Furthermore, in order to cope with collinearity problems caused by correlated process variables, the chemometric techniques such as principal component analysis (PCA) and partial least squares (PLS) can be used (Geladi and Kowalski, 1986). Wise and Gallagher (1996) reviewed some of the chemometric techniques and their application to chemical process monitoring and dynamic process modeling. They defined chemometrics as the science of relating measurements made on a chemical system to the state of the system via application of mathematical or statistical methods.

PCA is a tool for data compression and information extraction. PCA finds linear combinations of variables that describe major trends in a data set. Mathematically, PCA relies on the eigenvalue decomposition of the covariance or correlation matrix of the process data. For monitoring a process with PCA, control limits are set for two kinds of statistics, T^2 and Q. The T^2 statistic is the sum of normalized squared scores, and it is a measure of the variation within the PCA model. On the other hand, the Q statistic, which is the sum of squared residuals, is a measure of the amount of variation not captured by the PCA model. The residual is defined by the difference between the original data and data reconstructed with several principal components. The Q statistic based on the residual analysis was investigated by Jackson and Mudholkar (1979) and Jackson (1980).

Chemometric techniques are very useful for modeling and monitoring chemical processes, because a large number of variables are measured and are highly correlated. Kresta et al. (1991) demonstrated the usefulness of MSPC with applications to simulated data from a fluidized-bed reactor and an extractive distillation column. With multiway PCA and PLS (Wold et al., 1987), the MSPC functioning was extended to monitor time-varying batch processes. Multiway PCA was applied to simulated data obtained from a semibatch reactor (Nomikos and MacGregor, 1994) and also process data collected from an industrial batch polymerization reactor (Nomikos and MacGregor, 1995). Wise et al. (1999) compared PCA, multiway PCA, trilinear decomposition, and parallel factor analysis with their applications to a semiconductor etch process. Another extension to handle very large processes via multiblock PCA and PLS was made by MacGregor et al. (1994). Ku et al. (1995) used dynamic PCA to include process dynamics in a PCA model. Although PCA has shown its usefulness as a MSPC tool, few researchers have studied the theoretical basis of using PCA for monitoring dynamic processes. Wise et al. (1990) have shown that PCA can facilitate monitoring for processes with more measurements than states. PCA has also been used for faulty sensor identification (Dunia et al., 1996) and disturbance diagnosis (Raich and Cinar, 1996).

Many successful applications have shown the practicability of MSPC. However, the conventional MSPC method described earlier does not always function well, because it cannot detect the change of correlation among process variables as long as both T^2 and Q statistics are inside the control limits.

In the present work, a novel statistical process monitoring method is proposed in order to improve the performance of process monitoring. The method is based on the idea that a change of operating condition can be detected by monitoring a distribution of time-series data, which reflects the corresponding operating condition. To quantitatively evaluate the difference between two data sets, a new index representing the dissimilarity of process data is introduced. The performance of the proposed monitoring method and that of the conventional method are compared with their applications to simulated data collected from a 2×2 process and the Tennessee Eastman process.

Monitoring Based on Dissimilarity of Data Sets

The proposed monitoring method is based on the idea that a change of operating condition can be detected by monitoring the distribution of process data, because the distribution reflects the corresponding operating condition. In this section, a dissimilarity index is introduced in order to quantitatively evaluate the difference between distributions of process data. Then the monitoring method with the dissimilarity index is described.

Dissimilarity index

The concept of similarity or dissimilarity is often used for classifying a set of data. In cluster analysis, for example, the dissimilarity between two classes is measured by the difference between barycenters of the data, and two classes with the smallest degree of dissimilarity are combined to generate a new class.

To evaluate the difference between distributions of data sets, a classification method based on the Karhunen−Loeve (KL) expansion (Fukunaga and Koontz, 1970) is used in this work. The KL expansion is a well-known technique for feature extraction or dimensionality reduction in the pattern-recognition area, and it is mathematically equivalent to PCA. Therefore, both the proposed monitoring method and the conventional MSPC method are based on the same multivariate analysis technique. Those two methods use PCA to capture a correlation structure of process data, but they monitor different indexes on the basis of different concepts.

Consider the following two data sets:

$$X_i = \begin{bmatrix} x_{11}^{(i)} & x_{12}^{(i)} & \cdots & x_{1P}^{(i)} \\ x_{21}^{(i)} & x_{22}^{(i)} & \cdots & x_{2P}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N_i 1}^{(i)} & x_{N_i 2}^{(i)} & \cdots & x_{N_i P}^{(i)} \end{bmatrix}, \quad i = 1, 2 \tag{2}$$

where N_i is the number of samples of ith data set $\boldsymbol{X}_i$ and P is the number of variables. Here, each column of $\boldsymbol{X}_i$ is assumed to be mean-centered. The covariance matrices are given by

$$\boldsymbol{R}_i = \frac{1}{N_i - 1}\boldsymbol{X}_i^T\boldsymbol{X}_i \tag{3}$$

and the covariance matrix of the mixture of both data sets is given by

$$\begin{aligned}\boldsymbol{R} &= \frac{1}{N-1}\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix}^T\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix}\\ &= \frac{N_1-1}{N-1}\boldsymbol{R}_1 + \frac{N_2-1}{N-1}\boldsymbol{R}_2.\end{aligned} \tag{4}$$

By application of eigenvalue decomposition to $\boldsymbol{R}$, an orthogonal matrix $\boldsymbol{P}_0$ that satisfies

$$\boldsymbol{R}\boldsymbol{P}_0 = \boldsymbol{P}_0\Lambda \tag{5}$$

is derived. Here, Λ is a diagonal matrix whose diagonal elements are eigenvalues of $\boldsymbol{R}$. With a transformation matrix $\boldsymbol{P}$ defined as

$$\boldsymbol{P} = \boldsymbol{P}_0\Lambda^{-1/2} \tag{6}$$

the following equation can be derived.

$$\boldsymbol{P}^T\boldsymbol{R}\boldsymbol{P} = \boldsymbol{I} \tag{7}$$

When the data matrices $\boldsymbol{X}_i$ are transformed as

$$\boldsymbol{Y}_i = \sqrt{\frac{N_i-1}{N-1}}\,\boldsymbol{X}_i\boldsymbol{P}_0\Lambda^{-1/2} \tag{8}$$

the covariance matrices of the transformed data matrices

$$\begin{aligned}\boldsymbol{S}_i &= \frac{1}{N_i-1}\boldsymbol{Y}_i^T\boldsymbol{Y}_i\\ &= \frac{N_i-1}{N-1}\boldsymbol{P}^T\boldsymbol{R}_i\boldsymbol{P}\end{aligned} \tag{9}$$

satisfying the following equation:

$$\boldsymbol{S}_1 + \boldsymbol{S}_2 = \boldsymbol{I} \tag{10}$$

By application of eigenvalue decomposition to the covariance matrices

$$\boldsymbol{S}_i\boldsymbol{w}_j^{(i)} = \lambda_j^{(i)}\boldsymbol{w}_j^{(i)} \tag{11}$$

is derived. From Eqs. 10 and 11

$$\boldsymbol{S}_2\boldsymbol{w}_j^{(1)} = \left(1 - \lambda_j^{(1)}\right)\boldsymbol{w}_j^{(1)} \tag{12}$$

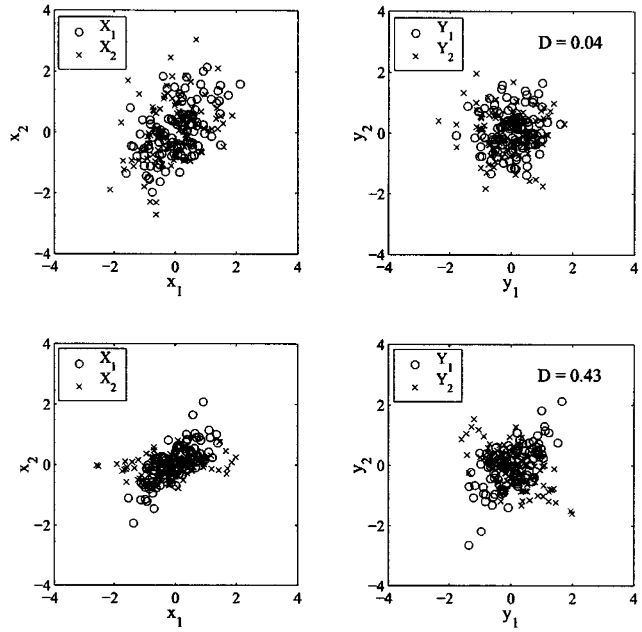

Figure 1. Linear transformation of data sets (from X_i to Y_i) for evaluating the dissimilarity.

The dissimilarity index D is 0.04 when data sets are similar (top). The dissimilarity index D is 0.43 when data sets are different (bottom).

can be derived. This equation means that the eigenvectors of $\boldsymbol{S}_2$ are the same as those of $\boldsymbol{S}_1$ and that the following relationship is satisfied

$$1 - \lambda_j^{(1)} = \lambda_j^{(2)} \tag{13}$$

As a result, since eigenvectors of a covariance matrix represent directions of principal components and eigenvalues are equivalent to the variances of principal components, both transformed data sets have the same set of principal components and the corresponding eigenvalues of the covariance matrices are oppositely ordered.

Thus, after the preceding transformation, the most important correlation for data set 1 is equivalent to the least important correlation for data set 2, and vice versa. When data sets are quite similar to each other, the eigenvalues $\lambda_j^{(i)}$ must be near 0.5. On the other hand, when data sets are quite different from each other, the largest and the smallest eigenvalues should be near one and zero, respectively. This mathematical explanation can be physically understood from Figure 1. Finally, the following index D is defined for evaluating the dissimilarity of data sets:

$$D = \frac{4}{P}\sum_{j=1}^{P}\left(\lambda_j - 0.5\right)^2 \tag{14}$$

The dissimilarity index D changes between zero and one. When two data sets are similar to each other, D must be near zero. However, D should be near one when data sets are quite different from each other. For example, the dissimilarity indexes in two cases are shown in Figure 1. When data sets are similar to each other, $\lambda_j^{(1)} = \{0.5091, 0.3646\}$, and as a result $D = 0.04$. On the other hand, when data sets are different from each other, $\lambda_j^{(1)} = \{0.8320, 0.1758\}$, and as a result $D = 0.43$. These examples show the validity of the dissimilarity index for quantitatively evaluating the difference between two data sets.

To detect changes in the distribution of process data, a reference distribution representing a normal operating condition should be defined, and the dissimilarity index between the reference distribution and a distribution representing the current operating condition should be monitored.

Monitoring method

A reference data set and a control limit must be determined in order to apply the proposed monitoring method. The following procedure is adopted.

(1) Acquire time-series data when a process is operated under a normal condition. Normalize each column (variable) of the data matrix, that is, adjust it to zero mean and unit variance.

(2) Determine the size (steps) of time window, w. Generate data sets with w samples from the data by moving the time window. Then, select a reference data set.

(3) Calculate the dissimilarity index D, and determine the control limit.

For on-line monitoring, the data matrix representing the current operating condition is updated by moving the time-window step by step, and it is scaled with the mean and the variance obtained at step 1. Then, the index D is calculated at each step. If the index is outside the control limit, the process is judged to be out of control.

Application 1: 2×2 Process

In this section, the proposed monitoring method as well as the conventional MSPC method are applied to monitoring problems of a simple 2×2 process. The simple process is used to obtain statistically meaningful results and for a fair comparison.

Example system

For simplicity, the following multivariate process (Ku et al., 1995) is considered

$$\boldsymbol{x}(t) = \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix} \boldsymbol{x}(t-1) + \begin{bmatrix} 1.0 & 2.0 \\ 3.0 & -4.0 \end{bmatrix} \boldsymbol{u}(t-1) \tag{15}$$

$$\boldsymbol{y}(t) = \boldsymbol{x}(t) + \boldsymbol{v}(t) \tag{16}$$

where $\boldsymbol{u}$ is the correlated input:

$$\boldsymbol{u}(t) = \begin{bmatrix} 0.811 & -0.266 \\ 0.477 & 0.415 \end{bmatrix} \boldsymbol{u}(t-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix} \boldsymbol{w}(t-1) \tag{17}$$

The inputs w_i in $\boldsymbol{w}$ are uncorrelated Gaussian signals with zero mean and unit variance. The output y_i in $\boldsymbol{y}$ is corrupted by uncorrelated Gaussian errors with zero mean and variance 0.1. The input $\boldsymbol{u}$ and the output $\boldsymbol{y}$ are measured, and those measurements are used for monitoring.

Table 1. Settings of Abnormal Conditions

Case	Type	Size
0	Normal condition	—
1	Mean shift of w_1	$0.0 \to 0.5$
2	Mean shift of w_1	$0.0 \to 1.0$
3	Mean shift of w_1	$0.0 \to 1.5$
4	Mean shift of w_1	$0.0 \to 2.0$
5	Mean shift of w_1	$0.0 \to 3.0$
6	Change of parameter from u_1 to x_2	$3.0 \to 2.5$
7	Change of parameter from u_1 to x_2	$3.0 \to 2.0$
8	Change of parameter from u_1 to x_2	$3.0 \to 1.0$

To compare the performance of the monitoring methods, mean shifts of w_1 and changes of a coefficient from u_1 to x_2 are investigated. The settings of those abnormal conditions are summarized in Table 1. One data set obtained from the normal operating condition was used to build a PCA model for the conventional MSPC and also to determine a reference data set for the proposed method. In addition, 200 data sets were used to determine control limits of monitored indexes. Different data sets were generated by changing seeds of the white noise $\boldsymbol{w}$, after which 1000 data sets were generated at each case shown in Table 1. One of 1000 realizations in cases 4 and 8 are shown in Figures 2 and 3, respectively. From these figures, it is confirmed that very small faults are investigated in this application.

Settings for monitoring

The control limit of each index is determined so that the number of samples outside the control limit is 1% of the entire samples while the process is operated under a normal condition. Therefore, the control limits represent 99% confidence limits without any assumptions of probability distributions of sample statistics.

On the basis of these control limits, each monitoring method is evaluated by the following steps.

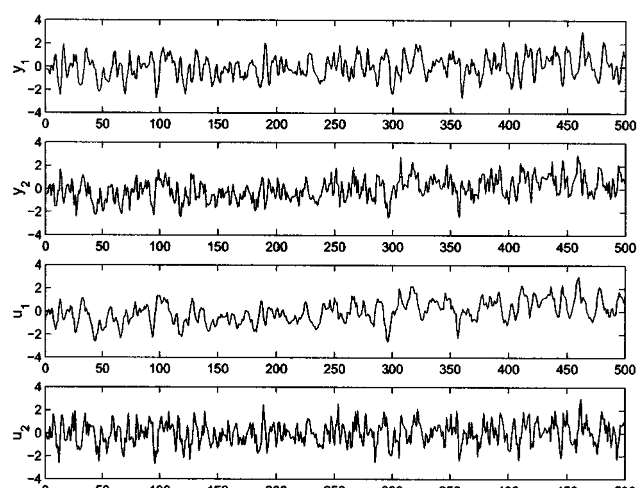

Figure 2. Time-series data in case 4.

The mean shift occurs at the 301st step.

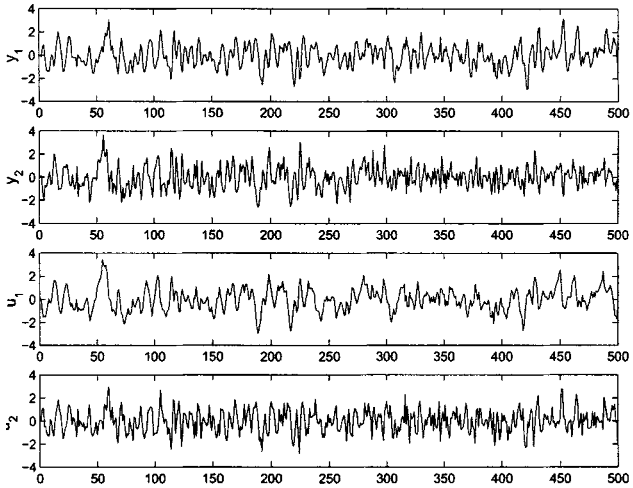

Figure 3. Time-series data in case 8.

The parameter change occurs at the 301st step.

(1) Each monitoring method is applied to the data in cases 0–8, and each index is calculated.

(2) For the data obtained after the occurrence of a fault, the percentage of the samples outside the control limit is calculated in each simulation. This percentage is termed "reliability" in the present work and is used for evaluation of each monitoring method.

(3) The average reliability of 1,000 simulations is calculated in each case.

Since the control limits are determined so that they represent 99% confidence limits, a monitoring method is regarded as successful in detecting the abnormal condition if the reliability is considerably higher than 1%. When the reliability is less than or close to 1%, however, the monitoring method is regarded as not functioning well. The reliability is affected by the number of samples used for calculating it. In the present work, 100 samples are used for calculating the reliability. Therefore, for example, when the reliability is 60%, the average run length (ARL) will be less than 41. ARL is a well-known index for evaluating the fault-detection performance (Montgomery, 1997).

Monitoring results and discussion

In the following sections, the conventional MSPC method based on PCA is referred to as cMSPC. The monitored indexes for cMSPC are T_i^2 and Q_i. The subscript i denotes the number of principal components retained in a PCA model. In addition, the proposed method is referred to as DISSIM, and the monitored index for DISSIM is the dissimilarity index D.

Static Monitoring. The vector of variables for static monitoring at step k is the following:

$$\left[\boldsymbol{y}^T(k) \quad \boldsymbol{u}^T(k) \right] \tag{18}$$

In this article, this type of monitoring is called static monitoring to differentiate it from another monitoring, in which past measurements as well as current measurements are used as monitoring variables. The static monitoring results of cMSPC are summarized in Table 2. The eigenvalues of the covariance matrix, which are the variances of principal components, are 1.86, 1.44, 0.63, and 0.06. Therefore, two or three principal components should be retained in the PCA model. As shown in Table 2, the best monitoring performance is achieved when three principal components are used. As expected, the reliability of cMSPC increases as the size of the mean shift or parameter change increases. In almost all cases, however, the reliability measures of cMSPC are less than 10%. These results indicate that cMSPC does not function well for the changes investigated here. In addition, since the reliability measures in case 0 are very close to 1%, the control limits are successfully determined.

Table 2. Reliability (%) of Static cMSPC: Applications to 2×2 Process

Index	Case w	0	1	2	3	4	5	6	7	8
T_2^2	—	1.1	1.3	2.1	3.5	6.5	16.9	1.0	0.9	1.1
Q_2	—	1.1	1.2	1.7	2.7	4.5	9.7	1.4	1.6	2.6
T_3^2	—	1.1	1.3	2.4	4.3	8.5	23.0	1.2	1.3	2.0
Q_3	—	1.0	1.0	1.1	1.2	1.6	2.3	1.6	3.2	9.5

The static monitoring results of DISSIM are summarized in Table 3. The design parameter of DISSIM is the time-window size. The optimal time-window size depends on the characteristics of abnormal conditions. For mean shifts and parameter changes, the best reliability is achieved when $w = 200$ and $w = 50$, respectively. The optimal size of time window also depends on the sampling interval of measurements and the number of monitored variables. To obtain a data set representing the operating conditions appropriately, one needs to use a sufficient number of samples. On the other hand, using an excessively large time window can result in the reduced speed of detecting changes in operating conditions. The reliability of DISSIM does not always increase as the size of the mean shift increases. This would be mainly because the index D is autocorrelated, and therefore it is more difficult to compute the reliability of DISSIM than cMSPC. In fact, the reliability of DISSIM in case 0 changes depending on the time-window size, although that of cMSPC is almost 1.0%. Running more simulations will make results more consistent, but the trend can be confirmed from Tables 2 and 3.

The results show that DISSIM can outperform cMSPC. For example, the reliability measures of DISSIM with $w = 100$ are considerably better than those of cMSPC using three principal components in cases 4, 5, and 8 for both mean shifts and parameter changes. It should be noted here that the reliability of DISSIM in case 0 is only 0.6%. This means that the

Table 3. Reliability (%) of Static DISSIM: Applications to 2×2 Process

Index	Case w	0	1	2	3	4	5	6	7	8
D	50	1.6	1.4	1.0	1.5	5.6	40.7	2.1	3.5	50.4
	100	0.6	0.3	1.3	4.0	18.0	57.6	0.9	2.0	16.2
	200	1.2	1.2	2.5	7.8	22.4	57.6	1.7	1.3	8.4

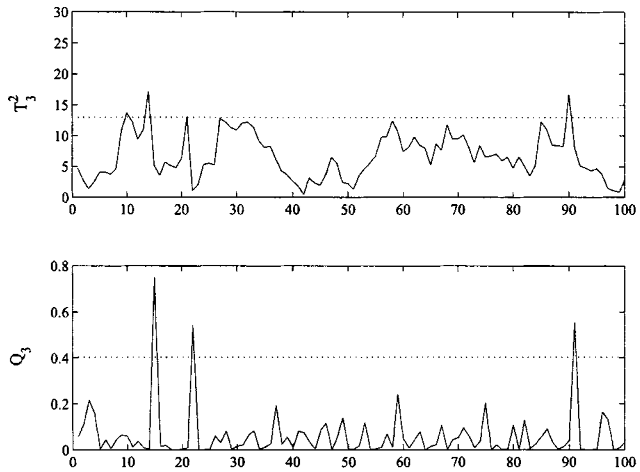

Figure 4. Monitoring results of the 2×2 process with static cMSPC in case 4.

T^2 and Q plots with the control limits representing 99% confidence limits (dotted line). Three principal components are retained in the PCA model.

control limit of D is too relaxed. If the control limit is tightened, the reliability of DISSIM with $w = 100$ increases and the superiority of DISSIM over cMSPC becomes clearer. On the other hand, this result indicates the difficulty of determining the control limit of D in comparison with T^2 or Q. The small changes, which are difficult to detect by cMSPC, can be detected by DISSIM because such abnormal operating conditions investigated in this application affect the correlation among variables and change the distribution of data. For changes to be detected in the process with cMSPC, T^2 or Q must exceed the control limit. In other words, changes in the correlation of process variables cannot be detected by cMSPC as long as measurements are inside the control envelope defined by the PCA model. DISSIM functions well even when the variation of data is decreased.

Since the reliability measures in case 0 are close to 1%, the control limits are successfully determined so that the probability of false alarms is almost 1% under normal operating conditions. Therefore, the preceding comparison based on the reliability is sufficiently fair.

One of the 1,000 monitoring results using cMSPC and DISSIM in case 4 is shown in Figures 4 and 5, respectively. The control limits, representing 99% confidence limits, are also shown in these figures. The mean shift occurs at the first

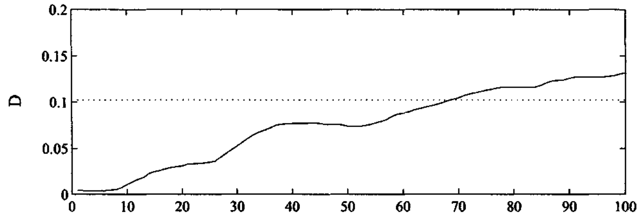

Figure 5. Monitoring results of the 2×2 process with DISSIM in case 4.

D plot with the control limit representing 99% confidence limit (dotted line). The time-window size is 100 steps.

step, and it can be detected by cMSPC much earlier than by DISSIM. However, the abnormal condition is not obvious in Figure 4, because the indexes, T_3^2 and Q_3, are below the control limits at almost all points. On the other hand, the dissimilarity index D increases gradually, and the occurrence of an abnormal condition is not in doubt after the 70th step. In general, the dissimilarity index D changes gradually in comparison with T^2 or Q, because a time window is used for calculating D. The use of a time window causes such a smoothing effect on the monitored index and can improve the performance of DISSIM, but it can also reduce the speed of detecting changes in the operating conditions. Therefore, the appropriate selection of time-window size is important for the effective functioning of DISSIM.

Dynamic Monitoring. In general, static monitoring does not function well for autocorrelated data. One useful approach for dealing with autocorrelated data involves identifying a time-series model such as the autoregressive moving-average (ARMA) model, using the model to remove the autocorrelation from the data, and applying SPC charts to the residuals. Using past measurements as monitored variables is also useful for capturing the correlation among process variables, because the dynamics can be taken into account (Ku et al., 1995). In this section, the measurements at one step before are used as monitored variables. The vector of process variables for dynamic monitoring at step k becomes the following:

$$\left[\boldsymbol{y}^T(k-1) \quad \boldsymbol{u}^T(k-1) \quad \boldsymbol{y}^T(k) \quad \boldsymbol{u}^T(k) \right] \tag{19}$$

A more general description of the vector for dynamic monitoring can be used. For example, if z is a column vector consisting of monitored variables

$$\left[z^T(k) \quad z^T(k-s_1) \quad z^T(k-s_2) \quad \cdots \right] \tag{20}$$

is a general description. For the best monitoring performance, s_i must be optimized. In this simple example, however, the use of measurements at one step before is found to be sufficient.

The dynamic monitoring results of cMSPC are summarized in Table 4. The best monitoring performance is achieved when six or seven principal components are used. The detection performance, that is, the reliability, of dynamic monitoring is considerably better than that of static monitoring shown in Table 2. For example, in case 8, the reliability of static cMSPC with Q_3 is only 9.5%, but the reliability of dynamic cMSPC with Q_6 is 65.5%. Therefore, the superiority of dynamic monitoring over static monitoring is obvious.

The dynamic monitoring results of DISSIM are summarized in Table 5. The performance of dynamic DISSIM is considerably better than that of static DISSIM. In particular, dynamic DISSIM can detect parameter changes quite successfully. In addition, the performance of dynamic DISSIM is considerably better than that of dynamic cMSPC. These results clearly show the advantages of the proposed method. The reason for the poorer performance of cMSPC is not because the reliability indexes of T^2 and Q are calculated independently. If one combined reliability value is calculated, then the control limits of both indexes should be relaxed to keep

Table 4. Reliability (%) of Dynamic cMSPC: Applications to 2×2 Process

Index	Case w	0	1	2	3	4	5	6	7	8
T_5^2	—	1.1	1.2	2.1	4.0	8.4	24.5	1.3	2.0	5.8
Q_5	—	1.1	1.0	1.4	1.8	2.7	6.3	3.5	17.4	46.9
T_6^2	—	1.1	1.1	1.9	3.7	7.8	23.7	1.6	3.8	14.2
Q_6	—	1.0	1.4	2.9	5.9	12.1	36.4	13.6	39.3	65.5
T_7^2	—	1.0	1.1	2.0	4.3	9.5	30.9	6.9	24.7	53.6
Q_7	—	1.0	1.7	3.7	7.9	15.2	40.2	4.7	18.0	45.1

Table 5. Reliability (%) of Dynamic DISSIM: Applications to 2×2 Process

Index	Case w	0	1	2	3	4	5	6	7	8
D	50	1.2	1.9	3.1	10.0	22.8	54.2	54.1	77.2	88.9
	100	0.7	1.1	1.9	11.6	35.6	72.1	39.1	73.4	89.7
	200	0.8	1.1	7.5	37.9	62.7	81.7	34.2	74.7	90.6

the combined reliability in the case 0 to be 1.0%. This relaxation will decrease the combined reliability in other cases.

One of the 1,000 monitoring results using cMSPC and DISSIM in case 6 is shown in Figures 6 and 7, respectively. The control limits are also shown in these figures. The parameter change occurs at the first step. It takes less than 30 steps for cMSPC to detect the change, and the detection using cMSPC is earlier than that using DISSIM. However, the indexes, T_6^2 and Q_6, are below the control limits at many steps. In particular, after the 55th step, cMSPC does not detect the parameter change. On the other hand, the dissimilarity index D of DISSIM is steadily increasing, that is, DISSIM can successfully capture the small change in the system.

One of the major differences between DISSIM and cMSPC is the number of samples to be used for calculating indexes. Only one current sample is used in cMSPC. In DISSIM, however, the number of samples are equivalent to the time-window size. Therefore, DISSIM is more robust against measurement noise than cMSPC. This is analogous to the relationship between Shewhart control charts and EWMA control charts. An EWMA control chart uses more samples, is more robust against noise, and functions better in detecting a small fault than a Shewhart control chart. In addition, DISSIM suffers from the delay because DISSIM has a smoothing effect caused by the use of a time window. It should be noted here, however, that it does not take longer than time-window length for DISSIM to detect faults or disturbances. In fact, the dynamic version of DISSIM takes only about 50 steps to detect the fault in case 6, as shown in Figure 7. The delay is smaller than the time-window size of 200 steps.

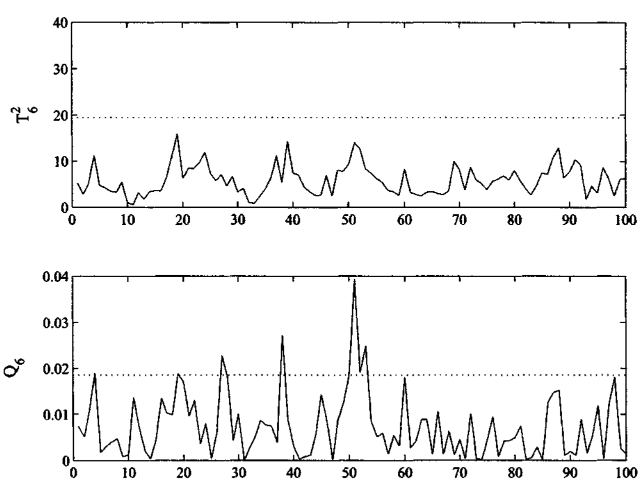

Figure 6. Monitoring results of the 2×2 process with dynamic cMSPC in case 6.

T^2 and Q plots with the control limits representing 99% confidence limits (dotted line). Six principal components are retained in the PCA model.

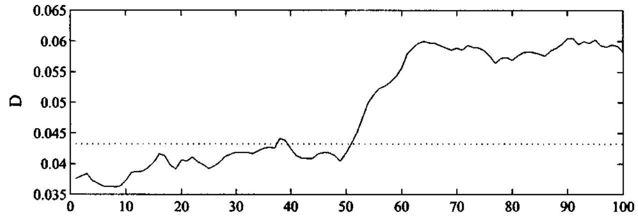

Figure 7. Monitoring results of the 2×2 process with dynamic DISSIM in case 6.

D plot with the control limits representing 99% confidence limit (dotted line). The time-window size is 200 steps.

Application 2: Tennessee Eastman Process

In this section, the proposed monitoring method as well as the conventional MSPC method are applied to the monitoring problems of the Tennessee Eastman process for investigating the practicability of DISSIM with application to a realistic chemical process.

Tennessee Eastman process

The simulator of the Tennessee Eastman process was developed by Downs and Vogel (1993). The process consists of five major unit operations: a reactor, a product condenser, a vapor–liquid separator, a recycle compressor, and a product stripper. Two products are produced by two additional exothermic reactions. The process has 12 manipulated variables, 22 continuous process measurements, and 19 composition measurements sampled less frequently. The simulator includes the set of programmed disturbances listed in Table 6. The control system used for dynamic simulations is the decentralized PID control system designed by McAvoy and Ye (1994), which is shown in Figure 8. The sampling interval of process variables is set at 3 min. A total of 16 variables, selected by Chen and McAvoy (1998) for monitoring purposes, are used for monitoring in this study. Those 16 variables are listed in Table 7.

Monitoring results and discussion

Several disturbances can easily be detected by univariate SPC, that is, by monitoring each measured variable independently. For example, the reliability of univariate SPC reached

Table 6. Process Disturbances for the Tennessee Eastman Process

Case	Disturbance	Type
1	A/C feed ratio, B composition constant	Step
2	B composition, A/C ratio constant	Step
3	D feed temperature	Step
4	Reactor cooling-water inlet temperature	Step
5	Condenser cooling-water inlet temperature	Step
6	A feed loss	Step
7	C header pressure loss—reduced availability	Step
8	A, B, C feed composition	Random variation
9	D feed temperature	Random variation
10	C feed temperature	Random variation
11	Reactor cooling-water inlet temperature	Random variation
12	Condenser cooling-water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling-water valve	Sticking
15	Condenser cooling-water valve	Sticking
16–20	Unknown	Unknown

99.2% in case 1 and 95.5% in case 2. Although it is very easy to detect such disturbances with other monitoring methods, the present work investigates disturbances that are relatively difficult to detect by univariate SPC.

All monitoring methods are unsuccessful in detecting the events in cases 4, 5, 12, and 15, because the disturbances or the faults in these cases can be easily compensated by the control system. The cascade control system can effectively compensate for these four upsets (McAvoy and Ye, 1994). Since the monitored variables listed in Table 7 do not include any manipulated variables, disturbances or faults cannot be detected while the control system functions quite well. For detecting the change in operating conditions in cases 4, 5, 12, and 15, manipulated variables should be used for monitoring. For example, the step change of reactor cooling-water inlet temperature in case 4 can be easily compensated by controlling the reactor cooling-water flow. Actually, a large shift in the reactor cooling-water flow indicates the possibility of a change in the reactor cooling-water inlet temperature. In addition, the monitoring performance might be improved by the use of measurements obtained from product analyzers as monitored variables.

Table 7. Process Variables Used for Monitoring

A feed
D feed
E feed
A and C feed
Recycle flow
Reactor feed rate
Reactor temperature
Purge rate
Product separator temperature
Product separator pressure
Product separator underflow
Stripper pressure
Stripper temperature
Stripper steam flow
Reactor cooling-water outlet temperature
Separator cooling-water outlet temperature

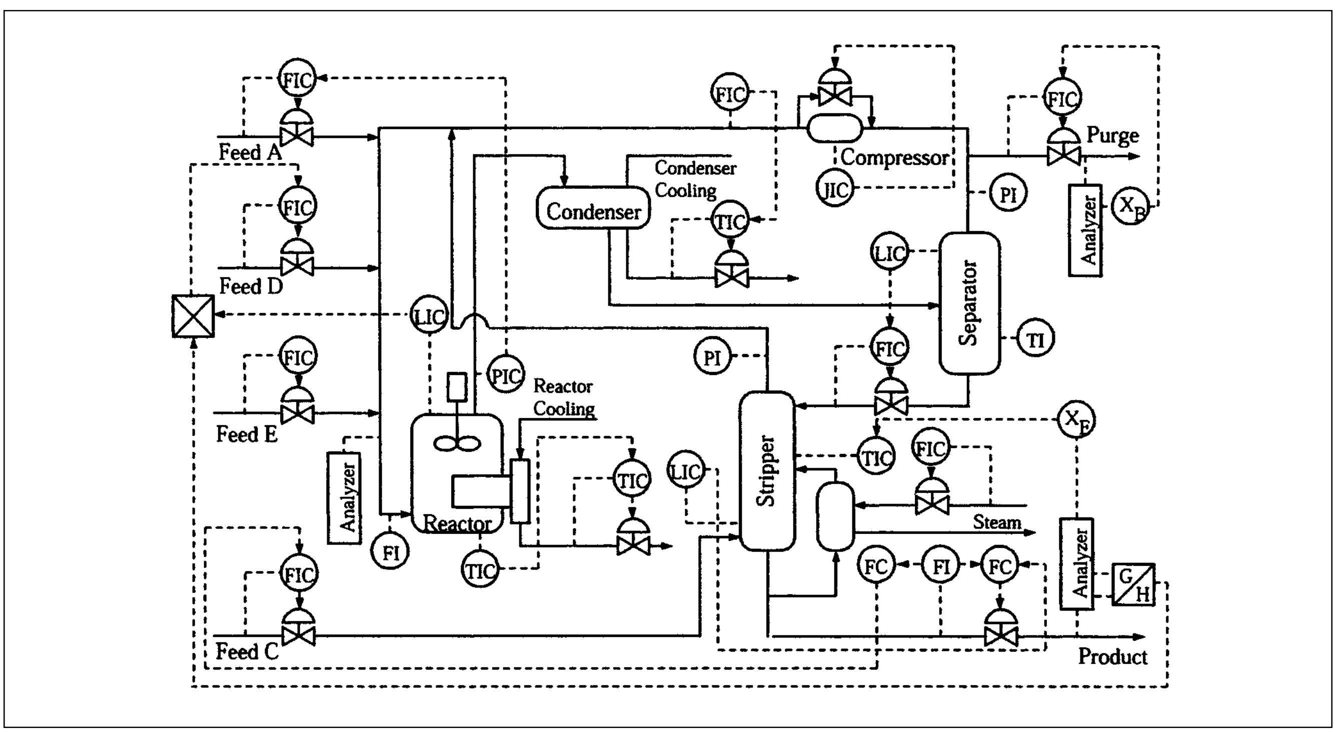

Figure 8. Control system of the Tennessee Eastman process.

Table 8. Reliability (%) of cMSPC: Applications to the Tennessee Eastman Process

Index	Case w	3	8	9	10	11	14	16	19	20
T_6^2	—	1.8	84.2	2.6	74.2	1.3	7.0	8.4	6.5	66.7
Q_6	—	2.7	86.1	2.9	80.0	3.6	42.3	9.2	5.3	66.9
T_{11}^2	—	1.7	84.6	3.1	76.5	0.8	6.2	9.0	10.3	67.4
Q_{11}	—	18.9	88.2	14.9	81.2	20.4	70.8	21.9	11.3	69.1

The static monitoring results of cMSPC are summarized in Table 8. The reliability was calculated by using 100 samples after the occurrence of a disturbance, and it is an average of 10 different simulations.

The number of principal components must be determined for cMSPC. Many procedures have been proposed for selecting the number of principal components to be retained in a PCA model (Jackson, 1991). For example, a practical method is to retain principal components with corresponding eigenvalues greater than one when data are normalized. In this application, six eigenvalues of the correlation matrix are greater than one. In addition, 68% of the variance in the data can be explained by the first 6 principal components, and 94% of the variance can be explained by 11 principal components. The results in Table 8 show that the practical procedure, by which 6 principal components are retained, does not function well in some situations for monitoring purposes. The reliability of cMSPC with 11 principal components is better than that of cMSPC with 6 principal components. In this application, the best performance was obtained by retaining 11 principal components in the PCA model.

The static monitoring results of DISSIM are summarized in Table 9. The time-window size strongly affects the monitoring performance of DISSIM. In this application, DISSIM with relatively small time windows, $w = 100$ and $w = 200$, does not function well, especially in cases 3, 9, and 11. In these cases, the reliability measures of DISSIM with small time windows are worse than those of cMSPC. Therefore, the time-window size should be more than 200 steps. The best performance of DISSIM is achieved when $w = 300$. The reliability of DISSIM is considerably better than that of cMSPC in cases 3, 9, 11, 14, 16, and 19. In other cases, the reliability of cMSPC and that of DISSIM are comparable to each other.

For example, in case 3 where the D feed temperature changes stepwise, the disturbance can be easily compensated by the control system, as reported by McAvoy and Ye (1994), and the deterministic changes of the monitored variables are not significant. As a result, it is difficult to detect the change in operating conditions by cMSPC, although the step change of the D feed temperature causes a small mean shift in both flow and outlet temperature of the reactor cooling water

Table 9. Reliability (%) of DISSIM: Applications to the Tennessee Eastman Process

Index	Case w	3	8	9	10	11	14	16	19	20
D	100	10.0	81.1	1.6	74.2	3.6	35.1	35.3	21.4	63.4
	200	0.0	74.2	0.0	59.7	0.0	0.0	3.9	0.0	50.8
	300	58.1	84.9	28.8	79.7	49.4	92.5	60.9	52.9	68.5

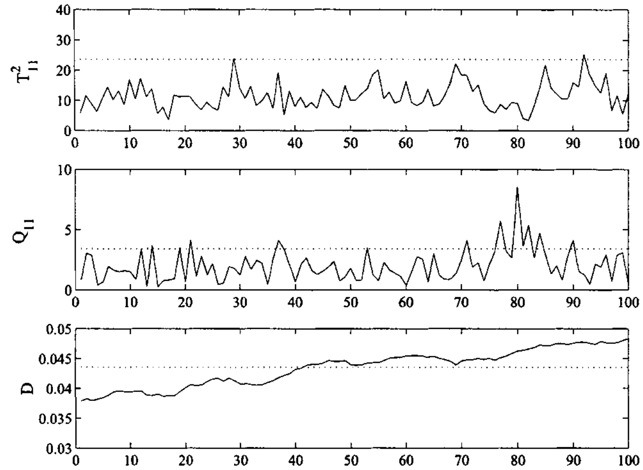

Figure 9. Monitoring results of the Tennessee Eastman process with cMSPC and DISSIM in case 3.

T_{11}^2 and Q_{11} of cMSPC and D of DISSIM are shown from the top (solid line) with their control limits (dotted limits).

through the function of the cascade control system for reactor temperature. However, DISSIM can detect the disturbance successfully. These results have clearly shown the advantages of DISSIM over cMSPC. DISSIM can detect the change of operating conditions even when faulty data are inside the control envelope defined by the PCA model, because it focuses on the distribution of process data. Figure 9 shows one of the 10 monitoring results.

For further improvement of the monitoring performance, dynamic monitoring can be used. The advantage of using dynamic monitoring for the Tennessee Eastman process was investigated by Kano et al. (2000a).

Conclusions

A novel statistical process monitoring method has been proposed in order to improve the performance of process monitoring. The proposed MSPC method is based in the idea that a change in operating condition can be detected by monitoring the distribution of process data, which reflects the corresponding operating condition. This method, termed DISSIM, is based on the dissimilarity of data sets. The dissimilarity index was introduced to quantitatively evaluate the difference between data sets.

The proposed monitoring method and the conventional MSPC method were applied to monitoring problems of a 2×2 process and the Tennessee Eastman process. The results show that the reliability of DISSIM is considerably better than that of cMSPC in many cases if the time-window size is appropriately selected. DISSIM can detect changes in operating conditions even when the deterministic changes in the monitored variables are not significant and the variances are not increased, because DISSIM monitors the distribution of process data. However, since DISSIM has a smoothing effect, which is caused by the use of a time window, DISSIM suffers from the delay. It should be noted, however, that the speed and reliability of fault detection can be adjusted by changing the time-window size. The selection of an appropriate time-window size is crucial for the effective functioning of DISSIM. The time-window size was selected by trial and error in this study. Since such a selection method is not practical, especially when abnormal data are not available, a systematic procedure for selecting the time-window size needs to be developed. Another remaining issue is theoretical analysis of the proposed dissimilarity index. From a practical viewpoint, it is important to develop a theoretical basis for selecting an appropriate control limit of the index D with less data. Furthermore, dynamic monitoring is quite effective for taking process dynamics into account and for coping with autocorrelated data. The performance of dynamic monitoring is considerably better than that of static monitoring.

For example, the fault-detection performance of cMSPC can be improved by using an EWMA-type approach. Such a technique will make cMSPC robust against measurement noise and also sensitive to small disturbances or faults. DISSIM has such advantages in nature because a time window is used. In other words, DISSIM is good at detecting small and slow changes in general. In addition, it should be emphasized that DISSIM can detect changes in correlation structure even when variation of data is decreased.

The present work focuses only on fault detection. However, the concept of dissimilarity or similarity also can be used for fault identification and diagnosis (Kano et al., 2000b). For example, when process data obtained from several past faulty operating conditions are available, a fault can be identified by comparing the similarity between data sets representing faulty operating conditions in the past and the data set representing an operating condition when the fault is detected. This study has shown the superiority, and practicability, of DISSIM over conventional MSPC. In addition, as described earlier, DISSIM can be used for fault identification and diagnosis. Therefore, a process monitoring system based on the dissimilarity of process data, which can detect and identify faults, would be promising.

Acknowledgments

The authors gratefully acknowlege the financial support from the Japan Society for the Promotion of Science (JSPS-RFTF96R14301). In addition, this research was partically supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Encouragement of Young Scientists (11750650).

Literature Cited

Chen, G., and T. J. McAvoy, "Predictive On-Line Monitoring of Continuous Processes," *J. Process Control*, **8**, 409 (1998).

Downs, J. J., and E. F. Vogel, "A Plant-Wide Industrial Process Control Problem," *Comput. Chem. Eng.*, **17**, 245 (1993).

Dunia, R., S. J. Qin, T. F. Edgar, and T. J. McAvoy, "Identification of Faulty Sensors Using Principal Component Analysis," *AIChE J.*, **42**, 2797 (1996).

Fukunaga, K., and W. L. G. Koontz, "Application of the Karhunen-Loeve Expansion to Feature Selection and Ordering," *IEEE Trans. Comput.*, **C-19**, 311 (1970).

Geladi, P., and B. R. Kowalski, "Partial Least-Squares Regression: A Tutorial," *Anal. Chim. Acta*, **185**, 1 (1986).

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

Jackson, J. E., "Principal Components and Factor Analysis: Part I—Principal Components," *J. Quality Technol.*, **12**, 201 (1980).

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

Kano, M., K. Nagao, H. Ohno, S. Hasebe, and I. Hashimoto, "Dissimilarity of Process Data for Statistical Process Monitoring," *Preprints of IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM)*, Vol. I, Pisa, Italy, p. 231 (2000a).

Kano, M., H. Ohno, S. Hasebe, I. Hashimoto, R. Strauss, and B. R. Bakshi, "Contribution Plots for Fault Identification Based on the Dissimilarity of Process Data," AIChE Meeting, Los Angeles, CA (2000b).

Kresta, J. V., J. F. MacGregor, and T. E. Marlin, "Multivariate 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 Component Analysis," *Chemometrics Intelligent Lab. Syst.*, **30**, 179 (1995).

MacGregor, J. F., C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process Monitoring and Diagnosis by Multiblock Methods," *AIChE J.*, **40**, 826 (1994).

McAvoy, T. J., and N. Ye, "Base Control for the Tennessee Eastman Problem," *Comput. Chem. Eng.*, **18**, 383 (1994).

Montgomery, D. C., *Introduction to Statistical Quality Control*, Wiley, New York (1997).

Nomikos, P., and J. F. MacGregor, "Monitoring Batch Processes Using Multiway Principal Component Analysis," *AIChE J.*, **40**, 1361 (1994).

Nomikos, P., and J. F. MacGregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, **37**, 41 (1995).

Raich, A., and A. Cinar, "Statistical Process Monitoring and Disturbance Diagnosis in Multivariable Continuous Processes," *AIChE J.*, **42**, 995 (1996).

Wise, B. M., N. L. Ricker, D. F. Veltkamp, and B. R. Kowalski, "A Theoretical Basis for the Use of Principal Component Models for Monitoring Multivariate Processes," *Process Control Qual.*, **1**, 41 (1990).

Wise, B. M., and N. B. Gallagher, "The Process Chemometrics Approach to Process Monitoring and Fault Detection," *J. Process Control*, **6**, 329 (1996).

Wise, B. M., N. B. Gallagher, S. W. Butler, D. D. White, Jr., and G. G. Barna, "A Comparison of Principal Component Analysis, Multiway Principal Component Analysis, Trilinear Decomposition and Parallel Factor Analysis for Fault Detection in a Semiconductor Etch Process," *J. Chemometrics*, **13**, 379 (1999).

Wold, S., P. Geladi, K. Esbesen, and J. Ohman, "Multi-Way Principal Components- and PLS-Analysis," *J. Chemometrics*, **1**, 41 (1987).

Manuscript received Aug. 21, 2000, and revision received Dec. 7, 2001.