Identification of Faulty Sensors Using Principal Component Analysis

Ricardo Dunia
Fisher-Rosemount Systems, Austin, TX 78754

S. Joe Qin and Thomas F. Edgar
Dept. of Chemical Engineering, University of Texas at Austin, Austin, TX 78712

Thomas J. McAvoy
Dept. of Chemical Engineering, University of Maryland, College Park, MD 20742

Even though there has been a recent interest in the use of principal component analysis (PCA) for sensor fault detection and identification, few identification schemes for faulty sensors have considered the possibility of an abnormal operating condition of the plant. This article presents the use of PCA for sensor fault identification via reconstruction. The principal component model captures measurement correlations and reconstructs each variable by using iterative substitution and optimization. The transient behavior of a number of sensor faults in various types of residuals is analyzed. A sensor validity index (SVI) is proposed to determine the status of each sensor. On-line implementation of the SVI is examined for different types of sensor faults. The way the index is filtered represents an important tuning parameter for sensor fault identification. An example using boiler process data demonstrates attractive features of the SVI.

Introduction

Modern chemical processes are well equipped with measurement sensors, such as temperature, flow rate, and pressure sensors. While some of the measurement instruments are used for closed-loop control, others are merely indicators for process monitoring. The availability of many sensors provides valuable redundancy for sensor fault detection and identification because the sensor measurements are highly correlated under normal conditions. These correlations are mainly due to physical and chemical principles governing the process operation, such as mass and energy balances. The task for sensor validation is to detect, identify, and reconstruct faulty sensors by examining the sensor measurements with a known model or prior knowledge.

Univariate and multivariate statistical process control techniques can be used to detect the following abnormal sensor conditions:

- The measurements reach unusual values. An unusual measurement is often caused by a major sensor failure. Univariate statistical methods (Mah and Tamhane, 1982) have been used to detect such faults with lower and upper limits. However, faults that are within the limits but do not follow the normal correlation among sensors are undetectable using this approach. For example, increasing the steam demand of a boiler should trigger an increase in the boiler fuel flow. The fuel flowmeter could fail in such a way that it does not respond to the flow increase, even though all sensor readings are within normal limits. Figure 1 illustrates the case where the sensor fault is within the normal range but does not match the normal correlation.
- Multiple sensors can deviate from normal correlations. Process measurements usually demonstrate strong correlation under normal conditions. These measurement correlations provide necessary redundancy to detect, identify, and reconstruct a faulty sensor. This type of sensor validation examines the difference between the process measurements and the estimates provided by models based on normal data. Statistical techniques such as principal component analysis (Wold et al., 1987), partial least squares (Geladi and Kowalski, 1986), and factor analysis (Basilevsky, 1994) are common approaches to building the correlation models.

Correspondence concerning this article should be addressed to S. J. Qin.

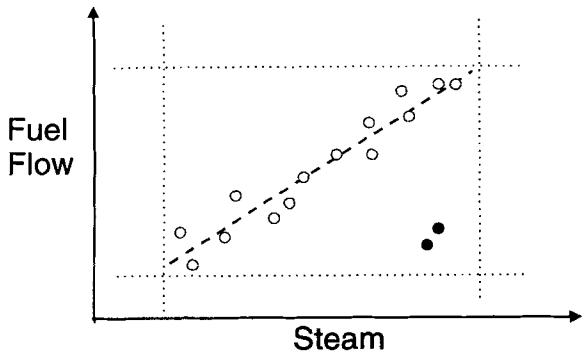

Figure 1. Correlation between the steam demand and the boiler fuel flow.

The good samples (○) are around the correlation line (– – –) and the abnormal samples (●) are within normal limits (· · · ·), but deviate from the correlation.

• The monitored process undergoes transient variations. Unexpected measurement transients, such as oscillations or unusual trends in batch processes, are due to abnormal operating conditions. The detection of this class of anomalies is often used on dynamic statistical models (MacGregor et al., 1994b) or Kalman filters (Da and Lin, 1995). Since most chemical processes are operated under slow transients for the sake of safety, these transients can usually be treated as pseudo-steady state and methods based on steady-state correlations may still be applied. Appropriate filtering techniques will help further eliminate possible transient effects (Piovoso et al., 1992).

Process monitoring and fault detection using principal component analysis (PCA) have been recently investigated in the literature (Tong and Crowe, 1995; MacGregor and Kourti, 1995). Due to correlation, a few principal components are usually sufficient to capture the data variance. Two kinds of abnormal conditions can be distinguished using PCA. These are:

• Failure of sensor correlations: In this situation the PCA model is no longer valid and the Euclidean norm of the residual vector increases significantly.

• Excessive normal variance: The variables used to define the operating variability are out of the normal range, as suggested by the historical data.

A faulty sensor usually breaks down the normal correlation with the remaining sensors. This feature can be used to identify the faulty sensor after an abnormal condition is detected. MacGregor et al. (1994a) used the contribution plot to identify process faults. Such a plot uses the residual of each sensor at every sample to identify the sensors related to a detected fault. The sensor with the largest error is considered faulty since it has a major contribution to the square prediction error used for fault detection. Tong and Crowe (1995) mentioned that the collective statistics derived from PCA allows one to track down the contributions of a fault. However, the existence of the sensor fault may result in incorrect estimation of other sensors and may lead to erroneous conclusions in the identification stage. To eliminate the effect of a faulty sensor in estimating other sensors, Wise and Ricker (1991) use a series of partial least squares (PLS) models to reconstruct faulty sensors among a set of correlated variables. Each model predicts one output using the remaining sensors.

Kramer (1992) used an autoassociative neural network to detect, identify, and reconstruct faulty sensors. This type of network is considered to be a nonlinear extension of PCA (Kramer, 1991). Fantoni and Mazzola (1994) employed an autoassociative neural net in a nuclear power plant to detect sensor faults. The reconstruction procedure applied by Kramer is based on the estimate in the output layer. However, the output estimate contains the effect of the faulty sensor.

This article focuses on identifying faulty sensors and reconstructing sensors based on measurement correlations. Sensor reconstruction is used to validate the original measurement by the model. It is assumed that one sensor has failed and the remaining sensors are used for reconstruction. This procedure is followed sequentially until all sensors have been validated. In the context of linear PCA models, concepts related to the capability of detection, identification, and reconstruction of a sensor fault are discussed and analyzed from a physical point of view. Five residuals based on variable reconstruction are defined and analyzed. Comparison of the different residuals shows that they can be expressed as a linear combination of two of them. We further present the propagation of the different sensor faults in the residuals. A sensor validity index (SVI) based on the most sensitive residual to sensor faults is proposed for sensor fault identification. The validity index can distinguish abnormal operational conditions from a single sensor fault. However, the on-line implementation of the validity index requires the use of a filter to avoid false alarms. An exponentially weighted moving average expression (EWMA) of the validity index and residuals is applied to filter the effect of transients and noise (Lucas and Saccucci, 1990) and to identify different types of sensor faults. Four types of sensor faults are considered in the present article: bias, complete failure, drifting and precision degradation. Figure 2 illustrates each of these sensor faults.

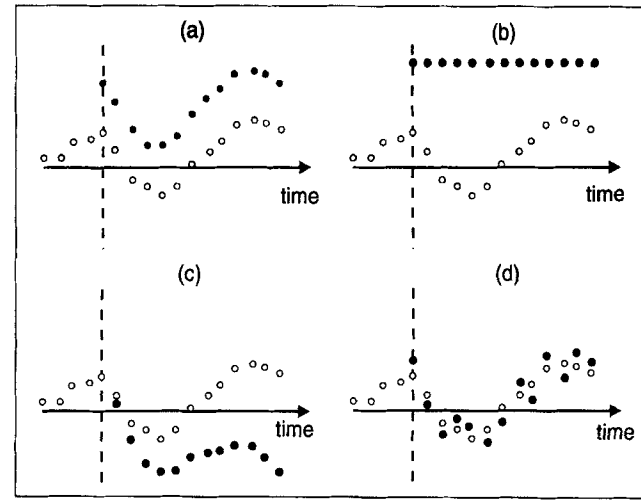

Figure 2. Type of faults found in process data.

The dashed line shows when the fault occurs. ○: data free of fault; ●: corrupted data for the following cases: (a) bias, (b) complete failure, (c) drifting, and (d) precision degradation.

After a brief introduction of PCA and its use in fault detection, this article discusses the reconstruction of each sensor by two approaches based on PCA. The utility of the reconstruction is subject to the degree of correlation among the sensors. A parameter called *unreconstructed variance* is calculated to estimate the goodness of the reconstruction and provides the basis for determining which measurements can be reconstructed. A group of residuals is also defined by considering the measurement reconstruction. The propagation of different sensor faults in these residuals is analyzed to determine a sensitive indicator in the identification of a sensor fault using the validity index. This index is based on the reconstruction procedure and allows one to identify the faulty sensor. The faulty sensor measurements are replaced by the reconstructed values used for identification and do not require additional calculation. An application to a boiler process illustrates the sensor validation procedure and the properties of the SVI. The procedure used for the sensor reconstruction step is presented, as well as advantages of using the SVI for sensor fault identification.

PCA and Fault Detection

Principal component analysis

This work makes use of PCA for sensor validation. Other modeling techniques based on partial least squares and factor analysis could also be applied to obtain empirical models for such a purpose. However, it is convenient to use PCA models because of their orthogonal properties. These properties simplify the fault-identification procedure and allow a better understanding of the sensor-validation problem.

In general, a normalized data matrix X of n samples (rows) and m variables (columns) can be decomposed as follows:

$$X = \hat{X} + E, \tag{1}$$

where the matrices $\hat{X}$ and E represent the modeled and unmodeled variations of X, respectively:

$$\hat{X} = TP^T = \sum_{i=1}^{l} t_i p_i^T \tag{2}$$

$$E = T_e P_e^T = \sum_{i=l+1}^{m} t_i p_i^T \tag{3}$$

and l represents the number of principal components. The matrices T and P are the score and loading matrices, respectively. The decomposition of X is such that the composed matrix $[P\ P_e]$ is orthonormal and $[T\ T_e]$ is orthogonal. The principal component projection reduces the original set of variables to l principal components. Because $[P\ P_e]$ is orthonormal, the cross covariance between $\hat{X}$ and E is zero:

$$\hat{X}E^T = \mathbf{0}. \tag{4}$$

Let the rows of $\hat{X}$ and E be represented by $\hat{\mathbf{x}}^T$ and $\mathbf{e}^T$, respectively. Equation 4 indicates that

$$\hat{\mathbf{x}}_i^T \mathbf{e}_j = 0 \qquad \forall i,\ \ j < n. \tag{5}$$

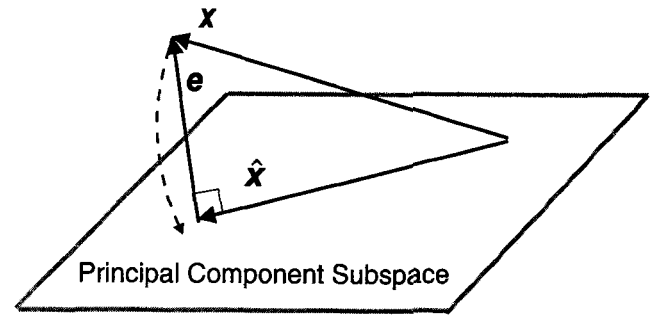

Figure 3. Sample vector x is projected to the principal component subspace.

The projection is $\hat{\mathbf{x}}$, where $\|\hat{\mathbf{x}}\| \le \|\mathbf{x}\|$ (see dashed line). The residual vector $\mathbf{e}$ is perpendicular to the subspace.

The previous expression shows that the row vectors of $\hat{X}$ and E are orthogonal, and these matrices define complementary subspaces of $\Re^m$. Using a generic sample vector, denoted by $\mathbf{x}$, we obtain the decomposition illustrated in Figure 3. This figure also shows that the magnitude of the sample vector is greater than or equal to its estimation. This magnitude reduction is part of the filtering feature of the principal component projection.

The projection to the estimation subspace, also known as the principal component subspace (PCS), is calculated by,

$$XPP^T = TP^T = \hat{X}. \tag{6}$$

The matrix PP^T is denoted by C and is usually rank deficient ($l < m$).

The relation between PCA and the correlation matrix R is given by:

$$R \equiv \frac{1}{n-1} X^T X = \frac{1}{n-1} \hat{X}^T \hat{X} + \frac{1}{n-1} E^T E. \tag{7}$$

The orthogonal property between the $\hat{X}$ and E is used in this equation.

If we use $\hat{R}$ to represent the correlation captured in the model,

$$\hat{R} \equiv \frac{1}{n-1} \hat{X}^T \hat{X} = PR_t P^T,$$

where R_t is the correlation matrix for the scores,

$$R_t \equiv \frac{1}{n-1} T^T T,$$

and R_e denotes the variation left in the residual,

$$R_e \equiv \frac{1}{n-1} E^T E,$$

Equation 7 becomes:

$$R = \hat{R} + R_e. \tag{8}$$

Fault detection using PCA

PCA exploits the use of the process degrees of freedom to detect abnormal conditions. Dunia et al. (1995) have explained why the process degrees of freedom are often restricted due to operational conditions and first principle relations. Tong and Crowe (1995) have summarized the use of PCA in the detection of gross errors. Their work focuses on the extension of standard statistics to PCA. Collective tests are considered in this article because of the multivariable nature of the process data and a multivariable normal distribution of the form $\boldsymbol{x} \simeq N(\boldsymbol{0}, \boldsymbol{R})$ is often assumed. Gnadesikan (1977) has presented a test to check the multivariable normality of the measurements.

Most of the concepts for multivariable data tests are due to Hotelling et al. (1947). The chi-squared distribution is a collective test based on Hotellings' T^2 statistics. It considers the sum of squares of independent normal random distributions of a set of normalized variables. However, PCA is used for correlated variables and the squared prediction error (SPE) is applied instead of the T^2 statistic to test the breakdown of the sensor correlations.

Jackson (1991) developed a test for the residuals obtained from PCA. Such a test suggests the existence of an abnormal condition associated with the breakdown of the correlation among variables when

$$SPE \geq cl(\alpha), \tag{9}$$

where $cl(\alpha)$ is the confidence limit for the $1-\alpha$ percentile in a normal distribution (Jackson, 1991). Wise and Ricker (1991) have used the same test in the detection of a sensor fault. MacGregor and Kourti (1995) summarize extensions of this detection approach using EWMA and CUSUM indicators. Such indicators are more sensitive to small and moderate changes in the mean or standard deviation than using individual measurements. False alarms are also reduced with the application of these indicators, but the detection is usually delayed due to the filtering effect. The delay in the detection of an abnormal condition requires the design of the confidence limits based on the average run length (ARL) (Crowder, 1987). The ARL is the average time required by the indicator to detect an abnormal condition, whether it is a mean or variance shift.

The approach used in this article applies an EWMA expression to the residuals and SPE for detection of abnormal conditions. The general expression for the residuals and SPE-EWMA is

$$\begin{bmatrix} \bar{\boldsymbol{e}} \\ \overline{SPE} \end{bmatrix}_k = \begin{bmatrix} (\boldsymbol{I}-\boldsymbol{\Lambda}) & 0 \\ 0 & (1-\lambda) \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{e}} \\ \overline{SPE} \end{bmatrix}_{k-1} + \begin{bmatrix} \boldsymbol{\Lambda} & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \boldsymbol{e} \\ SPE \end{bmatrix}_k, \tag{10}$$

where $\overline{SPE}$ is the index that detects a correlation breakdown and $SPE = \|\bar{\boldsymbol{e}}\|^2$; $\boldsymbol{\Lambda}$ and λ denote the forgetting factors for the residuals and SPE, respectively. $\boldsymbol{\Lambda}$ is adjusted to favor the identification of particular types of faults, while λ reduces the number of false alarms and allows us to detect moderate and small faults. Therefore, the diagonal matrix $\boldsymbol{\Lambda}$ has diagonal entries usually different from λ. $\boldsymbol{\Lambda} \to \boldsymbol{I}$ tends to favor the detection of variance changes in the data, while $\boldsymbol{\Lambda} \to \boldsymbol{0}$ makes the detection more sensitive to mean changes. Different diagonal entries can be defined in $\boldsymbol{\Lambda}$, depending on the type of fault to detect in each sensor.

The confidence limit for $\overline{SPE}$ is usually estimated from the training data. However, some theoretical expressions for the residuals allow one to derive relations among the failure magnitude and the confidence limits. The following example illustrates an analytical relation of these parameters based on the expectation of $\overline{SPE}$.

Example: Precision Degradation Fault. Let fault-free data be generated from the following expression:

$$\boldsymbol{x}_k^* = s_k \boldsymbol{p} + \boldsymbol{\nu}_k,$$

where $\boldsymbol{x}^*$ is the fault-free sample vector with m normalized entries, s is an independent random variable with zero mean; and $\boldsymbol{\nu}$ is the white noise vector with independent entries. The standard deviations of the noise entries are identical and $\boldsymbol{p} = \boldsymbol{1}(1/\sqrt{m})$, where $\boldsymbol{1}$ is a vector of ones. One principal component is used in modeling the fault free data. It is easy to show that the load vector is identical to $\boldsymbol{p}$ and the score variance is $\sigma_t^2 = \sigma_s^2 + \sigma_\nu^2$. Notice that a portion of the noise variance is captured in the principal component variance.

The corrupted data are given by:

$$\boldsymbol{x}_k = \boldsymbol{x}_k^* + f_k \xi_j,$$

where f_k is random with zero expectation and a variance of δ^2, and ξ_j is a unit norm vector with zeros in all entries except j. These conditions represent a precision degradation fault on the jth sensor.

The residual for the corrupted data is given by

$$\boldsymbol{e}_k = (\boldsymbol{I}-\boldsymbol{C})\boldsymbol{x}_k = (\boldsymbol{I}-\boldsymbol{C})(\boldsymbol{\nu}_k + f_k \xi_j)$$

and because f is independent of $\boldsymbol{\nu}$,

$$\mathcal{E}\{\|\boldsymbol{e}_k\|^2\} = (m-1)\left(\sigma_\nu^2 + \frac{\delta^2}{m}\right). \tag{11}$$

For a precision degradation fault, the current residual and past accumulated residuals are independent. Therefore, the expectation of SPE_k is given by

$$\mathcal{E}\{SPE_k\} = \{\mathcal{E}\|\bar{\boldsymbol{e}}_k\|^2\} = \mathcal{E}\{\|(\boldsymbol{I}-\boldsymbol{\Lambda})\bar{\boldsymbol{e}}_{k-1}\|^2\} + \mathcal{E}\{\|\boldsymbol{\Lambda}\boldsymbol{e}_k\|^2\}. \tag{12}$$

For $\boldsymbol{\Lambda} = \lambda_1 \boldsymbol{I}$,

$$\mathcal{E}\{SPE_k\} = (1-\lambda_1)^2 \mathcal{E}\{SPE_{k-1}\} + \lambda_1^2 \mathcal{E}\{\|\boldsymbol{e}_k\|^2\}. \tag{13}$$

The expectation of the bottom row in Eq. 10 gives

$$\mathcal{E}\{\overline{SPE}_k\} = (1-\lambda)\mathcal{E}\{\overline{SPE}_{k-1}\} + \lambda \mathcal{E}\{SPE_k\}. \tag{14}$$

Substitution of Eq. 13 in the previous expression provides the following second-order response for the expectation of $\overline{SPE}$ in the Z domain:

$$\mathcal{E}\{\overline{SPE}\}(Z) = G(Z)\mathcal{E}\{\|\mathbf{e}\|^2\}, \tag{15}$$

where

$$G(Z) = \frac{\lambda_1^2 \lambda}{\left[1 - Z^{-1}(1-\lambda_1)^2\right][1 - Z^{-1}(1-\lambda)]}. \tag{16}$$

The steady state of $\mathcal{E}\{\overline{SPE}\}$ is given as follows using the final value theorem and Eqs. 11 and 15:

$$\mathcal{E}\{\overline{SPE}\} = \frac{\lambda_1(m-1)(\sigma_v^2 + \delta^2/m)}{2-\lambda_1}.$$

Under the fault-free condition,

$$\mathcal{E}\{\overline{SPE}|\,\delta = 0\} = \frac{\lambda_1(m-1)\sigma_v^2}{2-\lambda_1}.$$

Therefore,

$$\mathcal{E}\{\overline{SPE}\} - \mathcal{E}\{\overline{SPE}|\,\delta = 0\} = \frac{\lambda_1(m-1)\delta^2}{(2-\lambda_1)m}.$$

To maximize the increase in $\mathcal{E}\{\overline{SPE}\}$ due to the precision degradation type of fault, $\lambda_1 = 1$. The confidence limit for $\overline{SPE}$ can be based on a critical value for δ^2,

$$cl \equiv \mathcal{E}\{\overline{SPE}|\,\delta = \delta_{\text{crit}}\} = \mathcal{E}\{\overline{SPE}|\,\delta = 0\} + \frac{(m-1)\delta_{\text{crit}}^2}{m}. \tag{17}$$

Note that λ does not affect the steady-state or confidence limit of $\mathcal{E}\{\overline{SPE}\}$. Therefore, it can be independently set between zero and one to avoid false alarms. Moreover, the parameter λ is adjusted by considering Eq. 15 in terms of the sampling time k,

$$\mathcal{E}\{\overline{SPE}_k\} = \mathcal{E}\{\overline{SPE}|\,\delta = 0\} + \frac{m-1}{m}\delta^2[1-(1-\lambda)^k], \tag{18}$$

where the fault is included at $k = 0$. The average number of samples that $\mathcal{E}\{\overline{SPE}_k\}$ takes to exceed cl, also known as the ARL, is

$$ARL = k, \quad \text{such that } \mathcal{E}\{\overline{SPE}_k\} \geq cl. \tag{19}$$

The substitution of Eqs. 17 and 18 in Eq. 19 gives

$$ARL \geq \frac{\ln\left(1 - (\delta_{\text{crit}}/\delta)^2\right)}{\ln(1-\lambda)}. \tag{20}$$

To illustrate the example with some numerical values, assume that $m = 10$ and the process operation requires the detection of a precision degradation fault of 6 times the noise variance or greater. Such a condition makes $\delta_{\text{crit}}^2 = 6\sigma_v^2$ and

$$cl = (m-1)\sigma_v^2\left(1 + \frac{6}{m}\right) = 14.4\sigma_v^2.$$

If λ is chosen as 0.1, ARL is around 10 for a fault of $9\sigma_v^2$ in variance. Figure 4 illustrates the typical trajectory for $\overline{SPE}$ using the conditions specified earlier. The precision degradation fault was included at $k = 0$ (dotted line) and the confidence limit is represented by the continuous horizontal line. Because $\lambda_1 = 1$, the expected trajectory for $\overline{SPE}$ (dashed line) is a first-order response according to Eq. 16.

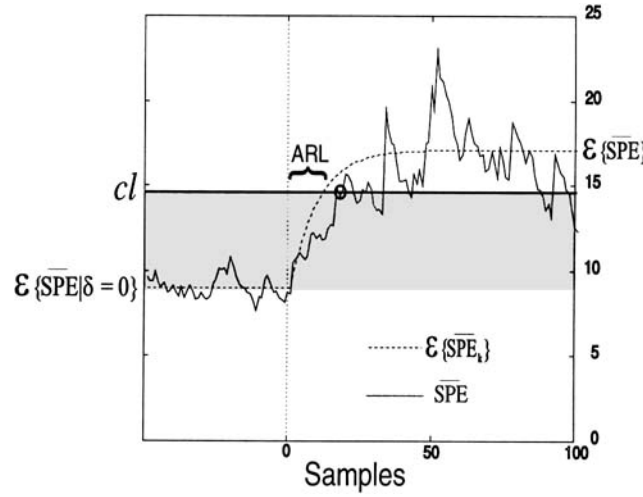

Figure 4. Detection of a process fault using EWMA for the $\overline{SPE}$.

The shaded area represents normal increase in the $\mathcal{E}\{\overline{SPE}\}$ and ○ denotes the detection location. The ordinate has been divided by σ_v^2.

Reconstruction Using PCA

The previous section explained how a fault is detected using an EWMA expression for the SPE. However, the rapid identification and reconstruction of the faulty sensor are critical in order to bring the operation back to normal. The approach to sensor identification proposed in this article assumes each sensor fails and reconstructs the sensor based on remaining measurements. We then validate the assumption by examining the residuals before and after reconstruction, and calculating a sensor validity index. If the faulty sensor is validated, a significant decrease in the residuals is expected. Since the tasks for identification and reconstruction are coupled together, we first present the reconstruction approaches assuming we know the faulty sensor and then discuss fault identification and the sensor validity index.

Given a principal-component model, the number of degrees of freedom is the difference between the number of correlated variables, m, and the number of principal components, l. To detect an abnormal condition it is required that $m - l \geq 1$. Furthermore, the number of degrees of freedom determines the number of variables that can be reconstructed, denoted by ρ. The identification of faulty measure-

ments by reconstruction requires that $(m-\rho)-l \geq 1$. This condition is necessary for identification and can be explained by assuming that the reconstruction of ρ sensors reduces the total number of independent sensors to $m^{\text{new}} = m - \rho$, and $m^{\text{new}} - l \geq 1$ is the new condition for detection. It should be noted that there is no unique criterion to select the number of principal components (PCs). However, one should choose l such that most of the variance in the data is captured. If l is too small, the residual will be large, which makes it unable to separate small faults from normal residuals. If $l = m$, the model reduces to an identity matrix and is useless.

This section presents two approaches to solving for the reconstructed variable, although the final results are equivalent. The goal of reconstruction is to estimate the ith variable from the remaining variables in vector $\mathbf{x}$. This reconstructed value is denoted by z_i and its calculation is demonstrated in this section.

Iterative approach

One may estimate the ith variable from $\mathbf{x}$ using Eq. 2. The drawback of this approach is that the faulty sensor contained in $\mathbf{x}$ is used in the estimate. To eliminate the effect of the faulty sensor, we feed back the prediction of the ith variable $(\hat{x}_i)$ to the input and iterate until it converges to a value z_i. The iteration can be represented by the following expression:

$$z_i^{\text{new}} = c_{ii} z_i^{\text{old}} + [\mathbf{x}_{-i}^T \quad 0 \quad \mathbf{x}_{+i}^T]\boldsymbol{c}_i = [\mathbf{x}_{-i}^T \quad z_i^{\text{old}} \quad \mathbf{x}_{+i}^T]\boldsymbol{c}_i, \quad (21)$$

where

$$\boldsymbol{C} = \boldsymbol{P}\boldsymbol{P}^T = [\boldsymbol{c}_1 \ \boldsymbol{c}_2 \cdots \boldsymbol{c}_m] \qquad \boldsymbol{c}_i^T = [c_{1i} \ c_{2i} \cdots c_{mi}]$$

and $\mathbf{x}^T$ represents a row of the matrix $\boldsymbol{X}$ and the subscripts $-i$, $+i$ denote a vector formed by the first $i-1$ and the last $m-i$ elements of the original vector, respectively. We will show that the iteration always converges and the converged value z_i can be calculated with one formula without actually iterating.

To show the convergence of Eq. 21, we need to show that c_{ii} is inside the unit circle, or on the unit circle only if $\boldsymbol{c}_i = \boldsymbol{\xi}_i$, in which case $z_i^{\text{new}} = z_i^{\text{old}}$ at the first iteration. In this case the ith sensor is not correlated to any other sensors and cannot be reconstructed. Using the orthogonal properties of $\boldsymbol{P}$, we have

$$[\boldsymbol{P} \ \boldsymbol{P}_e][\boldsymbol{P} \ \boldsymbol{P}_e]^T = \boldsymbol{C} + \boldsymbol{P}_e\boldsymbol{P}_e^T = \boldsymbol{I}. \quad (22)$$

The diagonal entry of the previous expression gives

$$c_{ii} + \sum_{h=l+1}^{m} p_{ih}^2 = 1, \quad (23)$$

which suggests that c_{ii} is between zero and one. If $c_{ii} = 1$, the ith row of $\boldsymbol{P}_e$ is $\mathbf{0}$, making the ith row and column of $\boldsymbol{P}_e\boldsymbol{P}_e^T = \boldsymbol{I} - \boldsymbol{C}$ also $\mathbf{0}$, which indicates that $\boldsymbol{c}_i = \boldsymbol{\xi}_i$. Therefore, the convergence of the iterative approach is guaranteed. The asymptotic value for z_i, regardless of the initial condition for $c_{ii} < 1$, is

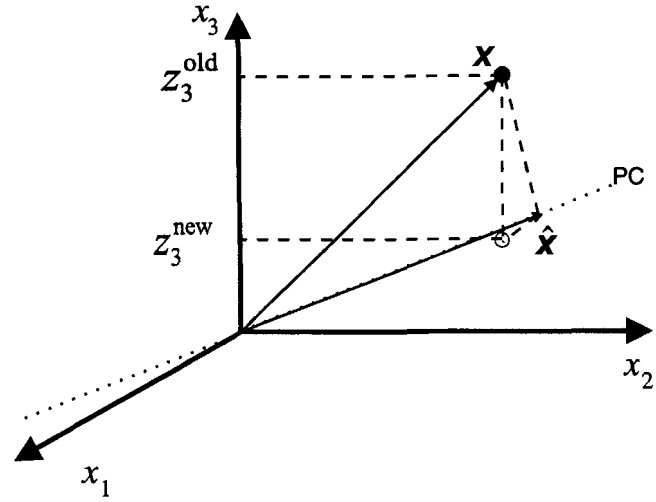

Figure 5. Geometric interpretation of the iterative approach.

The vector $\mathbf{x}$ is projected on the PC line and a new x_3 coordinate, denoted by z_3^{new}, is calculated.

$$z_i = \frac{[\mathbf{x}_{-i}^T \ 0 \ \mathbf{x}_{+i}^T]\boldsymbol{c}_i}{1 - c_{ii}} \quad (24)$$

and $z_i = x_i$ for $c_{ii} = 1$. It is interesting to note that if we treat the faulty sensor as having a missing value, the missing data replacement approach available in the literature (Martens and Naes, 1989) would suggest the following formula for replacement:

$$z_i = \boldsymbol{\xi}_i^T \boldsymbol{P}(\boldsymbol{P}_i^T\boldsymbol{P}_i)^{-1}\boldsymbol{P}_i^T\mathbf{x}, \quad (25)$$

where $\boldsymbol{P}_i^T = [\boldsymbol{P}_{-i}^T \ \mathbf{0} \ \boldsymbol{P}_{+i}^T]$. In the Appendix we show that the iterative approach is equivalent to the one proposed by Martens and Naes for missing data. However, Eq. 24 does not require matrix inversion and the reconstruction is obtained without actually iterating.

Figure 5 shows the geometric interpretation of the iterative approach for three correlated variables and one principal component. The point • represents $\mathbf{x}$, and the reconstructed variable is x_3. The iterative approach projects $\mathbf{x}$ on the PC line and the x_3 coordinate (z_3^{new}) is used to define the new location of $\mathbf{x}$, denoted in the figure by ○. The procedure continues until x_3 converges to its final value z_3. Notice that the remaining coordinates are kept constant. For a linear PCA model, we find the final solution of the iteration as in Eq. 24.

Optimization approach

This approach was implemented by Wise and Ricker (1991) for a set of reconstructed variables. For the reconstruction of the ith variable, the optimization procedure reduces to

$$\min_{z_i} \|\mathbf{x}_i^*\|_{I-C}^2$$

where $\mathbf{x}_i^*$ is identical to $\mathbf{x}$ except for the ith entry, z_i. Since $\boldsymbol{I} - \boldsymbol{C}$ is positive semidefinite, the derivative of the objective

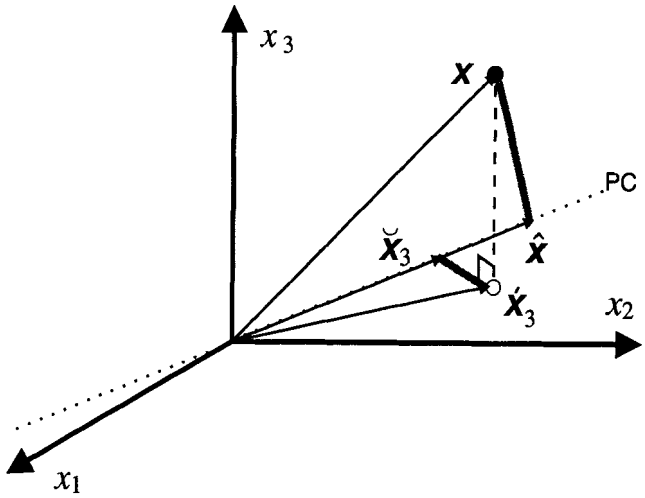

Figure 6. Geometric perspective of the optimization approach.

The point • with coordinates $\boldsymbol{x}$ slides parallel to the x_3-axis to the optimal location ○ of coordinates $\acute{\boldsymbol{x}}_3$. The initial and final magnitude of the objective function are represented by the gray bars.

function with respect to z_i leads to a necessary condition for a minimum:

$$\xi_i^T(\boldsymbol{I}-\boldsymbol{C})\acute{\boldsymbol{x}}_i=[\boldsymbol{x}_{-i}^T\ z_i\ \boldsymbol{x}_{+i}^T](\xi_i-\boldsymbol{c}_i)=0. \tag{26}$$

This expression leads to the same solution as Eq. 24.

To visualize this approach we will assume similar conditions to those for Figure 5 ($l=1$, $m=3$, and x_3 is reconstructed). The minimization approach slides the point • along the x_3-axis to a location ○, as is illustrated in Figure 6. The condition imposed by Eq. 26 is equivalent to choosing the location ○ such that the new residual $(\acute{\boldsymbol{x}}_3-\breve{\boldsymbol{x}}_3)$ is perpendicular to the sliding direction, ξ_3,

$$\xi_3^T(\acute{\boldsymbol{x}}_3-\breve{\boldsymbol{x}}_3)=0, \tag{27}$$

where $\breve{\boldsymbol{x}}_3$ defines the new location of $\hat{\boldsymbol{x}}$ when x_3 is reconstructed. From the geometric interpretation we see that the iterative procedure converges when the condition imposed by the optimization approach is satisfied.

Unreconstructed variance

The reconstruction of correlated variables was defined earlier for two different procedures leading to the same result. However, it is important to quantify the amount of variation that can be reconstructed given a principal component model and the correlation matrix.

An expression that summarizes the reconstruction procedure is obtained from Eq. 24:

$$z_i=\frac{1}{1-c_{ii}}[\boldsymbol{c}_{-i}^T\ 0\ \boldsymbol{c}_{+i}^T]\boldsymbol{x}. \tag{28}$$

The unreconstructed variance of variable i, denoted by u_i, is given by the variance of x_i-z_i:

$$u_i\equiv\mathrm{var}(x_i-z_i)=\mathrm{var}\left(\xi_i^T\boldsymbol{x}-\frac{1}{1-c_{ii}}[\boldsymbol{c}_{-i}^T\ 0\ \boldsymbol{c}_{+i}^T]\boldsymbol{x}\right)$$

$$\approx\frac{1}{(1-c_{ii})^2}\|\xi_i-\boldsymbol{c}_i\|_{\boldsymbol{R}}^2$$

$$=\frac{1}{(1-c_{ii})^2}\xi_i^T(\boldsymbol{I}-\boldsymbol{C})\boldsymbol{R}(\boldsymbol{I}-\boldsymbol{C})\xi_i=\frac{\xi_i^T\boldsymbol{R}_e\xi_i}{(\xi_i^T\boldsymbol{P}_e\boldsymbol{P}_e^T\xi_i)^2}. \tag{29}$$

Since each variable x_i is normalized to unit variance, $u_i=1$ if we let $z_i=\bar{x}_i$. Therefore, $u_i>1$ is the case for poor reconstruction and the use of the mean of x_i instead of Eq. 28 actually gives a smaller u_i. This situation can happen when the reconstructed sensor exhibits a lack of correlation to others.

The next example provides an expression for the case of m variables with identical correlation.

Example. Consider the following correlation matrix:

$$\xi_i^T\boldsymbol{R}\xi_j=\begin{cases}r, & i\neq j\\ 1, & i=j\end{cases}\quad\text{where } i,j\leq m. \tag{30}$$

The largest eigenvalue for such a correlation matrix is $\sigma_{t1}^2=1+(m-1)r$ and $\boldsymbol{p}_1=(1/\sqrt{m})\boldsymbol{1}$. The use of one principal component in Eq. 29 leads to,

$$u_i=(1-r)\frac{m}{m-1}. \tag{31}$$

In Eq. 31 the unreconstructed variance decreases with m and r. To achieve $u_i<1$, we need $r>1/m$. If $r<1/m$, using the mean of x_i will be a better replacement than Eq. 24.

Reconstruction Residuals and Fault Propagation

Residuals based on variable reconstruction

The reconstructed vector obtained in the previous section can be used to define a set of residuals for process monitoring. Such residuals can detect a deviation between the individual measurements and the normal trend of the remaining variables. The magnitude of this deviation will be used to identify a faulty sensor.

To obtain a diverse group of residuals based on the reconstructed measurements, we define the matrix $\boldsymbol{G}_i$ from Eq. 28,

$$\boldsymbol{G}_i^T=[\xi_1\ \xi_2\cdots\boldsymbol{g}_i\cdots\xi_m],$$

where

$$\boldsymbol{g}_i^T=\frac{1}{1-c_{ii}}[\boldsymbol{c}_{-i}^T\ 0\ \boldsymbol{c}_{+i}^T]. \tag{32}$$

$\boldsymbol{G}_i$ allows one to calculate the adjusted measurement, $\acute{\boldsymbol{x}}_i$, using the input vector, $\boldsymbol{x}$,

$$\acute{\boldsymbol{x}}_i^T=\boldsymbol{x}^T\boldsymbol{G}_i^T=[x_1\cdots z_i\cdots x_m]. \tag{33}$$

Table 1. Description of the Different Weights Used to Define the Residuals

Residual	H	Description
A	$I-C$	Measurements—model estimation
B	$I-CG_i$	Measurements—model estimation of adjusted measurements
C	G_i-C	Adjusted measurements—model estimations
D	$I-G_i$	Measurement—adjusted measurement
E	$(I-C)G_i$	Adjusted measurement—model estimation of adjusted measurement

A set of residuals $\{A, B, \ldots, E\}$ similar to the SPE and based on G_i are defined,

$$A, B, \cdots, E = \|\boldsymbol{H}\boldsymbol{x}\|^2. \tag{34}$$

Table 1 presents H as a function of G_i for five different cases. Residual A is the typical SPE. These residuals can be expressed as a linear combination of any two residuals. Table 2 shows residuals B, C, and E as a linear combination of residuals A and D. In the next section we investigate the properties of the different residuals in the presence of a fault to determine a sensitive indicator for faulty sensors.

Fault propagation in the residuals

In the Introduction we defined different types of sensor faults. These faults can be expressed in terms of the scalar f on variable j,

$$\boldsymbol{x}_k = \boldsymbol{x}_k^* + f_k \xi_j, \tag{35}$$

where f denotes the faulty term and the superscript $*$ denotes the uncorrupted portion. The description of f depends on the nature of the fault, as illustrated in Table 3 and Figure 2.

The propagation of sensor faults inside the different residuals is useful for selecting an appropriate indicator for fault identification. To analyze how different faults propagate inside each type of residual, we substitute Eq. 35 in Eq. 34:

$$\|\boldsymbol{H}(\boldsymbol{x}_k^* + f_k \xi_j)\|^2 = \|\boldsymbol{x}_k^*\|_W^2 + 2\boldsymbol{w}_j^T \boldsymbol{x}_k^* f_k + w_{jj} f_k^2, \tag{36}$$

where

Table 2. Relation Among Different Residuals

B	=	$A + \dfrac{(\hat{x}_i - x_i)^2 c_{ii}}{(1-c_{ii})^2}$	=	$A + Dc_{ii}$
C	=	$A + \dfrac{(\hat{x}_i - x_i)^2 (2c_{ii}-1)}{(1-c_{ii})^2}$	=	$A + D(2c_{ii}-1)$
D	=	$\dfrac{(\hat{x}_i - x_i)^2}{(1-c_{ii})^2}$		
E	=	$A - \dfrac{(\hat{x}_i - x_i)^2}{1-c_{ii}}$	=	$A - D(1-c_{ii})$
$C+A$	=	$B+E$		

Table 3. Expressions for Different Types of Faults

Sensor Fault	Description	Representation (δ is a Constant)
Bias	Measurement affected by a constant term	$f_k = f_{k-1} \quad f_0 = \delta$
Complete failure	Measurement is a constant	$f_k + \xi_j^T \boldsymbol{x}_k^* = \delta$
Drifting	Autocorrelated term with constant forcing function added to the measurement	$f_k = f_{k-1} - \delta$ $f_0 = 0$
Precision degradation	Generated from a probabilistic distribution	$\mathcal{E}(f_k) = 0$ $\mathcal{E}\{f_k^2\} = \sigma_f^2 = \delta^2$

$$\boldsymbol{W} = \boldsymbol{H}^T\boldsymbol{H} = [\boldsymbol{w}_1 \ \cdots \ \boldsymbol{w}_m], \qquad \boldsymbol{w}_i^T = [w_{1i} \ \cdots \ w_{mi}]. \tag{37}$$

The righthand side of Eq. 36 consists of three terms. From left to right they represent the residual term due to nonpredictable variation of the data, the residual due to correlation between the fault and the measurement, and the residual attributed to the fault. For dynamic systems the vectors $\boldsymbol{x}_{k-i}$, where $i = 1, \ldots, n_d$ can be appended to $\boldsymbol{x}_k$. However, the fault identification procedure will be subject to multiple reconstructions and it may be desirable to use dynamic principal-component models (McBrayer, 1995).

We estimate the expectation of a faulty sensor with uncorrupted data decomposed into:

$$\boldsymbol{x}_k^* = \boldsymbol{P}\boldsymbol{t}_k + \boldsymbol{\nu}_k. \tag{38}$$

The matrix $\boldsymbol{P}$ contains l orthonormal columns, where

$$\mathcal{E}(\boldsymbol{t}_k) = \boldsymbol{0}, \qquad \xi_i^T \mathcal{E}\{\boldsymbol{t}_k \boldsymbol{t}_k^T\} \xi_j = \begin{cases} 0, & i \neq j \\ \sigma_{ti}^2, & i = j \end{cases}$$

$$\mathcal{E}\{\boldsymbol{v}_k\} = \boldsymbol{0}, \qquad \xi_i^T \mathcal{E}\{\boldsymbol{v}_k \boldsymbol{v}_k^T\} \xi_j = \begin{cases} 0, & i \neq j \\ \sigma_v^2, & i = j \end{cases}$$

and

$$\xi_i^T \mathcal{E}\{\boldsymbol{t}_k \boldsymbol{\nu}_k^T\} \xi_j = 0, \qquad \forall i, j. \tag{39}$$

The entries of $\boldsymbol{t}$ are placed in descending order defined by their variances:

$$\sigma_{ti}^2 \geq \sigma_{tj}^2 \quad \text{for} \quad i < j. \tag{40}$$

The expectation of residual A for faulty sensors, represented by $\mathcal{E}\{A|f\}$, is illustrated in Table 4. This table shows

Table 4. Effect of Faults on the Expectation of Residual A for l Principal Components

Type of Faults	$\mathcal{E}\{A\|f\}$
Bias	$\sigma_v^2(m-l) + \delta^2(1-c_{jj})$
Complete failure	$\sigma_v^2(m-l) + (\sigma_j^2 - 2\sigma_v^2 + \delta^2)(1-c_{jj})$
Drifting	$\sigma_v^2(m-l) + (\delta k)^2(1-c_{jj})$
Precision degradation	$\sigma_v^2(m-l) + \delta^2(1-c_{jj})$

Table 5. Effect of Sensor Faults on the Expectation of Residual D for l Principal Components

Type of Faults	$\mathcal{E}\{D_i\|f\}$ $i \neq j$	$\mathcal{E}\{D_i\|f\}$ $i = j$
Bias	$\sigma_v^2 \frac{1}{1-c_{ii}} + \delta^2 \left(\frac{c_{ij}}{1-c_{ii}}\right)^2$	$\sigma_v^2 \frac{1}{1-c_{jj}} + \delta^2$
Complete failure	$\sigma_v^2 \frac{1}{1-c_{jj}} + (\sigma_f^2 - 2\sigma_v^2 + \delta^2)\left(\frac{c_{ij}}{1-c_{ii}}\right)^2$	$\sigma_v^2 \frac{c_{jj}}{1-c_{jj}} + \sigma_f^2 + \delta^2$
Drifting	$\sigma_v^2 \frac{1}{1-c_{ii}} + (\delta k)^2 \left(\frac{c_{ij}}{1-c_{ii}}\right)^2$	$\sigma_v^2 \frac{1}{1-c_{jj}} + \delta^2$
Precision degradation	$\sigma_v^2 \frac{1}{1-c_{ii}} + \delta^2 \left(\frac{c_{ij}}{1-c_{ii}}\right)^2$	$\sigma_v^2 \frac{1}{1-c_{jj}} + \delta^2$

that all expressions have σ_v^2 and δ^2 as factors when the right number of principal components are chosen, and the ratio δ/σ_v is important for the detection of a faulty sensor. Therefore large noise variance will make it difficult to detect a fault.

The expectation of residual D for faulty sensors depends on the reconstructed variable i. The subscript j denotes where the fault is located. Table 5 lists the expectation of the residual D for all possible f. In the absence of a fault $\mathcal{E}\{D_i|f=0\} = \sigma_v^2/(1-c_{ii})$, which suggests that the residual variance is amplified in the reconstructed measurement. For $f \neq 0$ the scalar c_{ij} plays an important role in the calculation of $\mathcal{E}\{D_i|f\}$ when $i \neq j$. This condition intuitively indicates that c_{ij} measures the effect of a faulty variable j on the reconstruction of variable i.

The similarity among bias, drifting, and precision degradation faults is observed in Tables 4 and 5. These three types of faults always exhibit a consistent structure in the expectation of residuals A and D_i. In the case of a complete failure, the residuals are correlated with $\mathbf{x}^*$ and the variance of the measurement (σ_f^2) affects the expectation of A and D_i. A qualitative explanation of these relationships is based on the correlation between the measurements and the residuals. For bias, drift, or precision degradation, the faulty sensor still reflects the variations of the fault-free portion of the measurement. However, the residual variations are not correlated with the fault-free data. The opposite is true for the complete failure type of fault.

Based on $\mathcal{E}\{A|f\}$, $\mathcal{E}\{D_i|f\}$, and the relations shown in Table 2 we calculate the expectation of residuals B_i, C_i, and E_i. Among these residuals the most suitable one to identify the faulty sensor is the residual E_i because this residual completely eliminates δ^2 when the true faulty sensor is chosen ($i=j$). Table 6 shows that all the expressions obtained for $\mathcal{E}\{E_i|f\}$ when $i=j$ are free of faulty terms and equal to $\sigma_v^2(m-l-1)$ regardless of the type of fault. This last expression is also identical to the case where $i \neq j$ and $f=0$. The condition imposed by the number of degrees of freedom is implicit in the calculation of such a residual. Furthermore, the factor $m-l-1$ illustrates the reduction of one degree of freedom of E_i compared to A. This is because the ith variable has been reconstructed from the remaining variables.

Identification by a Sensor Validity Index

A sensor fault can be identified by using indicators based on other correlated measurements. The previous section shows that a sensor fault propagates in the different residuals except in residual E when the faulty sensor is reconstructed.

This section concentrates on the identification of a single faulty sensor, but the algorithm can be serially implemented to identify multiple faults that do not occur simultaneously. Nevertheless, it is unusual to have simultaneous sensor faults unless, for example, a power failure occurs, which is usually prevented by redundant power supplies. Therefore, we con-

Table 6. Effect of Sensor Faults on the Expectation of Residual E for l Principal Components

Type of Faults	$\mathcal{E}\{E_i\|f\}$ $i \neq j$	$\mathcal{E}\{E_i\|f\}$ $i = j$
Bias	$\sigma_v^2(m-l-1) + \delta^2\left(1-c_{jj}-\frac{c_{ij}^2}{1-c_{ii}}\right)$	$\sigma_v^2(m-l-1)$
Complete failure	$\sigma_v^2(m-l-1) + (\sigma_f^2 - 2\sigma_v^2 + \delta^2)\left(1-c_{jj}-\frac{c_{ij}^2}{1-c_{ii}}\right)$	$\sigma_v^2(m-l-1)$
Drifting	$\sigma_v^2(m-l-1) + (\delta k)^2\left(1-c_{jj}-\frac{c_{ij}^2}{1-c_{ii}}\right)$	$\sigma_v^2(m-l-1)$
Precision degradation	$\sigma_v^2(m-l-1) + \delta^2\left(1-c_{jj}-\frac{c_{ij}^2}{1-c_{ii}}\right)$	$\sigma_v^2(m-l-1)$

sider the case of no simultaneous sensor faults, which helps isolate a sensor problem from an abnormal operating condition where a group of variables breaks the normal correlation. Another reason to choose the serial identification scheme is to avoid dealing with a large number of combinations of faulty sensors.

Design and properties

The previous section illustrates how different faults propagate inside the residuals for steady-state random processes. It was shown that $\mathcal{E}\{E_i\}$ clearly indicates the effect of the sensor fault when the correct measurement is chosen. Here a quantitative index or statistic should be defined as an objective measure of sensor performance. This index should have a precise span range regardless of the number of principal components, noise, variable variances, or type of faults. It should also distinguish the abnormal operating conditions from the sensor fault situation. The ratio between residuals E and A can provide these desired properties for the identification of a sensor fault,

$$\eta_i^2 = \frac{E_i}{A} = 1 - (1 - c_{ii})\frac{D_i}{A}. \tag{41}$$

We call η_i, the *sensor validity index* (SVI), and it ranges between [0,1] because D, E, and A are positive. To analyze in detail the previous equation, we can express A as

$$A = \sum_{h=1}^{m} e_h^2 = \sum_{h=1}^{m} D_h(1 - c_{hh})^2. \tag{42}$$

Substitution in Eq. 41 gives

$$\eta_i^2 = 1 - (1 - c_{ii})\frac{D_i}{\sum_{h=1}^{m} D_h(1 - c_{hh})^2}. \tag{43}$$

A validity index close to one indicates that the sensor variations follow the variations experienced by the remaining sensors. When the sensor is faulty, η_i is close to zero. However, if several variables break the correlation captured in the principal-component model the ratio D_i/A goes to zero. From Eq. 41 the index will go to one. This result indicates that the index can distinguish between an abnormal operational condition and a faulty sensor.

A contribution plot approach based on residual A has been recently used in the literature (MacGregor et al., 1994a; Tong and Crowe, 1995). This approach compares the contribution of each variable in the SPE when a fault has been detected. Because this approach is based on residual A, the sensor fault propagates to other variable estimates, which increases the residual contribution of other sensors. As a consequence it may increase the chance of erroneous identification.

The identification algorithm presented in this article makes use of reconstructed values to calculate the residual E_i. Because the ith sensor, which is assumed faulty, is not used in reconstruction, the effect of the fault does not appear in its estimate. This procedure is followed sequentially for all variables. A considerable reduction in the residual E_i suggests that the ith variable is faulty. The advantages of the *sensor validity index* based on residual E are (1) it has normalized values between zero and one, making it a well-defined index that simplifies the selection of the alarm threshold; (2) it is sensitive to sensor faults due to the following reason: if the ith sensor is faulty, the residual A increases but E_i is not affected because the ith measurement is not used in such an indicator, making $\eta = E_i/A$ to approach zero.

Two considerations should be taken into account for the use of η to identify faulty sensors. The first is the estimation of the threshold based on historical data. The lack of a theoretical calculation for the confidence region when a distribution is assumed for the measurement is because of the dependence between D and A. Such a dependence makes the expectation of the ratio of the residuals different from the ratio of their expectations. The second point to consider is the oscillations experienced by the validity index when no sensor fault is present in the system. These oscillations are due to the tendency of the ratio to always penalize the variables with the largest error. The use of a filter for the validity index is necessary to eliminate such oscillations and the possibility of erroneous identification.

EWMA for fault identification

In the previous section we defined an index that allows one to identify sensor faults in a reliable way. Here we analyze the effect of filtering the residuals as well as the validity index in the identification of the different types of sensor faults.

Effect of the Residual-EWMA on Fault Identification. Filtering the residuals affects not only the detection but also the identification of the faulty sensor. Λ is adjusted for the identification of mean or variance change in the measurements. To quantify the filtering effect on the residuals we define a filter based on a moving window $\boldsymbol{E}_k$, which consists of the past n_w residuals starting at the current sample k. This window is updated at every sample by shifting the rows of $\boldsymbol{E}_k$ ahead to replace the residual in the last row. The average expression is given by:

$$\bar{\mathbf{e}}_k = \frac{1}{n_w}\boldsymbol{E}_k^T\mathbf{1} = (\boldsymbol{I} - \boldsymbol{C})\frac{1}{n_w}\boldsymbol{X}_k^T\mathbf{1} = (\boldsymbol{I} - \boldsymbol{C})\bar{\mathbf{x}}_k, \tag{44}$$

where the bar indicates an average of the past n_w samples, including the current sample k. A vector of weights can be used instead of $\mathbf{1}$ to give more importance to recent samples inside the window.

Substitution of Eq. 44 in the residuals means, $\bar{A}$ and $\bar{D}_i$, gives

$$\bar{A} = \|\bar{\mathbf{e}}\|^2 \tag{45}$$

$$\bar{D}_i = \left(\frac{\xi_i^T\bar{\mathbf{e}}}{1 - c_{ii}}\right)^2. \tag{46}$$

Notice that the previous expressions follow the relations defined in Table 2. The expectation of the residuals based on filtered errors retains the same structure as the ones for individual residuals.

Table 7. $\bar{\delta}$ as a Function of δ, n_w, and k for Different Types of Faults

Type of Fault	$k \geq n_w - 1$	$k < n_w - 1$
Bias	δ	$\frac{k\delta}{n_w - 1}$
Complete failure	δ	$\frac{k\delta}{n_w - 1}$
Drifting	$\left(k - \frac{n_w - 1}{2}\right)\delta$	$\frac{k(k+1)\delta}{n_w 2}$
Precision degradation	$\frac{\delta}{\sqrt{n_w}}$	$\frac{\sqrt{k+1}}{n_w}\delta$

For a set of faulty measurements defined by Eqs. 35 and 38, we get

$$\bar{\mathbf{e}}_k = (\mathbf{I} - \mathbf{C})\left(\bar{\boldsymbol{\nu}}_k + \bar{f}_k \boldsymbol{\xi}_j\right). \tag{47}$$

The vector $\bar{\boldsymbol{\nu}}_k$ still represents independent events with zero expectation but reduced variance. The effect of $\bar{f}_k$ for the different types of failures is now expressed as a function of $\bar{\delta}_k$. Table 7 shows how $\bar{\delta}_k$ is related to δ. Note that the window could momentarily include a portion of uncorrupted measurements when $k \leq n_w - 1$. This temporal effect is illustrated in Figure 7 where the window at sample $k_1 < n_w - 1$ includes measurements free of faults, while the window at sample $k_2 > n_w - 1$ only considers corrupted measurements. The adjustment of the standard deviation of the noise and the substitution of $\bar{\delta}$ in the place of δ validate the use of Tables 4 to 6 for filtered residuals.

The filter on the residuals reduces the variance of the noise, helping the detection of small δ for faults with autocorrelated residuals. However, the expressions for the case of precision degradation in Table 7 show that it is difficult to identify this fault using a residual filter with large n_w. Another effect of using large windows is the increase in the number of samples for which $k + 1 < n_w$. This delays the influence of f_k on the residuals for any fault because n_w divides all terms in the second column of Table 7.

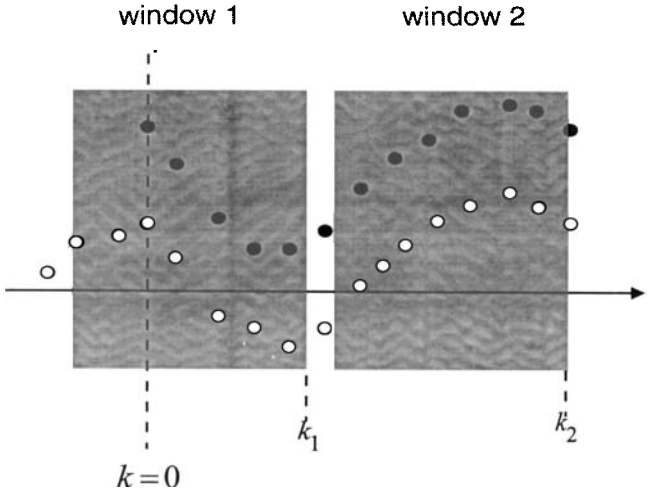

Figure 7. Window defined for the filter can momentarily include good and corrupted measurements like in the case of window 1.

The storage of the past n_w residuals for the calculation of a mean indicator could result in an excessive waste of computer memory. A residual-EWMA is obtained by taking the difference of two consecutive means defined in Eq. 44:

$$\bar{\mathbf{e}}_k^T - \bar{\mathbf{e}}_{k-1}^T = \frac{1}{n_w}\mathbf{1}^T\mathbf{E}_k - \frac{1}{n_w}\mathbf{1}^T\mathbf{E}_{k-1} = \frac{1}{n_w}\left(\mathbf{e}_k^T - \mathbf{e}_{k-n_w+1}^T\right). \tag{48}$$

An approximation of $\mathbf{e}_{k-n_w+1}$ relative to its corresponding mean $\bar{\mathbf{e}}_{k-1}$ provides a residual-EWMA identical to Eq. 10 when $\boldsymbol{\Lambda} = (1/n_w)\boldsymbol{I}$. However, different window sizes (forgetting factor entries) can be applied for each sensor residual.

EWMA for the Validity Index. The implementation of a filter for η^2 reduces the oscillations of the validity index. The use of a η^2-EWMA results in the following state-space representation:

$$\begin{bmatrix} \bar{\mathbf{e}} \\ \overline{SPE} \\ \bar{\eta}^2 \end{bmatrix}_k = \begin{bmatrix} (\mathbf{I} - \boldsymbol{\Lambda}) & 0 & 0 \\ 0 & (1-\lambda) & 0 \\ 0 & 0 & (\mathbf{I} - \boldsymbol{\Gamma}) \end{bmatrix} \begin{bmatrix} \bar{\mathbf{e}} \\ \overline{SPE} \\ \bar{\eta}^2 \end{bmatrix}_{k-1} + \begin{bmatrix} \boldsymbol{\Lambda} & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \boldsymbol{\Gamma} \end{bmatrix} \begin{bmatrix} \mathbf{e} \\ SPE \\ \eta^2 \end{bmatrix}_k \tag{49}$$

where $\boldsymbol{\Gamma}$ is the forgetting factor matrix for the validity index and,

$$\eta_i^2 = 1 - \frac{\bar{e}_i^2}{(1 - c_{ii}) \sum_{h=1}^{m} \bar{e}_h^2}. \tag{50}$$

As in the detection stage, the factors for the residual and index are not necessarily identical. However, $\boldsymbol{\Gamma}$ and the threshold for $\bar{\eta}$ are adjusted such that identification is delayed with respect to detection by a reasonable number of samples.

Notice that the vector in Eq. 49 has been augmented with $\bar{\eta}^2$. The last two rows in this equation represent monitored outputs. No input has been added to the system, and Eq. 50 makes the dynamics of $\bar{\eta}^2$ nonlinear. Moreover, the ratio defined in such an equation makes impossible the calculation of a theoretical ARL for identification of the faulty sensor. However, $\bar{\eta}$ is bounded and insensitive to other abnormal conditions than sensor faults. The initial conditions are usually assumed zero for the residuals and SPE, and $1 - [1/(m-1)]$ for $\bar{\eta}^2$. This last condition is taken from the ratio of expectations of the reconstructed residuals and could temporarily lead to an index greater than one. Nevertheless, the effect of such initial conditions will eventually disappear due to the use of forgetting factors.

Application to a Boiler Process

Nine sensors with 400 samples were collected from a boiler process. Eighty percent of the data are used to analyze the

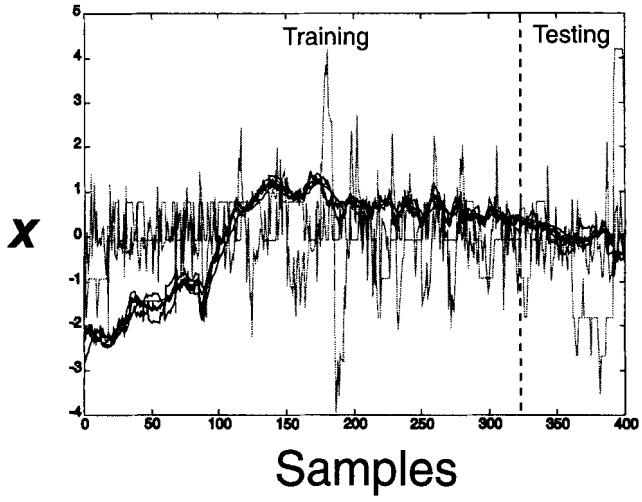

Figure 8. Correlated data in the boiler process.

Notice that most of the variables follow a similar pattern.

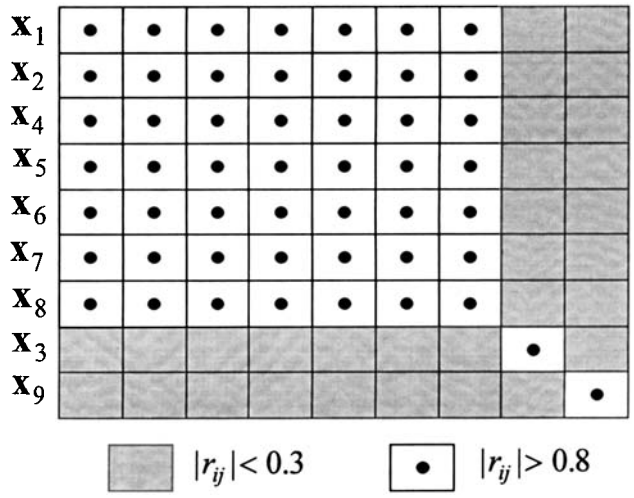

Figure 9. Correlation block for the boiler process data.

All variables except x_3 and x_9 can be reconstructed.

PCA model, and the remaining samples are considered for testing. Figure 8 shows that most of the normalized variables follow the same trend. This qualitatively confirms the correlation among the variables.

Not all the variables in a plant are correlated to others. A group of correlated variables in the process defines a sensor correlation block. The initial step for sensor validation using PCA is to reorder the entries in the sample vector $\mathbf{x}$ such that correlation blocks are defined. A composition of column–row permutations is applied to the original correlation matrix to group sensors by their correlation coefficients. The matrix $\boldsymbol{R}^{\text{new}}$ represents the new correlation matrix with permuted column–rows such that larger r_{ij} elements are closest to the diagonal, which is achieved by minimizing

$$\min_{V} \sum_{i=1}^{m} \sum_{j=1}^{m} \mathrm{f}(r_{ij}^{\text{new}})|i-j|, \tag{51}$$

where

$$\boldsymbol{R}^{\text{new}} = \boldsymbol{V}^{T}\boldsymbol{R}\boldsymbol{V}. \tag{52}$$

The matrix $\boldsymbol{V}$ represents a composition of column permutations,

$$\boldsymbol{V} = \prod \boldsymbol{S}_{ij},$$

and $\boldsymbol{S}_{ij}$ provides the permutation between the i and j columns when it post multiplies $\boldsymbol{R}$,

$$\boldsymbol{S}_{ij} = [\,\xi_1 \;\cdots\; \xi_{i-1}\;\; \xi_j \;\; \xi_{i+1} \;\cdots\; \xi_{j-1}\;\; \xi_i \;\; \xi_{j+1} \;\cdots\; \xi_m\,].$$

To keep $\boldsymbol{R}^{\text{new}}$ symmetric the permutations are applied to both sides of $\boldsymbol{R}$ in Eq. 52. Because there is a significant gap between [0.3 0.8] in the correlation coefficients, the nonlinear function f is defined by

$$\mathrm{f}(r_{ij}) = \begin{cases} 1 & r_{ij} > 0.8 \\ 0 & r_{ij} < 0.3. \end{cases}$$

The resulting $\boldsymbol{R}^{\text{new}}$ is illustrated in Figure 9. Seven variables define a correlation block in which all $r_{ij} > 0.8$. Variables x_3 and x_9 only correlate with themselves and are not considered in the reconstruction stage. The validation of such uncorrelated variables is based on univariate tests. The unreconstructed variance is not used to define the sensor validation block in this example because the correlation coefficient gap shows which sensors are correlated and which are not.

The PCA model is built using the training set for the seven correlated variables. Figure 10 shows the percentage of cumulative variance of the training set as a function of the number of principal components, where

$$Cv(l) \approx 100 \frac{\sum_{j=1}^{l} \sigma_{tj}^2}{\sum_{j=1}^{m} \sigma_{tj}^2}\%. \tag{53}$$

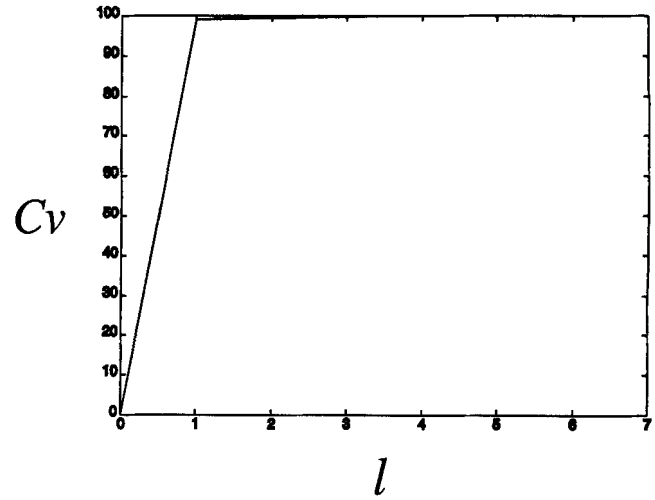

Figure 10. Cumulative variance for the boiler process data.

More than 99% of the data variation can be explained for $l = 1$.

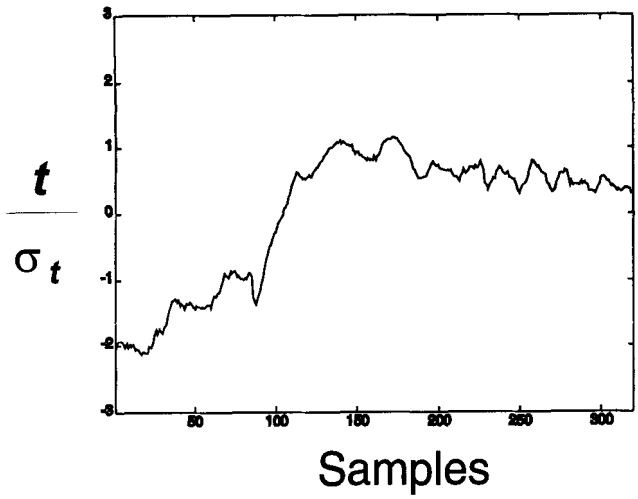

Figure 11. Normalized scores for the training data.

Notice that $Cv(1) > 99\%$, which suggests using only one principal component. The unreconstructed variance using one principal component for the variables selected is

$$u^T = [0.0029 \quad 0.0066 \quad \text{n/a} \quad 0.0032 \quad 0.0152 \quad 0.0170 \quad 0.0059 \quad 0.0165 \quad \text{n/a}]. \quad (54)$$

All entries are much less than one, which indicates reliable reconstruction for the number of principal components chosen.

Figure 11 illustrates the normalized scores in time. For one principal component the univariate normal distribution test is applied instead of the chi-squared collective test for the scores. The confidence region is ± 3.

The norm of the projection on the residual subspace is used to detect the breakdown of the variable correlations. Figure 12 (top) shows the residuals' squared norm for the training data. The confidence limit, given by the horizontal line, is defined by Jackson (1991). The temporal violations of the

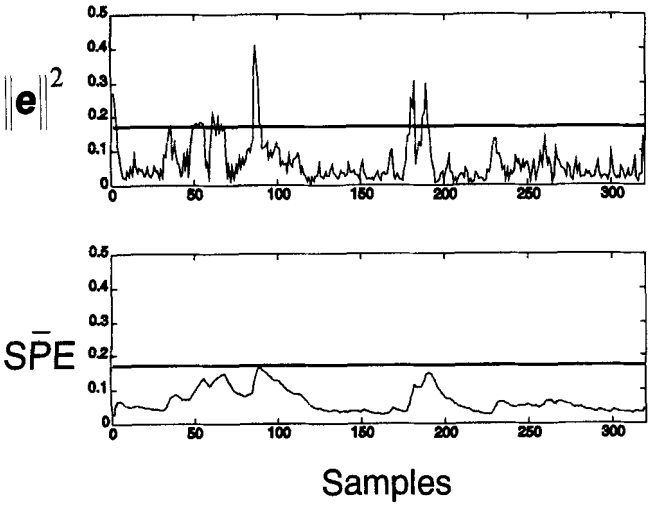

Figure 12. *SPE* for training data (top); the false alarm are eliminated by using $\overline{SPE}$ (bottom).

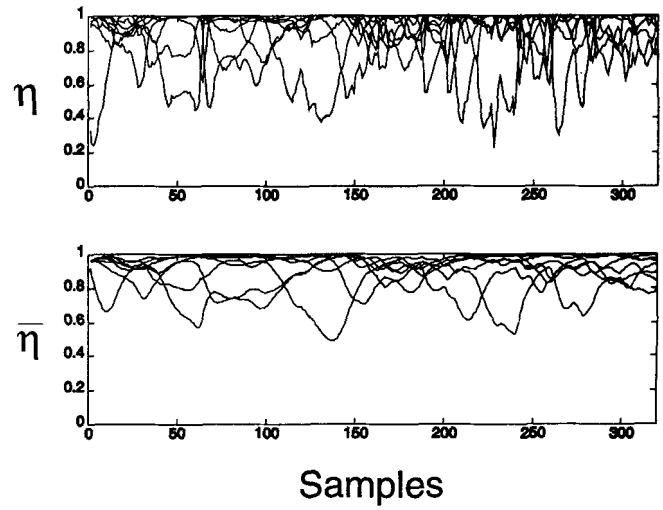

Figure 13. Validity index for training data before and after filtering.

Most of the high-frequency oscillations are eliminated by using η^2-EWMA (bottom).

confidence limits are false alarms. Comparing Figures 11 and 12, we observe that these false alarms correspond to significant transitions in the score's plot and can be avoided by using the SPE-EWMA described in Eq. 10. The forgetting factors chosen were $\lambda = 0.1$ and $\Lambda = 0.1I$, which reduces the false alarms and favors the identification of changes in the data mean, respectively. The bottom plot of Figure 12 shows how the SPE-EWMA eliminates the oscillations.

Figure 13 illustrates the profiles for the validity index. η^2-EWMA with $\Gamma = 0.1I$ is applied to reduce oscillations. Notice how some sensors are drastically penalized due to their large residuals. A threshold of 0.6 was chosen for a reasonable speed of identification.

Figures 14 to 17 illustrate all four faults introduced in the testing data at sample 350. The vertical dotted lines show where the fault has occurred. The symbols ○ and • represent the time of fault detection and identification, respectively. The validity index not only identifies the abnormal sensor, but also polarizes the effects of faults by raising the indices of the good sensors near unity. Table 8 summarizes the detec-

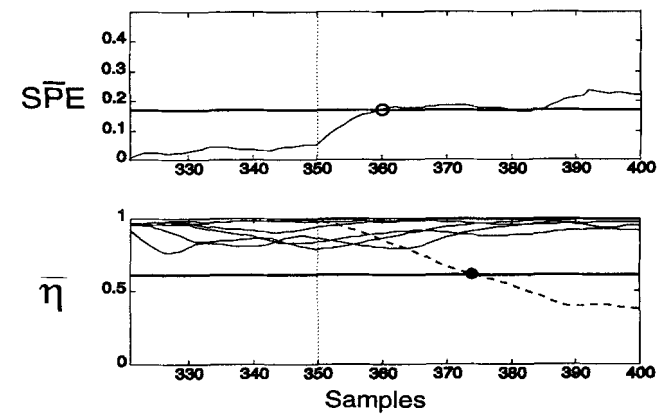

Figure 14. $\overline{SPE}$ detects, and $\bar{\eta}$ identifies a bias in the sensor 1.

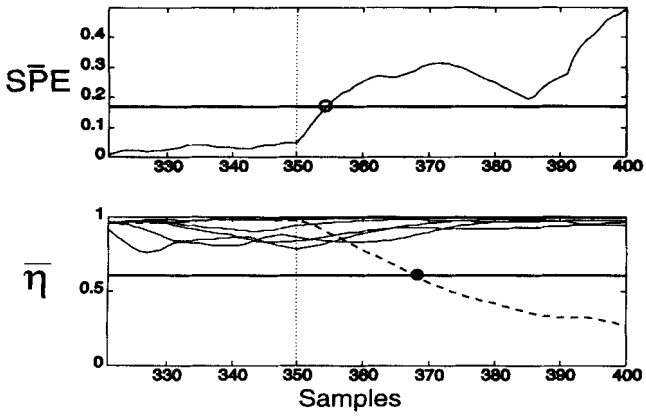

Figure 15. $\overline{SPE}$ detects, and $\overline{\eta}$ identifies a complete failure in sensor 2.

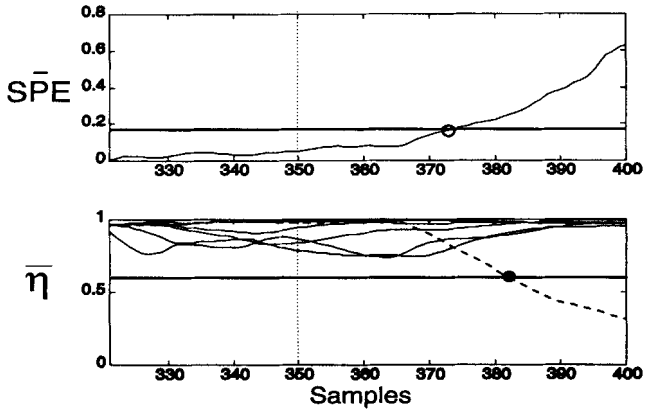

Figure 16. $\overline{SPE}$ detects, and $\overline{\eta}$ identifies a drift in the sensor 4.

tion and identification stages for sensor validation in these figures. The identification occurs a few samples after the detection. At this time the corrupted measurements are replaced by the reconstructed value obtained from Eq. 24. This allows the detection of subsequent faults that may occur from other sensors. The detection of precision degradation faults is possibly only for large δ because of the residual-EWMA. Therefore, another set of indicators with $\Lambda = I$ were used to detect such a type of fault. In general, two sets of indicators

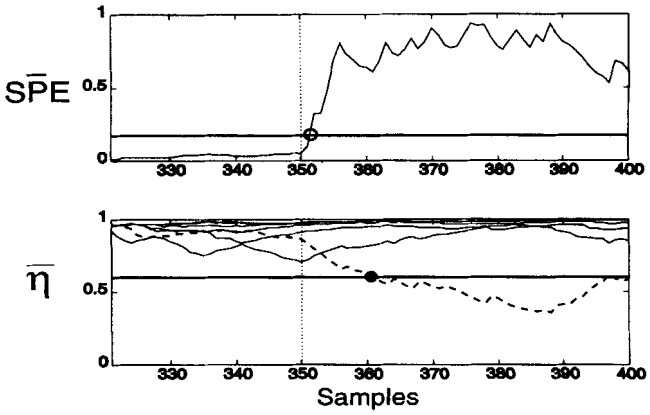

Figure 17. *SPE* detects, and $\overline{\eta}$ identifies the precision degradation of sensor 5.

with different forgetting factors are used to detect and identify faults related to variance and mean changes in the data. The case of two sensors violating simultaneously the PCA model is illustrated in Figure 18. Notice that this situation could happen due to an abnormal operating condition, which causes two sensors to break the correlation.

Conclusions and Future Considerations

This article makes use of PCA to develop a model that captures correlation among different sensors. Such a model is used to reconstruct a sensor measurement from the remaining sensors. Other modeling techniques may provide more reliable reconstructed values than PCA by rejecting the noise in the measurements and considering only the cross-correlation terms in the correlation matrix. However, the orthonormal decomposition of PCA permits one to understand and develop clear criteria for identification by reconstruction.

Two reconstruction techniques based on successive substitution and optimization are shown to provide the same result. The unreconstructed variance provides an estimation of the reliability of the reconstructed value. Different residuals were defined based on reconstructed values in order to obtain a reliable, sensitive indicator for fault identification. This is the sensor validity index that has a specific span between zero and one. The on-line implementation of such an index requires use of filters. The forgetting factor for residual-EWMA adjusts the sensitivity of the index to mean or vari-

Table 8. Number of Samples Required for Detection and Identification Measured from the Time the Fault Occurs

Fault	δ	Figure	Detection	Identification	Comments
Bias	0.35	14	10	24	From $\mathcal{E}\{A\} = cl(\alpha)$ we get $\delta = 0.35$ as the minimum expected bias in detection
Complete failure	0.5	15	8	18	From $\mathcal{E}\{A\} = cl(\alpha)$ we expect to detect any complete failure for reasonable noise levels
Drifting	0.02	16	24	32	This fault usually takes the longest for detection
Precision degradation	1.5	17	2	10	The residual-EWMA was eliminated
Abnormal operation	0.8	18	2	—	This condition was simulated as a bias for sensors 6 and 7

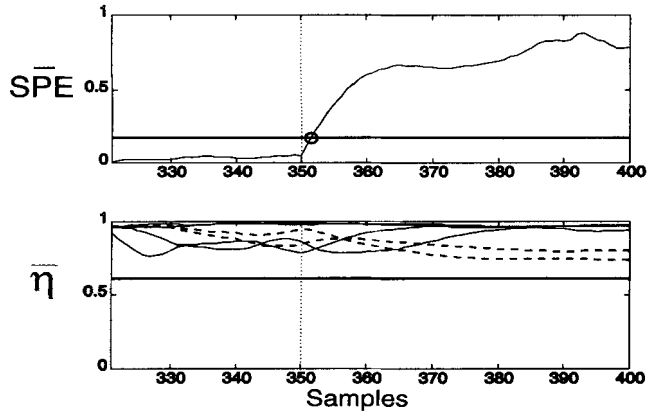

Figure 18. $\overline{SPE}$ detects an abnormal condition and $\bar{\eta}$ shows that it is caused by sensors 6 and 7.

ance changes. The EWMA for the SPE and η^2 reduce the false alarms and oscillations of the indicators, respectively.

This sensor fault detection and identification method was successfully implemented on a boiler process. The validity indices show a rapid identification after detection and a clear separation between the faulty sensor and the normal sensors. In general, two sets of forgetting factors can be applied for the detection and identification of mean and variance shift faults.

As future work, a validity index can be extended to nonlinear reconstruction of the sensors. Using nonlinear PCA instead of using linear PCA to define the model, we can apply autoassociative neural nets (Dong et al., 1995) trained using principal curve factors (Dong and McAvoy, 1994).

Notation

$Cv(\cdot)$ = cumulative variance
$\boldsymbol{C}$ = model matrix
c = column vector of $\boldsymbol{C}$
G = transfer function
$\boldsymbol{I}$ = identity matrix
$N(\cdot,\cdot)$ = normal distribution
$\boldsymbol{p}$ = row vector of $\boldsymbol{P}$
p = column vector of $\boldsymbol{P}$
r = correlation coefficient
$\boldsymbol{t}$ = row vector of $\boldsymbol{T}$
t = column vector of $\boldsymbol{T}$
$\boldsymbol{w}$ = column vector of $\boldsymbol{W}$
$\boldsymbol{W}$ = weigh matrix
x = column vector of $\boldsymbol{X}$
Z = z-transform variable
$\|\cdot\|$ = Euclidean norm
δ = fault magnitude
σ = standard deviation

Superscripts

T = transpose
$\bar{}$ = average
$\hat{}$ = estimation
$'$ = reconstructed data
$\breve{}$ = estimation using reconstructed data

Literature Cited

Basilevsky, A., *Statistical Factor Analysis and Related Methods*, Wiley-Interscience, New York (1994).

Crowder, S. V., "A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts," *Technometrics*, **29**, 401 (1987).

Da, R., and C. Lin, "Sensor Failure Detection with a Bank of Kalman Filters," *Proc. of the American Control Conf.*, p. 1122 (1995).

Dong, D., and T. J. McAvoy, "Nonlinear Principal Component Analysis—Based on Principal Curves and Neural Networks," *Proc. of the American Control Conf.*, p. 1284 (1994).

Dong, D., T. J. McAvoy, and L. J. Chang, "Emission Monitoring Using Multivariable Soft Sensors," *Proc. of the American Control Conf.*, p. 761 (1995).

Dunia, R., J. Qin, and T. F. Edgar, "Multivariable Process Monitoring Using Nonlinear Approaches," *Proc. American Control Conf.*, p. 756 (1995).

Fantoni, P., and A. Mazzola, "Applications of Autoassociative Neural Networks for Signal Validation in Accident Management," *Proc. of the IAEA Specialist Meet. on Advanced Information Methods and Artificial Intelligence in Nuclear Power Plant Control Rooms*, International Atomic Energy Agency, New York (1994).

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

Gnanadesikan, R., *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York (1977).

Hotelling, H., C. Eisenhart, M. Hastay, and W. A. Wallis, *Multivariate Quality Control*, McGraw-Hill, New York (1947).

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

Kramer, M. A., "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks," *AIChE J.*, **37**, 233 (1991).

Kramer, M. A., "Autoassociative Neural Networks," *Comput. Chem. Eng.*, **16**, 313 (1992).

Lucas, J. M., and M. S. Saccucci, "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, **32**, 1 (1990).

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

MacGregor, J. F., P. Nomikos, and T. Kourti, "Multivariable Statistical Process Control of Batch Processes Using PCA and PLS," *IFAC, ADCHEM*, International Federation of Automatic Control, p. 525 (1994b).

MacGregor, J. F., and T. Kourti, "Statistical Process Control of Multivariable Processes," *Contr. Eng. Pr.*, **3**(3), 403 (1995).

Mah, R. S. H., and A. C. Tamhane, "Detection of Gross Errors in Process Data," *AIChE J.*, **28**, 828 (1982).

Martens, H., and T. Naes, *Multivariable Calibration*, Wiley, New York (1989).

McBrayer, K., "Description of Bias Identification Method used with NDDR," Tech. Rep., Chemical Engineering Dept., Univ. of Texas at Austin (1995).

Ogata, K., *Discrete Time Control Systems*, Prentice Hall, Englewood Cliffs, NJ (1987).

Piovoso, M. J., K. A. Kosanovich, and R. K. Pearson, "Monitoring of Process Performance in Real Time," *Proc. Amer. Control Conf.*, p. 2359 (1992).

Tong, H., and C. M. Crowe, "Detection of Gross Errors in Data Reconciliation by Principal Component Analysis," *AIChE J.*, **41**, 1712 (1995).

Wise, B. M., and N. L. Ricker, "Recent Advances in Multivariate Statistical Process Control: Improving Robustness and Sensitivity," *Proc. IFAC. ADCHEM Symp.*, International Federation of Automatic Control, p. 125 (1991).

Wold, S., K. Esbensen, and P. Geladi, "Principal Component Analysis," *Chem. Intell. Lab. Syst.*, **2**, 37 (1987).

Appendix

For simplicity, assume the last sensor is missing/faulty. Substitution of $i = m$ in Eq. 24 gives:

$$z_m = \frac{[\boldsymbol{x}_{-m}^T \quad 0]\boldsymbol{c}_m}{1 - c_{mm}},$$

which in terms of $P^T = [P^T_{-m} \quad p_m]$ is

$$z_m = \frac{[x^T_{-m} \quad 0]Pp_m}{1 - p^T_m p_m} = \frac{x^T_{-m} P_{-m} p_m}{1 - p^T_m p_m}.$$

Substitution of z_m in the original sample vector gives

$$\acute{x}_m = \begin{bmatrix} x_{-m} \\ z_m \end{bmatrix} = \begin{bmatrix} I_{m-1} \\ \dfrac{p^T_m P^T_{-m}}{1 - p^T_m p_m} \end{bmatrix} x_{-m}.$$

The model estimation using the reconstructed value z_m is denoted by $\check{x}_m$,

$$\check{x}_m = PP^T \acute{x}_m = P\left(P^T_{-m} + \frac{p_m p^T_m P^T_{-m}}{1 - p^T_m p_m} \right) x_{-m}$$

$$= P\left(I_l + \frac{p_m p^T_m}{1 - p^T_m p_m} \right) P^T_{-m} x_{-m}.$$

Using the matrix inversion lemma (Ogata, 1987),

$$\check{x}_m = P(I_l - p_m p^T_m)^{-1} P^T_{-m} x_{-m}.$$

Since

$$P^T P = I_l = P^T_{-m} P_{-m} + p_m p^T_m,$$

we obtain

$$\check{x}_m = P(P^T_{-m} P_{-m})^{-1} P^T_{-m} x_{-m},$$

which is the expression used by Martens and Naes (1989) when the last measurement is missing. The vector x and the matrix P can be rearranged to keep the last measurement as the reconstructed one. Therefore, the iterative approach and the calculation of missing data proposed by Martens and Naes provide the same reconstructed values.

Manuscript received Oct. 18, 1995, and revision received Apr. 5, 1996.