Instrumentation Design and Upgrade for Principal Components Analysis Monitoring

Estanislao Musulin,[†] Miguel Bagajewicz,[‡] José M. Nougués,[†] and Luis Puigjaner*,[†]

Chemical Engineering Department, Universitat Politècnica de Catalunya, ETSEIB, Av. Diagonal 647, E-08028 - Barcelona, Spain, and University of Oklahoma, 100 East Boyd T-335, Norman, Oklahoma 73019

In this paper, a methodology to design instrument networks optimally selected for the use of PCA-based monitoring in process plants is presented. Sensors are placed or signals are selected to minimize the sum of investment cost and the cost of not detecting a fault. This optimization problem is solved using Genetic Algorithms (GA). It is shown that for existing sensor networks, the use of fewer signals than the available ones can be, in some cases, advantageous for monitoring purposes.

Introduction

Nowadays chemical process industry produces a huge amount of information readily available through Distributed Control Systems (DCS) and data acquisition systems. However, if it is not appropriately processed, this information is of limited use. Many industrial applications have shown the ability of techniques associated with Multivariate Statistical Process Monitoring (MSPM) to monitor chemical processes.[1] Among these methods, the PCA that builds a linear statistical model of the process using the correlation present in process data is the most widely used.

The first descriptions of PCA were introduced by Pearson[2] and Hotelling.[3] Recently PCA methods have been extended to perform fault detection and diagnosis.[4,5] PCA can be applied without a model of the process, in an easy and fast way, and without excessive computational requirements. It performs a dimensionality reduction of the process variables generating new noncorrelated variables called Principal Components (PCs). PCs and some related statistics can be used online as part of a classification algorithm or an expert system to perform fault diagnosis.[6] Nevertheless, because of uncertainty, noise, disturbances, and nonlinearities, fault detection and isolation present many difficulties.[6−8]

One of the key issues for the success of MSPM is the quality of data. In addition, fault detection and isolation depends on the number and placement of the different sensors in a system. Therefore, it is important that sensors are placed in the right location.

Articles that discuss the sensor placement problem based on fault modeling have been published. They use the information provided by a Direct Graph (DG) model[9] and Signed Direct Graph (SDG) model.[10] These studies analyze the behavior of process signals when faults occur to locate sensors in order to optimize fault isolation. Bagajewicz[11] presents a series of techniques to locate instrumentation that are mostly based on the assumption that data reconciliation is the monitoring paradigm used to obtain unbiased estimators of process variables with the required precision.[12] In a more recent article,[13] he also presents a technique to isolate faults together with the use of redundancy through data reconciliation to obtain estimators of process variables.

In most of the articles dealing with PCA it is assumed that the sensor network is already in place. Thus, the relationship between sensor location and PCA performance is rarely discussed in the literature. Recently, Wang et al.[14] utilized the algorithm proposed by Raghuraj et al.[9] to select the set of sensors so as to optimize fault detection and isolation. Although PCA is used to test the algorithm results, the performance of PCA monitoring, the importance of the faults, and the cost of the sensors are not taken into account during the sensor location procedure.

In this work, a design method for sensor placement and signal selection that improves fault detection based on PCA is presented. It considers sensors cost and the different relative impact of faults/disturbances. PCA and the means to identify a good quality PCA are first reviewed. Next, a fault size penalization (FSP) model is introduced. It penalizes the sensor network taking into account the size of the faults at the detection time. Then, a GA is used to improve the sensor network performance by minimizing the sensors cost and, more importantly, the FSP. Finally, the procedure is validated using two relevant case studies.

Model Building Performance Measures

Consider a matrix $\mathbf{X}$ (of dimension $m \times n$) containing data corresponding to m samples of n process variables. Each column of $\mathbf{X}$ is supposed to follow a normal probability distribution and must be normalized with zero mean and unit variance. Let $\bar{\mathbf{X}}$ be this normalized data matrix, and $\mathbf{R}$ its corresponding correlation matrix:

$$\mathbf{R} = \frac{\bar{\mathbf{X}} \cdot \bar{\mathbf{X}}^T}{m - 1} \qquad (1)$$

Then, performing singular value decomposition on $\mathbf{R}$,

$$\mathbf{R} = \mathbf{U} \mathbf{D}_\lambda \mathbf{U}^T \qquad (2)$$

a diagonal matrix $\mathbf{D}_\lambda = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$ where λ_i are the eigenvalues of $\mathbf{R}$ sorted in decreasing order $\lambda_1 > \lambda_2 > \cdots > \lambda_n$, is obtained. We note that $\mathbf{R}$ is square and symmetric, so that the generalized SVD, which is

---
* To whom correspondence should be addressed. Tel.: +34-93-401.66.78. Fax: +34-93-401.09.79. E-mail: Luis.Puigjaner@upc.es.
† Universitat Politècnica de Catalunya.
‡ University of Oklahoma.

applied to nonsymmetric matrices, reduces to the above presented case. The corresponding eigenvectors p_I (columns of $\mathbf{U}$) are the principal components and form an orthonormal base in $\mathscr{R}^n$. It is possible to divide the principal components in two orthogonal sets, $\mathbf{P} = [p_1, p_2, ..., p_A]$ and $\bar{\mathbf{P}} = [p_{A+1}, p_{A+2}, ..., p_n]$. The first containing most of the process variance and the second describing the variance due to the noise (called the residual subspace). Methods to select the proper number of PCs are reviewed by Himes et al.[15]

A reduction of dimensionality is made by projecting every normalized sample vector x' in the principal component space generated by $\mathbf{P}$, obtaining $t = \mathbf{P}^T x'$, which is called the principal score vector. Because of the normalization of $\bar{\mathbf{X}}$, when the process is in control, x' is projected at the origin, or in its proximity. Then, the state of the process can be monitored using the Hotelling's statistic, which is a measure of the Mahalanobis distance to the origin of the new space,

$$T^2 = ||\mathbf{D}_\lambda^{-1/2}\mathbf{P}^T x'|| \qquad (3)$$

The test consists of declaring normal operation if $T^2 \leq \delta_T^2$, where δ_T^2 is the control confidence limit of T^2. However, monitoring based only on this statistic is not sufficient because it only shows the deviations in the principal components subspace. If new data are not contained in the reference set of data, the projection of the new observation will move to the residual space, and it is possible that the anomaly is not detected by the T^2 test. To detect this kind of deviation the Squared Predictive Error statistic (SPE), also called Q statistic, is usually used. SPE is defined as the sum of the quadratic error between x' and the reconstructed signal after the dimensionality reduction

$$\Delta x' = x' - \mathbf{P}\mathbf{P}^T x' = (\mathbf{I} - \mathbf{P}\mathbf{P}^T)x' = \tilde{\mathbf{C}}x' \qquad (4)$$

Thus,

$$\mathrm{SPE} = ||\tilde{\mathbf{C}}x'||^2 \qquad (5)$$

if $\mathrm{SPE} \leq \delta_{\mathrm{SPE}}^2$, where δ_{SPE}^2 is the control limit for SPE, the data vector is declared normal.

The threshold limits δ_T^2 and δ_{SPE}^2 can be obtained in different ways.[17,16] In this work thresholds are fixed in such a way that a certain percentage (typically 1% and 5%) of the statistics values are under them.

Fault Detectability

Wang et al.[19] derive sufficient conditions so that a fault is detectable, that is, the SPE or T^2 statistics exceed the control limit. They are summarized next.

T^2 **Detectability.** Let us suppose that in a process there exist J faults considered of importance. If a set of located sensors is given, it is possible to describe each fault F_j in the fault set $\{F_j\}_{j=1}^J$ by a corresponding subspace in $\mathscr{R}^n$. This subspace is called a fault subspace.[18] Let $\{\Theta_j\}_{j=1}^J$, be a set of orthonormal basis that describes the fault subspaces, then $\Theta_j \in \mathscr{R}^{n \times l_j}$ where n is the number of sensors and l_j is the dimension of the jth fault subspace (i.e. l_j is the number of dimensions in which the fault impacts). In the case of unidimensional faults, that is, faults that affect only one process variable, Θ_j is a vector of dimension n with unit norm. In case of a single fault with multiple dimensions, that is, faults that affect several process variables, Θ_j is

chosen to be orthonormal of dimension $n \times l_j$ and each column of Θ_j is zero except for a 1 (or -1 depending on the sign of the deviation) in the affected variable. The vector of coordinates of a fault $f_j \in \mathscr{R}^{l_j}$ is defined as the deviations that the fault F_j causes in the process variables. Note that the value of the elements of f_j may change depending on how the actual fault develops. The fault magnitude is defined as $||f_j||$.

When a fault occurs the vector of measured process variables x' can be expressed as follows

$$x' = x_0 + \Theta_j \; f_j \qquad (6)$$

where x_0 represents the measurements under normal conditions and the second term represents the deviation introduced by the fault. Therefore, the value of T^2 can be evaluated using the following expression

$$T^2 = ||D_\lambda^{-1/2}\mathbf{P}^T x'||^2 = ||D_\lambda^{-1/2}\mathbf{P}^T(x_0 + \Theta_j \; f_j)||^2 \quad (7)$$

Since, the expression

$$||D_\lambda^{-1/2}\mathbf{P}^T x'|| \geq |||D_\lambda^{-1/2}\mathbf{P}^T x_0|| - ||D_\lambda^{-1/2}\mathbf{P}^T \Theta_j \; f_j||| \quad (8)$$

is true, and $||D_\lambda^{-1/2}\mathbf{P}^T x_0|| \leq \delta_T$ by construction, it follows that the sufficient condition for fault detectability using the T^2 statistic is

$$||D_\lambda^{-1/2}\mathbf{P}^T \Theta_j \; f_j|| \geq 2\delta_T \qquad (9)$$

A more restrictive sufficient condition is found taking into account that

$$||D_\lambda^{-1/2}\mathbf{P}^T \Theta_j||\cdot||f_j|| \geq ||D_\lambda^{-1/2}\mathbf{P}^T \Theta_j \; f_j|| \geq 2\delta_T \quad (10)$$

which is equivalent to

$$||f_j|| \geq ||f_{T_j}|| = 2\sigma_{\max}^{-1}(D_\lambda^{-1/2}\mathbf{P}^T \Theta_j)\delta_T \qquad (11)$$

SPE Detectability. When a fault occurs, the SPE statistic is given by

$$\mathrm{SPE} = ||\tilde{\mathbf{C}}x'||^2 = ||\tilde{\mathbf{C}}x' + \tilde{\mathbf{C}}\Theta_j \; f_j||^2 \qquad (12)$$

However,

$$||\tilde{\mathbf{C}}x'|| \geq |||\tilde{\mathbf{C}}x_0|| - ||\tilde{\mathbf{C}}\Theta_j \; f_j||| \qquad (13)$$

and since $||\tilde{\mathbf{C}}x'|| \leq \delta_{\mathrm{SPE}}$; then

$$||\tilde{\mathbf{C}}\Theta_j \; f_j|| \geq 2\delta_{\mathrm{SPE}} \qquad (14)$$

But since

$$||\tilde{\mathbf{C}}\Theta_j||\cdot||f_j|| \geq ||\tilde{\mathbf{C}}\Theta_j \; f_j|| \qquad (15)$$

one can write the following more restrictive condition:

$$||f_{\mathrm{SPE}_j}|| = 2\sigma_{\max}^{-1}(\tilde{\mathbf{C}}\Theta_j)\delta_{\mathrm{SPE}} \qquad (16)$$

where $||f_{T_j}||$ and $||f_{\mathrm{SPE}_j}||$ are said to be the critical fault magnitudes (CFMs). They are approximations to the minimum fault magnitude $||f_j||$ detectable in the T^2 and SPE tests, respectively.

Based on the detectability conditions one can formulate the optimization problem that allows selecting the best set of sensors. The problem can be expressed as one that minimizes both the sensor costs and the cost

associated to the CFMs. This minimization should take into account the different importance of the faults.

Fault Size Penalization Model

We define a sensor network using a binary vector Q_i. Each element of Q_i represents the presence or absence of a sensor. The length of Q_i is equal to the number of process variables that can be potentially measured (N)

$$Q_i = [q_1 \ q_2 \ ... \ q_{N-1} \ q_N] \tag{17}$$

Let suppose that $\{\bar{\mathbf{X}}_j\}_{j=1}^J$ is a set of data matrices, each matrix corresponding to a given abnormal state of the process. The columns of $\bar{\mathbf{X}}_j$ are the sensor measurements under abnormal state. Since they are normalized with the mean and variance of the normal state, some of them are deviated and their means are not null.

To evaluate the performance of the sensor network Q_i, the Minimum Critical Fault Magnitudes ($MCFM_{i,j}$) are defined as follows

$$MCFM_{i,j} = \min\{||f_{T_{i,j}}||, ||f_{SPE_{i,j}}||\} \tag{18}$$

where $||f_{T_{i,j}}||$ and $||f_{SPE_{i,j}}||$ are the CFMs corresponding to each F_j when PCA is performed using the signals selected by Q_i. $MCFM_{i,j}$ represents approximately the lower fault magnitude of F_j to be detectable either by the T^2 or SPE tests.

The performance of Q_i is penalized by a weighed sum of the $MCFM_{i,j}$; this penalization is called Fault Size Penalization FSP_i and is given by

$$FSP_{i,j} = w_{i,j} \cdot MCFM_{i,j} \tag{19}$$

$$FSP_i = \sum_{j=1}^J FSP_{i,j} \tag{20}$$

where the weights $w_{i,j}$ should be selected reflecting the importance of F_j in relation to other abnormal states.

As a rule, faults must be detected before process variables reach any value that can cause shut downs, hazardous situations, and so on. Therefore, it must be taken into account that the process can only deal with a certain range of process variable values. The limits of the process variables are called Process Deviation Limits (PDL). On the other hand, some small disturbances (e.g. due to process noise) might not be of interest. If knowledge about these limits is available, they can be introduced into the FSP model by defining two penalization bounds:

(i) $f_{\sup_{i,j}}$ is defined as the size of F_j at which the fault has to be compulsory detected when Q_i is used. If the MCFM related to this fault and sensor network is greater than this limit, the FSP is considered infinite.

(ii) If the MCFM is under a lower limit $f_{\inf_{i,j}}$, the related FSP is zero. That is to say, a deviation of size inferior of $f_{\inf_{i,j}}$ is considered as noise, and its lack of detection is not penalized.

To obtain $f_{\sup_{i,j}}$, the process data matrices ($\mathbf{X}_j$) are inspected in order to find the process data vectors when any process variable exceeds one of the prespecified limits (PDL). Lets x°_j be the process data vectors sampled in this situation for each of the faults (one for each fault). Therefore, the normalized versions of x°_j are the fault parameters vectors of F_j at which the fault has to be compulsory detected (say $\hat{f}^\circ_j$) (see eq 6). If all the
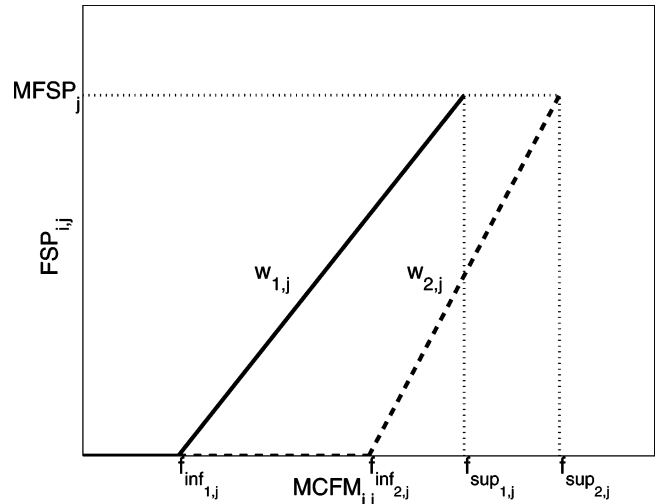

Figure 1. FSP model for the same fault and two different sensor networks.

process variables are measured, $f_{\sup_{i,j}}$ can be defined directly as the norm of f°_j. However, if only some of them are measured (i.e. the sensor Q_i is used), $f_{\sup_{i,j}}$ can be calculated as

$$f_{\sup_{i,j}} = ||\mathbf{f}^\circ_j * Q_i|| \tag{21}$$

where "*" is the Hadamark product of vectors (element by element multiplication).

On the other hand, a conservative estimation of $f_{\inf_{i,j}}$ is

$$f_{\inf_{i,j}} = ||[\sigma_1, ..., \sigma_N] * Q_i|| \tag{22}$$

where σ_k is the variance of the kth column of the normalized data matrix under normal operational conditions $\bar{\mathbf{X}}$. Since the data vectors are normalized

$$f_{\inf_{i,j}} = ||[1, ..., 1] * Q_i|| = \sqrt[n_i]{n_i} \tag{23}$$

where n_i is the number of sensors used in Q_i. Therefore, with this estimation, $f_{\inf_{i,j}}$ depends only on the number of selected sensors and is independent of the fault.

Finally, the FSP model can be summarized as follows:

$$FSP_{i,j} = \begin{cases} 0 & MCFM_{i,j} < f_{\inf_{i,j}} \\ w_{i,j}(MCFM_{i,j} - f_{\inf_{i,j}}) & f_{\inf_{i,j}} < MCFM_{i,j} < f_{\sup_{i,j}} \\ +\infty & MCFM_{i,j} > f_{\sup_{i,j}} \end{cases} \tag{24}$$

Although the FSP bounds change depending on the selected sensor network Q_i (eqs 21 and 22), the penalization associated with a given deviation has to remain the same and must be independent of Q_i. Thus, it should be only a function of f°_j. Then, a new set of constants is defined as the penalization assigned to Q_i when it detects F_j at the size $f_{\sup_{i,j}}$. This constant, which is called Maximum Fault Size Penalization (MFSP), is defined as the penalization assigned to Q_i when $MCFM_{i,j} = f_{\sup_{i,j}}$ (Figure 1). Therefore, the penalization weigths are defined by

$$w_{i,j} = \frac{\text{MFSP}_j}{(f_{\text{sup}i,j} - f_{\text{inf}_{i,j}})} \qquad (25)$$

Besides the minimization of the FSP, the cost of the located sensors should be taken into account during sensor network design. Thus, the cost of every sensor is estimated (c_k), and a sensor cost vector is defined as follows:

$$C = [c_1 \; c_2 \; ... \; c_{N-1} \; c_N] \qquad (26)$$

Consequently, the cost of the sensor network can be written as

$$\text{SNC}_i = \sum_{k=1}^{N} q_k \cdot c_k \qquad (27)$$

Finally, the objective function (OF) is defined as the sum of both penalty indexes, FSP and SNC:

$$\text{OF}_i = \text{FSP}_i + \text{SNC}_i \qquad (28)$$

This OF can be seen as a global penalty index of the sensor network. Note that both partial indexes and the OF depend on the selected set of signals. To improve the integration of the penalty indexes, this global objective function may be extended to be a true cost associated to Q_i. To do that it is sufficient to define the MFSP$_j$ in terms of an economic penalization. In this way, the global penalty index of the sensor network is expressed in economics terms. We note, in addition, that one can consider that faults have some other effects that are not always measurable in economic terms. Thus, one can consider different weights in a goal programming approach to investigate how the sensor network needs to be constructed for a different relative importance of cost in the objective.

The model developed can be used to evaluate the performance of sensor networks of existing plants or simulated ones. Moreover, the model can be integrated into an optimization algorithm to attain the optimal performance of the sensor network. Optimization algorithms can be used either to select signals of processes with an existing sensor network or to decide the best location of sensors when designing the plant.

Genetic Algorithm Approach

We chose a Genetic Algorithm (GA) technique to solve this problem, which is part of a vast class of stochastic algorithms that consider many points in the search space simultaneously and therefore have a reduced chance of converging to a local optima.[20]

In the classical GA formalism Figure 2, a set of N_{ind} candidate solutions (population) are generated randomly. The potential solution in the multidimensional search space (individual) is coded as a vector, called chromosome (which consists of a string of genes, each representing a single feature).

The goodness of each individual in the population is evaluated by utilizing a prespecified fitness criterion Φ. Upon assessing fitness of all the chromosomes in the population, a new generation of individuals is created from the current population by using the selection, crossover, and mutation operators.[21,22]

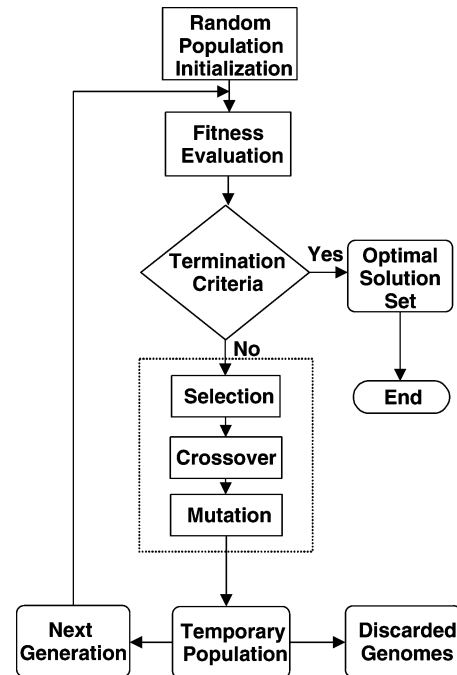One of the main steps in a GA approach is the codification of the individuals. In the proposed algorithm

Figure 2. GA optimization loop.

each chromosome represents a sensor network Q_i and is codified as in eq 17.

The individuals are considered to be more fit if their penalty index OF value is lower. Therefore the fitness is evaluated as the inverse of OF

$$\Phi_i = \frac{1}{\text{OF}_i} \qquad (29)$$

The roulette-wheel selection and multipoints crossover operators have been used. Half of the individuals ($N_{\text{ind}}/2$) has been selected to be crossed in each generation. The mutation probability P_m is usually much lower than the crossover probability P_c ($P_m = Q_S/L_{\text{ind}}$). Once mutation has been applied, a new set of $N_{\text{ind}}/2$ individuals is available. This new set of individuals is then merged with the best fitted $N_{\text{ind}}/2$ individuals of the previous generation to obtain a new generation of N_{ind} individuals. The algorithm is stopped when the number of generations reaches a predefined maximum value, and when the current population does not give sufficient improvement compared with the population G generations before (typically $G = 5$).

To calculate the fitness of each individual (eqs 11, 16, 18, 21, 22, 25, and 29) the following tasks have to be performed in advance: (i) sensors cost estimation, (ii) $f°_j$ and MFSP$_j$ definition, and (iii) fault subspace estimation (Θ_j).

The first two tasks have been already discussed. The estimation of Θ_j for sensor faults, which are in general unidimensional, is straightforward. However, in process faults, which are in general propagated to several sensors, the fault subspace is multidimensional, and its estimation becomes more difficult. Methods such as DG[9] and SDG[10] can be used. The estimation of the fault propagation has not been done in this paper using these techniques; instead of that an outlier detection method is proposed to investigate such fault propagation. This method has shown to be good enough to apply the proposed algorithm in two case studies. In any case, a fault subspace defined this way, by containing essential
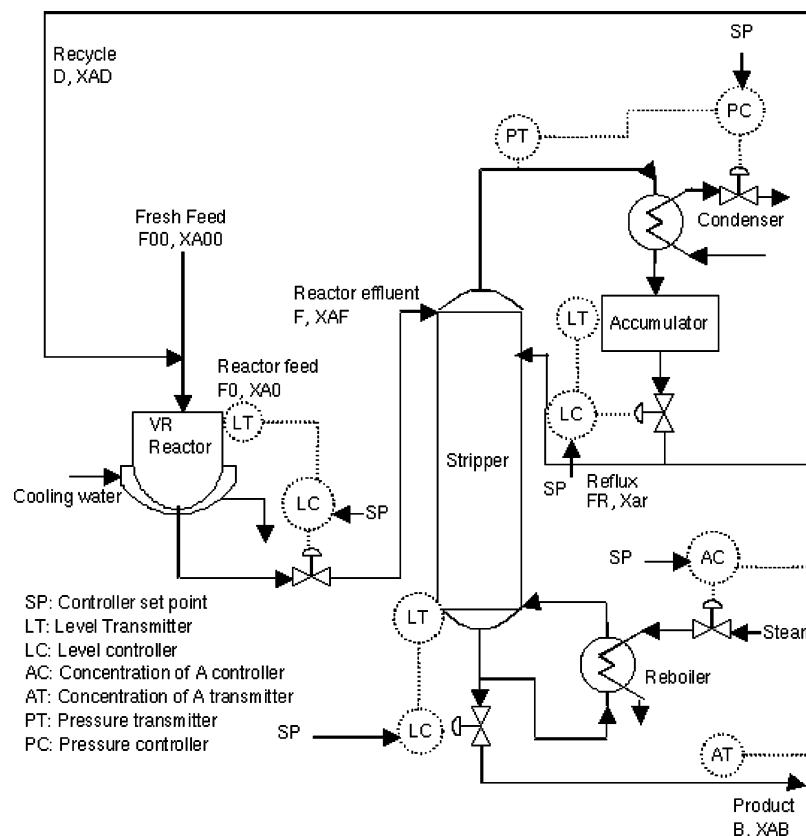
Figure 3. Case study I: process flow-sheet.

Table 1. Case Study I: Faults Definitions

fault name	description	MFSP$_j$ [$]
F_1	V_R level sensor drift	10000
F_2	V_R level sensor bias	10000
F_3	steam leak in the reboiler	5000
F_4	failure in reactor's output pump	10000

characteristics of the deviations, is good enough to evaluate the CFMs[19] and then the MCFM.

Case Studies

Case Study I: Chemical Plant with Recycle. The first case study was proposed by Belanger and Luyben.[23] The plant has two main operating units: a reactor and a stripper (Figure 3). Fresh feed, consisting of reactant *A* and some product *B*, is fed to the reactor. The reactor is a continuous stirred tank. The irreversible reaction that occurs is of first order: $A \rightarrow B$. The reactor output is fed to the stripper, where most of unreacted *A* is separated from product *B*. The plant's product, with a small mol fraction of *A* (X_{AB}) is obtained at the stripper's bottom. The stripper's output at the top is recycled to the reactor. The physical properties of the components are the same except for the relative volatility. The reaction rate is given by $r = V_R \times k \times X_f$, where *r* is the rate of reaction, V_R is the holdup of the reactor, *k* is the reaction rate constant, and X_f is the mol fraction of *A* in the reactor. To apply the proposed methodology in this case study, data from a process simulator built in Matlab have been used.

The considered fault conditions and the penalization (MFSP$_j$) can be seen in Table 1. Table 2 lists the 13 potentially measured process variables, the estimated costs of the associated sensors and the deviation limits for the process variables. As was explained in section

Table 2. Case Study I: Potentially Located Sensors and Variables Characteristics

index	description	c_k	SS value	PDL (%)
q_1	column input composition	5000	0.376	+10, −5
q_2	distillate flow	500	1.4 kmol·min^{-1}	±10
q_3	condenser valve	50	40.91%	
q_4	condenser level	50	10.14 Kmol	±10
q_5	product flow	500	1.8 kmol·min^{-1}	±10
q_6	column base valve	50	42.4%	±10
q_7	column base level	50	10.18kmol	±10
q_8	steam flow	500	12.63 kmol·min^{-1}	±10
q_9	steam valve	50	52.7%	
q_{10}	output composition	50	0.1	±10
q_{11}	reactor level	50	750.64 Kmol	±10
q_{12}	reactor output flow	50	3.2 kmol·min^{-1}	±10
q_{13}	reactor input flow	500	3.2 kmol·min^{-1}	±10

[*], process data ($\mathbf{X}_j$) are compared with these limits in order to find the data vector when any of the limits is exceeded ($x°_j$). Figure 4 shows the profile of X_f, which is the first variable that goes out of bounds for F_1, F_2, and F_4, and its deviation limits for all the fault cases. Note that X_f is not controlled. A process data vector is obtained for each fault case $x°_j$ in these situations. Then, a normalized version of them $f°_j$ is easily obtained (subtracting the process variables mean and dividing by the process variables variance). In the case of F_3, the limits are not exceeded for any variable. This is due to the fact that the control valve hides the leak and the disturbance is not propagated to other process variables. Therefore, the economic impact of this fault is directly related to losses of vapor. In this case study a 1% vapor leakage is considered as the maximum limit. $x°_3$ corresponds to a data vector obtained when a leakage of this size occurs.

Usually, the presence of noise in the data matrices involves the fact that deviations due to noise are
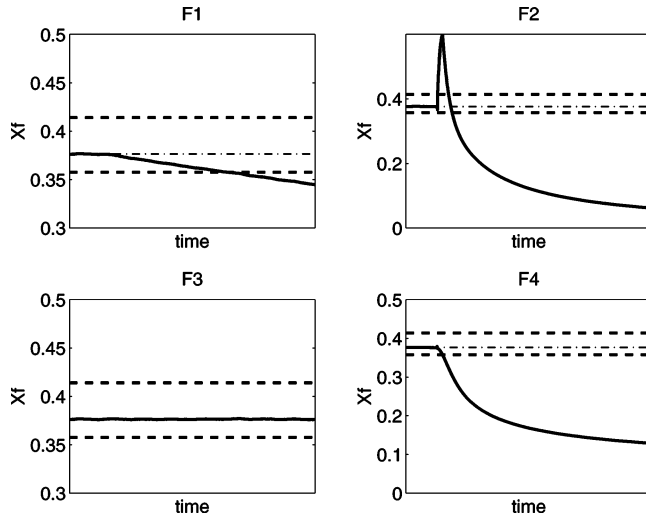
Figure 4. Case study I: deviation of the column input composition in all the considered fault cases.

attributed to the faults. Therefore, the robustness of the evaluation of the fault propagation is improved by analyzing a filtered version of each of the process signals. The signals stored in the normalized data matrices are filtered as follows:

$$x'F(t) = \text{mean}(x'(t), x'(t - Ts), x'(t - 2Ts),$$
$$x'(t - 3Ts), x'(t - 4Ts)) \quad (30)$$

Next, the signals are tested using a simple robust scaling algorithm based on the well-known "3δ edit rule".[24] This rule states that when $|x'_F| > 3$ the variable is considered to be deviated from its normal state. The test is then performed on the filtered version of $x^o{}_j$. Results are presented in Table 3. Using this information, the fault subspaces can be easily defined. They are

$$\Theta_1 =$$

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}^T$$
$$(31)$$

$$\Theta_2 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T$$
$$(32)$$

Table 3. Case Study I: Fault Incidence

	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}
F_1	-1	-1	-1	-1	0	0	0	-1	-1	0	0	-1	-1
F_2	-1	1	1	1	-1	-1	-1	1	1	-1	1	1	1
F_3	0	0	0	0	0	0	0	1	1	0	0	0	0
F_4	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	1	-1	-1

$$\Theta_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T \quad (33)$$

$$\Theta_4 =$$

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}^T$$
$$(34)$$

When F_4 occurs, most of the variables are deviated. The fault parameter vector at which F_4 has to be compulsory detected results

$$f^o{}_4 = [89,52,76,63,52,61,52,64,64,0,12658,58,28] \quad (35)$$

Sensor q_{11} is the most influenced by this fault and because this sensor is not expensive (Table 1) compared to MFSP$_4$ (Table 2), it is expected that the algorithm will select it. Moreover, from Θ_3 it can be seen that only q_8 and q_9 are influenced by F_3. Thus, one of them should be in the selected sensor network.

The algorithm was applied with an initial population of 130 (i.e. $10 \cdot N$) individuals, $Q_S = 0.7$ and $N_G = 10$. The sensor network representing the best individual is as follows

$$Q_{S1} = [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 1 \ 0 \ 0] \quad (36)$$

and its related costs are FSP$_{S1}$ = \$8508, SC$_{S1}$ = \$600, and OF$_{S1}$ = \$9108. As it was expected, q_9 and q_{11} have been selected in order to detect F_3 and F_4. Only one additional sensor (q_2) is located to detect all the faults in an optimal way. To sum up, the algorithm suggests picking the sensors of the distillate flow and opening of the steam valve and reactor level. It is important to note the correlation between the result obtained and the set of faults initially defined.

Next, the performance of Q_{S1} was compared with another sensor network (Q_A), which contains additional sensors used to measure other signals disturbed by the defined faults

$$Q_A = [0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 1] \quad (37)$$

The costs are FSP$_A$ = \$1 069 462, SC$_A$ = \$7000, and OF$_A$ = \$1 076 462. It can be seen that not only the sensors cost increases but also the FSP, which imply an inferior detection performance. Figure 5 shows the
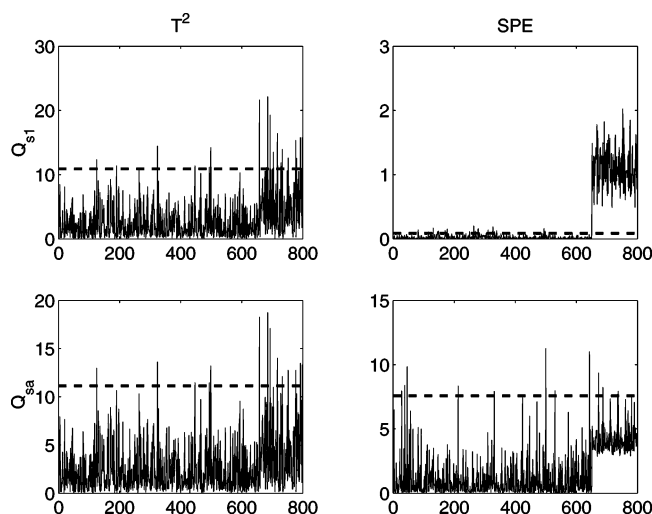
Figure 5. Case study I: comparison between of the detection performance of Q_{S1}, Q_A, and Q_{S2} in case of a steam leak of 0.1% in the reboiler.

comparison in the monitoring performance of these two individuals in case of a small steam leak in the reboiler (0.1%). Only the PCA monitoring system that uses Q_{S1} detects the fault because its SPE statistic clearly exceeds the 99% control limit. The statistics generated by using Q_A are also deviated, but these deviations are not enough to exceed the control limits. It can be noted that the use of more signals add more variance to the model. Then, the control limits end up being higher for Q_A. Most of the cost related to Q_{S1} are due to the FSP, and the sensors costs are negligible. The algorithm was run again considering that the sensors costs are 10 times higher than the previous values and the MFSP are 10 times lower. The solution is again the same

($Q_{S2} = Q_{S1}$), and the new related costs are as follows: $FSP_{S2} = \$851$, $SC_{S2} = \$6000$, and $OF_{S2} = \$6851$. These results show that in this case there is not a notable tradeoff between the two objectives (i.e. low sensor costs and good detection performance). On the contrary, the sensor network with the best detection performance (i.e low FSP) has also a low cost (i.e. low SC).

Case Study II: Tennessee Eastman Process. The proposed scheme was also applied to a well-known benchmark, the Tennessee Eastman Process (TE).[25] It involves the production of two products, G and H, from four reactants, A, C, D, and E. Additionally, there are two side reactions that occur and an inert B. All the reactions are irreversible and exothermic. There are 41 measured process variables and 12 manipulated variables.

In this work the decentralized control system proposed by McAvoy and Ye[26] has been used (Figure 6). The MATLAB simulation procedure follows that of Singhal.[27,28] The process disturbances listed in Table 4 have been simulated from time = 2 h to time = 50 h with a sample time of Ts = 1 min. The complete set of data includes measured and actuators signals

$$\mathbf{X} = \begin{bmatrix} \text{measurements} \\ \text{actuators signals} \end{bmatrix} \qquad (38)$$

We used our method to analyze the fault propagation. Table 4 shows an estimation of the costs associated to each fault. As discussed, these costs do not have to be rigorous, but the relative importance of the faults has to be taken into account. The sensors that are used for control purposes are considered of null cost.

The initial population for the GA is set to 520 (i.e. $10 \cdot N$), $N_G = 50$, $Q_S = 0.7$, and $G = 10$.
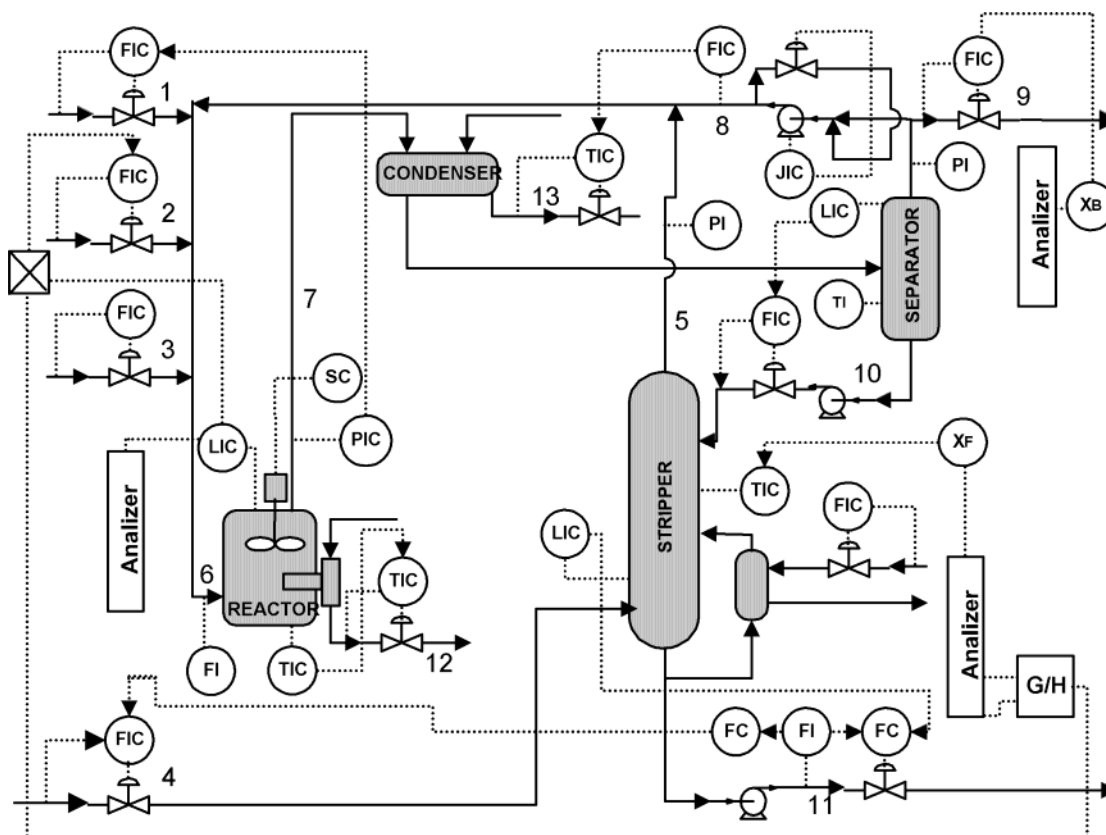

Figure 6. Tennessee Eastman process flow-sheet.

Table 4. Disturbances and Faults in the TE Process

disturbance	description	type	MFSP$_j$ [$]
IDV(1)	A/C feed ratio, B composition constant (stream 4)	step	7500
IDV(2)	B composition, A/C feed ratio constant (stream 4)	step	25000
IDV(3)	D feed temperature (stream 2)	step	95000
IDV(4)	reactor cooling water inlet temp	step	2500
IDV(5)	condenser cooling water inlet temp	step	10000
IDV(6)	A feed loss (stream 1)	step	500000
IDV(7)	C header pressure loss-reduced availability (stream 4)	step	500000
IDV(8)	A, B, C feed composition (stream 4)	random variation	15000
IDV(9)	D feed temperature (stream 2)	random variation	35000
IDV(10)	C feed temperature (stream 4)	random variation	5000
IDV(11)	reactor cooling water inlet temperature	random variation	15000
IDV(12)	condenser cooling water inlet temperature	random variation	35000
IDV(13)	reaction kinetics	slow drift	1000
IDV(14)	reactor cooling water valve	sticking	3750

Table 5. Reliability for Q_A and Q_S (Ts = 1 min)

|	SPE reliability		T^2 reliability	
IDV	Q_A	Q_S	Q_A	Q_S
1	99.86	93.81	99.72	100.00
2	99.58	91.02	98.78	99.58
3	1.77	3.63	1.39	23.20
4	3.68	1.07	1.97	2.38
5	1.90	0.65	1.21	1.37
6	99.96	100.00	99.96	100.00
7	99.96	91.91	99.96	100.00
8	99.17	80.01	98.89	99.28
9	5.14	3.27	4.65	5.12
10	85.80	21.00	21.17	76.56
11	78.97	59.96	23.91	26.29
12	70.91	35.40	35.33	55.56
13	94.30	72.81	91.74	95.72
14	28.84	37.36	3.29	11.60

The resulting individual (Q_S) has sensors in the positions 15, 21, 22, 32, 43, and 50. Q_S uses only sensors already used by the control system; therefore, it involves low investment costs.

To analyze its performance, Q_S has been compared with a sensor network proposed in CG98. This individual is called Q_A and contains sensors in the positions 1, 2, 3, 4, 5, 6, 9, 10, 11, 13, 14, 16, 18, 19, 21, and 22. One PCA model has been built for each sensor-network and implemented in the TE simulator.

From the point of view of the proposed fitness evaluation, Q_A is unfeasible, (i.e. FSP → ∞), because it cannot detect IDV(3). Another comparison has been made using a reliability index in the different cases. The reliability has been defined as the percentage of sample statistics out of the control limits. Control limits have been set to 99% of confidence so that 1% of sample statistics are outside the control limit in normal operation. The reliability was calculated for all disturbances, for both statistics (SPE and T^2) and for both individuals (Q_A, Q_S). Table 5 and Figures 7 and 8 show how the sensor network Q_S can clearly detect IDV(3) which is not detected by Q_A. On the other hand, Q_A has a better performance (higher reliability) in disturbances IDV-(10), IDV(11), and IDV(12), but Q_S can also clearly detect them. It should be noted that neither Q_S nor Q_A detect IDV(4) and IDV(9). That is mainly due to the fact that the necessary sensors to detect them are more expensive than the FSP (at the simulated size).

Conclusions

A performance index for sensor networks has been introduced. It deals with two objectives, the sensor network costs and the detection capabilities of a PCA
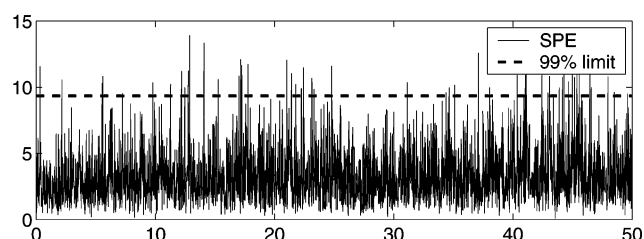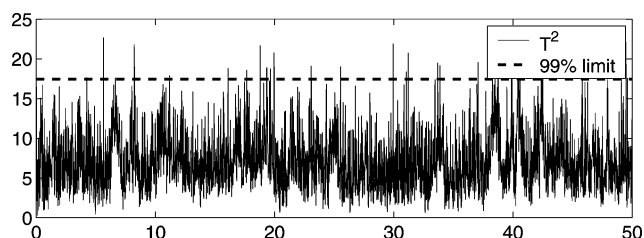
Figure 7. PCA detection results of IDV(3) (at time = 2 h) when Q_A is used.
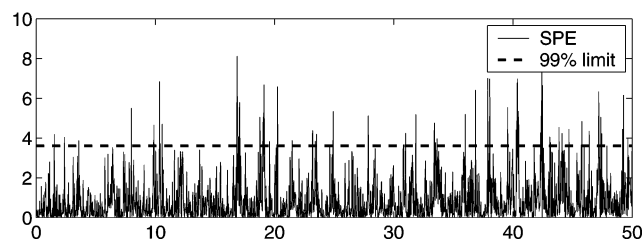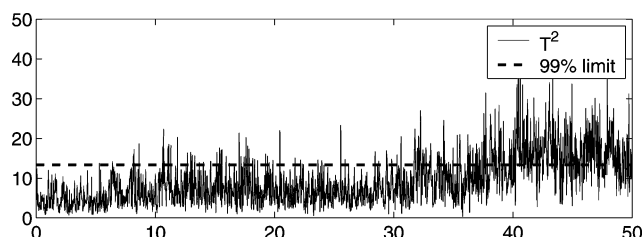
Figure 8. PCA detection results of IDV(3) (at time = 2 h) when Q_S is used.

monitoring system that use its signals. A genetic algorithm that maximizes the performance index has been presented. Finally, the complete methodology has been tested using two relevant case studies.

Although it may seem that there is a tradeoff between the two objectives, it is shown that it is possible to improve fault detection performance of PCA monitoring systems without increasing sensor network costs. Adding sensors in the network does not imply an improvement in fault detection; on the contrary, it could be in detriment of the detection performance.

Nomenclature

A = reactant
B = product
CFMs = critical fault magnitudes
C = sensor costs matrix
D_λ = eigenvalues diagonal matrix
F = fault
FSP = fault size penalization
FC_{sup} = superior fault cost
f_j = fault parameter vector
$f°_j$ = fault parameters vectors of compulsory detection
$f_{sup_{i,j}}$ = size at which F_j have to be detected when Q_i is used
$f_{inf_{i,j}}$ = maximum size of F_j considered as noise when Q_i is used
f_{SPE} = critical fault magnitude in the residual subspace
f_T = critical fault magnitude in the PC subspace
l = fault subspace dimension
J = number of considered faults
k = reaction rate
L_{ind} = individual length
N = number of potentially located sensors
m = number of samples in $\mathbf{X}$
MCFM = minimum critical fault magnitude
n_i = number of variables measured by the sensor network Q_i
n = number of measured process variables
N_{ind} = number of individuals in each generation
OF = objective function
$\mathbf{P}$ = principal component matrix
PDL = process deviations limits
P_c = crossover probability
P_m = mutation probability
Q = sensor network
$\mathbf{R}$ = data correlation matrix
r = reaction rate
SNC = total sensor cost
SPE = squared predicted error statistic
T^2 = Hotteling's statistic
V_R = reactor hold up
w = FSP by unit of fault magnitude
$\mathbf{X}$ = data matrix
$\bar{\mathbf{X}}$ = normalized data matrix
$\mathbf{x}'$ = normalized sample vector
X_{AB} = mol fraction of A in the stripper bottom stream
X_{A00} = mol fraction of A in the fresh feed
X_f = mol fraction of A in the reactor
$\mathbf{x}_F$ = filtered sample vector

Greek Characters

α = relative volatility
$\Delta x'$ = residual vector
δ^2_{SPE} = SPE confidence limit
$\delta^2_T\, T^2$ = confidence limit
Φ = fitness function
Θ_j = fault orthonormal base

Subscripts

i = sensor network
j = fault
k = sensor position

Acronyms

DCS = distributed control systems
DG = direct graph
FSP = fault size penalization
GA = genetic algorithms
MSPM = multivariate statistical process monitoring
PCA = principal component analysis
PCs = principal components
SDG = signed direct graph

Literature Cited

(1) Kourti, T.; Lee, J.; MacGregor, J. Experiences with industrial applications of proyection methods for multivariate statistical process control *Comput. Chem. Eng.* **1996**, *20*, S745–S750.

(2) Pearson, K. On lines and planes of closest fit to systems of point in space. *Philos. Mag.* **1901**, *6*, 559–572.

(3) Hotelling, H. Analysis of a complex of statistical variables into principal components. *Educ. Psychol.* **1933**, *24*, 417–441, 498–520.

(4) Nomikos, P.; MacGregor, J. Monitoring batch processes using multiway principal component analysis. *AIChE J.* **1994**, *40*, 1361–1375.

(5) Dunia, R.; Qin, S.; Edgar, T.; McAvoy, T. Identification of faulty sensors using principal component analysis. *AIChE J.* **1996**, *42*, 2797–2812.

(6) Vedam, H.; Venkatasubramanian, V. PCA-SDG based process monitoring and fault diagnosis. *Control Eng. Pract.* **1999**, *7*, 903–917.

(7) Kramer M. A. NonLinear PCA using Autoassociative NN. *AIChE J.* **1991**, *37*, 233–243.

(8) Dong, D.; McAvoy, J. Nonlinear principal component analysis-based on principal curves and neural networks. *Comput. Chem. Eng.* **1996**, *20*, 65–77.

(9) Raghuraj, R.; Bhushan, M.; Rengaswamy, R. Locating sensors in complex chemical plants based on fault diagnosis observability criteria. *AIChE J.* **1999**, *45*, 310–322.

(10) Bhushan, M.; Rengaswamy, R. Design of sensor network based on the signed directed graph of the process for efficient fault diagnosis. *Ind. Eng. Chem. Res.* **2000**, *39*, 999–1019.

(11) Bagajewicz, M. *Design and upgrade of process plant instrumentation*; CRC: 2000.

(12) Bagajewicz, M. A review of techniques for instrumentation design and upgrade in process plants. *Can. J. Chem. Eng.* **2002**, *80*, 3–16.

(13) Bagajewicz, M. Review of Recent Results in Instrumentation Design and Upgrade for Process Plants. In *Proceedings of the 4th IFAC Workshop on On-Line Fault Detection & Supervision in the Chemical Process Industries*, June 2001.

(14) Wang, H.; Song, Z.; Wang, H. Statistical process monitoring using improved PCA with optimized sensor locations. *J. Process Control* **2002**, *12*, 735–744.

(15) Himes, D. M.; Storer, R. H.; Georgakis, C. Determination of the number of principal components for disturbance detection and isolation. In *Proceedings of the Americal Control Conference*; IEEE Press: Piscataway, NJ, June 1994.

(16) Jackson, J. *A user's guide to principal components analysis*; John Wiley: New York, 1991.

(17) Jackson, J.; Mudholkar, G. Control procedures for residuals associated with principal components analysis. *Technometrics* **1979**, *21*, 341–349.

(18) Dunia, R.; Qin, S. Subspace approach to multidimensional fault identification and reconstruction. *AIChE J.* **1998**, *44*.

(19) Wang, H.; Song, Z.; Li, P. Fault detection behaviour and performance analysis of principal component analysis based process monitoring methods. *Ind. Chem. Res.* **2002**, *41*, 2455–2464.

(20) Karr, C. Fuzzy control of pH using genetic algorithms. *IEEE Trans. Fuzzy Syst.* **1993**, *1*, 146–153.

(21) Holland, J. *Adaptation in natural and artificial systems*; The University of Michigan Press: 1975.

(22) Goldberg, D. *Genetic algorithms in search, optimization, and machine learning*; Addison-Wesley: 1989.

(23) Belanger, P. W.; Luyben W. L. Inventory control in processes with recycle. *Ind. Eng. Chem. Res.* **1997**, *1*, 706−716.

(24) Chiang, L.; Pell, R.; Seasholtz, M. Exploring process data with the use of robust outlier detection algorithms. *J. Process Control* **2003**, *13*, in press.

(25) Downs, J.; Vogel, E. A plant-wide industrial process control problem. *Comput. Chem. Eng.* **1993**, *17*, 245−255.

(26) McAvoy, T.; Ye, N. Base control problem for the Tennessee Eastman problem. *Comput. Chem. Eng.* **1994**, *18*, 383−413.

(27) Singhal, A. Tennessee Eastman plant simulation with base control system of McAvoy and Ye. Available at http://chemen-gr.ucsb.edu/ceweb/computing/TE/tesim.pdf, accesed in Dec 2003.

(28) Cheng, G.; McAvoy, T. Predictive on-line monitoring of continuous processes. *J. Process Control* **1998**, *8*, 409−420.