A Systematic Approach for the Design of Optimal Monitoring Systems for Large Scale Processes

David Zumoffen and Marta Basualdo*

Grupo de Informática Aplicada a Ingeniería de Procesos (GIAIP), Centro Franco-Argentino de Ciencias de la Información y de Sistemas (CIFASIS-CONICET-UNR), 27 de Febrero 210 bis, S2000EZP Rosario, Argentina, and Universidad Tecnológica Nacional. Facultad Regional Rosario, Argentina

In this work a new concept for designing an efficient monitoring system for large scale chemical plants is presented. It is considered that the monitoring problem must be solved integrated with the optimal sensor location together with the plant-wide control structure design. The solution of these problems involves deciding among a great number of possible combinations between the input−output variables. It is done supported by the application of genetic algorithm (GA). The key new idea is to propose an adequate objective function, within the GA, that takes into account a fault detectability index based on combined statistics. Additionally, by using a specific penalty function, it is possible to drive the search to the less expensive structure, that is by using the lowest number of sensors. The well-known benchmark case of the Tennessee Eastman plant (TE) is chosen for testing this methodology and for discussion purposes. Since several authors have studied the TE case, the results obtained here can be rigorously compared with those already published. All of the previous works considered that every TE output variables were available for the abnormal events detection for designing the monitoring system.

1. Introduction

The monitoring systems for large scale processes play an important role in supervising the state of a plant. Hundreds and thousands of variables must be handled each sampling time. The increasing demand for obtaining more efficient and profitable process generates complex control structures and highly interconnected plants. In this context, to achieve a suitable monitoring system it is necessary to account with an efficient technique for decision making.[1−3] The methodology proposed in this work is integrated to a global, generalized, and systematic analysis for determining an optimal sensor location and plant-wide control structure selection, given in a previous work of the authors.[13] In that work only a steady-state model of the process was accounted without using heuristic concepts for reducing the problem dimensionality. In this work an optimal monitoring system, based on historical normal and abnormal behavior database of the process, is proposed. In Figure 1 the block diagram with light gray background displays the procedure given in Zumoffen and Basualdo.[13] Meanwhile, the dark gray background represents the systematic approach detailed here. Therefore, Figure 1 summarizes the integration between both techniques for analyzing, which is the real benefit of using them. The PCA-based monitoring systems are widely used in industrial processes as well as academic research[4−6] due to their excellent properties for handling noise and large databases. One of the problem analyzed here is, which is the minimum number of signals to be chosen for developing an efficient PCA monitoring system? Usually, this question is not addressed because it is assumed that the overall plant information is available for developing the PCA model. It is remarkable that some of the previous works in the literature addressed this complex problem in a partial way. For example, Yue and Qin[7] presented the combined statistic using PCA for fault reconstruction tasks on rapid thermal annealing (RTA) process. Only fault detectability with square prediction error, *SPE*, or Hotelling, T^2, statistics were analyzed. Nothing was said about the use of combined statistics for fault detectability calculations and the optimal signal selection for PCA implementation. Additionally, on this last area, several works have been published[8−10] based on Kalman filtering in steady-state, integer optimization routines and observability index. It can be remarked that, these approaches were applied in two ways, assuming that the process was working in open loop or with an existing plant-wide control structure in the process. It can be mentioned that in Musulin et al.[11,12] was presented a signal selection method with PCA-based monitoring using only SPE and T^2 statistics. The approach was applied on the Tennessee Eastman (TE) process where the signals selection was done as a trade off between cost and precision aspects. That strategy was tested in two case studies. First, a chemical plant with recycle is used accounting faults and disturbances, and second, the TE process is performed by analyzing some of the suggested disturbances without taking

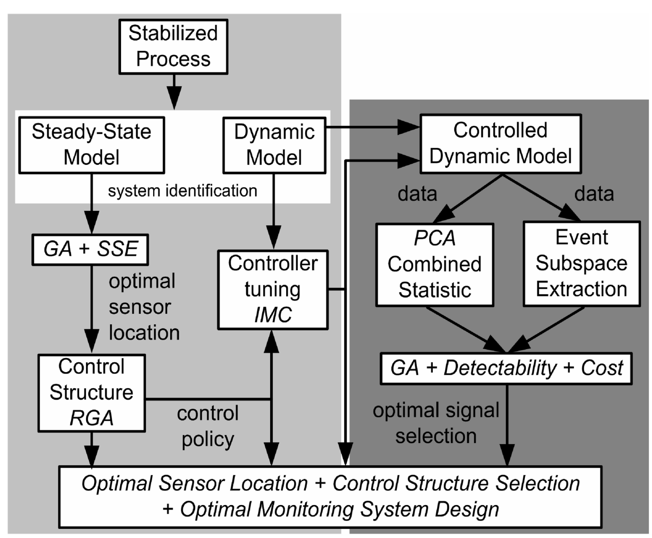


Figure 1. Proposed strategy.

* To whom correspondence should be addressed. Tel.: +54-341-4237248 int 304. Fax: +54-341-482-1772. E-mail: basualdo@cifasis-conicet.gov.ar.

10.1021/ie9017836
Published on Web 01/15/2010

into account the potential faults at different hardware elements. In this work a new approach to improve the PCA-based monitoring systems performance is presented. Its optimal design is thought to maximize the abnormal events detectability and to minimize the investment costs. The detectability indexes are useful for determining which set of variables could provide the best fault classification. Particularly, one of the contribution of this work is to calculate this index by using a combined statistic. Because of the great dimensionality problem associated with the large chemical plants, the best variables selection is performed with the help of genetic algorithms (GA). In addition, this approach can be driven to use those measurements provided by the optimal sensor location analysis previously done. Thus, the objective function, within the GA, also accounts a cost that penalizes the use of new measurement devices which were not considered to be part of the obtained control structure. So, it provides a more profitable selection.

Finally, on the basis of the arguments given above, it is important to remark that the proposed approach here has several differences with respect to analogous previous strategies:[11,12] First, the global treatment of the complete problem, which has never been analyzed in this way, supported by a well-defined objective function to be optimized through genetic algorithm and, second, the monitoring design based on the detectability index calculated with the combined statistics. It is proposed here based on the successful results shown in previous works.[1−3] On the other hand, the overall monitoring system is developed with adaptive PCA (APCA) techniques, which present better behavior than conventional ones in plants with frequent changes on the operating conditions. The APCA, classification, and estimation modules were discussed in detail in Zumoffen and Basualdo[1,2] and Zumoffen et al.[3]

2. Background Information and Tools

2.1. PCA-Based Monitoring. The principal component analysis (PCA) is the main tool used here for performing fault detection. PCA is a projection-based method that facilitates a reduction of data dimension. This analysis begins by considering the data matrix $\mathbf{X}$ of $m \times n$ containing m samples of n process variables collected under normal operation. Assuming that $\bar{\mathbf{X}}$ is the normalized version of $\mathbf{X}$, to zero mean and unit variance with the scale parameter vectors $\mathbf{b}$ and $\mathbf{s}$ as the mean and variance vector of the process variables in the data matrix, respectively. The normalized data matrix can be represented as

$$\bar{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{1}$$

where $\mathbf{T} \in \mathscr{R}^{m\times A}$ and $\mathbf{P} \in \mathscr{R}^{n\times A}$ are the latent and principal components matrices, respectively, and A is the number of principal components retained in the model $\mathbf{P}$. The residual matrix $\mathbf{E}$ represents the associated error since only $A \ll n$ principal components were selected.

$\mathbf{P}$ can be obtained by means of singular value decomposition (SVD) from the normalized data correlation matrix $\mathbf{R}_c = \bar{\mathbf{X}}\cdot\bar{\mathbf{X}}^{\mathrm{T}}/(m-1) = \mathbf{U}\mathbf{D}_\lambda\mathbf{U}^{\mathrm{T}}$, by selecting only the first A columns of $\mathbf{U}$. This factorization produces a diagonal matrix $\mathbf{D}_\lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, where λ_i are the eigenvalues of $\mathbf{R}_c$ sorted in decreasing order ($\lambda_1 > \lambda_2 > \cdots > \lambda_n$) and the corresponding columns of $\mathbf{U}$ are the eigenvectors $\mathbf{p}_i$ and so-called the principal components. Thus, $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_A]$ and $\mathbf{D}_A = \mathrm{diag}(\lambda_1, \cdots, \lambda_A)$. A reduction of dimensionality is made by projecting every normalized sample vector $\bar{\mathbf{x}}(k)$ (of dimension $n \times 1$) in the principal component space generated by $\mathbf{P}$, obtaining $\mathbf{t}(k) = \mathbf{P}^{\mathrm{T}}\bar{\mathbf{x}}(k)$, which is called the principal score vector.

Different approaches for selecting the A principal components retained[14] can be chosen. In this work the cumulative percent variance (CPV) is used, eq 2. This index measures the percent variance captured by the first A principal components.

$$\mathrm{CPV}(A) = \frac{\sum_{j=1}^{A} \lambda_j}{\mathrm{trace}\ (\mathrm{R_c})} 100 \tag{2}$$

in this case a search between $1 \leq A \leq n$ is made in order to be satisfied the condition $\mathrm{CPV}(A) \geq \delta_{\mathrm{cpv}}$ with the minimum A. Where δ_{cpv} is a percentage value; if it achieves lower values it means that only a few principal components retained are needed, and if it is close to 100%, it means that $A \approx n$.

For generating quality control charts in a multivariable online monitoring process with PCA, two statistics are widely used, the Hotelling's T^2 and the squared prediction error (SPE) Q. Considering the actual process measurements at their normalized version $\bar{\mathbf{x}}(k)$, k being the actual sampling time, these statistic are defined as

$$T^2(k) = ||\mathbf{D}_A^{-1/2}\mathbf{P}^{\mathrm{T}}\bar{\mathbf{x}}(k)||^2, Q(k) = ||\tilde{\mathbf{C}}\bar{\mathbf{x}}(k)||^2 \tag{3}$$

where $\tilde{\mathbf{C}} = \mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}}$ and $\Delta\bar{\mathbf{x}}(k) = \tilde{\mathbf{C}}\ \bar{\mathbf{x}}(k)$ is the prediction error. The test declares normal operation if $T^2(k) \leq \delta_{T^2}$ for Hotelling's statistic and $Q(k) \leq \delta_Q$ for SPE statistic, where δ_{T^2} and δ_Q are the control or confidence limits for the above statistics, respectively. Supposing Gaussian distribution the control limits can be approximated by $\delta_i = \mu_i + \nu\cdot\sigma_i$,[15,16] where μ_i and σ_i are the mean and variance values for the statistic i computed from the normal data matrix ($i = T^2, Q$) and $\nu = 2,3$ according to the 95% or 99% confidence level, respectively.

$\mathbf{P}$ is computed using the normal data matrix which has information about the *common-cause* variations at the surroundings of the process operation point. The Hotelling's statistic T^2 for a new incoming data sample is a measurement of its distance with respect to the origin of the dominant variation subspace. This origin and its proximities delimit the *in-control zone*. The case where the principal score vectors remain at the surroundings of the in-control zone (Hotelling's statistic under the confidence limit), even though an abnormal event has happened, suggests that it cannot change enough over the dominant variation subspace. To avoid this lose of detection of abnormal events information, the use of the Q and T^2 statistics, working together, in a combined way is proposed.[7] Hence, a detectability analysis is done based on them, as shown in section 3, which represents a new result from this work. This approach improves the detection properties such as detection times, less false alarms occurrence, and missed detections. The combined index $z(k)$ is defined as

$$z(k) = \frac{T^2(k)}{\delta_{T^2}} + \frac{Q(k)}{\delta_{\mathrm{SPE}}} \tag{4}$$

Note that $z(k)$, $T^2(k)$, and $Q(k)$ are the computed statistics for the actual measurement $\bar{\mathbf{x}}(k)$. Under these assumptions normal operation conditions can be declared if $z(k) \leq \delta_z$, where δ_z is the new control limit. A conservative selection could be $\delta_z = 2$, according to the false alarms occurring when the FD system is injurious. A modified version of this approach integrated with adaptive PCA techniques, fault detection, diagnosis, and estimation system which supports fault tolerant control can be found in Zumoffen and Basualdo[1,2] and Zumoffen et al.[3]

2.2. Fault Detectability Index Based on T^2 and Q. Fault detectability index can be suitably obtained using an additive fault model[7] representation of the abnormal process data. This methodology can be applied to both Hotelling and the square prediction error statistics. Yue and Qin[7] presented an approach to develop this index in the T^2 and Q cases but nothing was said about the combined statistic fault detectability index. In the next sections an extension for computing it is proposed.

The normalized process measurement $\bar{\mathbf{x}}_*(k)$ when a fault is present can be written as

$$\bar{\mathbf{x}}_*(k) = \bar{\mathbf{x}}_0(k) + \Theta_j \mathbf{f}_j \tag{5}$$

Two additive effects can be observed in this fault model: $\bar{\mathbf{x}}_0(k)$ that considers the normal behavior case and $\Theta_j\mathbf{f}_j$ as the fault contribution to the actual measurements. Θ_j is the fault subspace for the fault j, where $j = 1, \cdots, J$ are the fault types and $\mathbf{f}_j$ is the fault components vector. The fault subspace extraction procedure is addressed properly in the appendix.

$$\begin{bmatrix} \bar{x}_*^1(k) \\ \bar{x}_*^2(k) \\ \vdots \\ \bar{x}_*^m(k) \end{bmatrix} = \begin{bmatrix} \bar{x}_0^1(k) \\ \bar{x}_0^2(k) \\ \vdots \\ \bar{x}_0^m(k) \end{bmatrix} + \begin{bmatrix} \theta_j^{11} & \theta_j^{12} & \cdots & \theta_j^{1r} \\ \theta_j^{21} & \theta_j^{22} & \cdots & \theta_j^{2r} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_j^{m1} & \theta_j^{m2} & \cdots & \theta_j^{mr} \end{bmatrix} \begin{bmatrix} f_j^1(k) \\ f_j^2(k) \\ \vdots \\ f_j^r(k) \end{bmatrix} \tag{6}$$

where m is the amount of measurement points, and r is the fault components vector length for the fault type j. The columns of Θ have zero entries except for the measurement affected by the fault; in this case the entry is 1 or -1 depending of the fault direction.

Considering the fault detection process with Hotelling statistic results

$$T^2(k) = \|\mathbf{D}_A^{-1/2}\mathbf{P}^{\mathrm{T}}\bar{\mathbf{x}}_*(k)\|^2 = \|\mathbf{D}_A^{-1/2}\mathbf{P}^T\bar{\mathbf{x}}_0(k) + \mathbf{D}_A^{-1/2}\mathbf{P}^T\Theta_j\mathbf{f}_j\|^2 \tag{7}$$

and accounting the triangle inequality it can be expressed as

$$\|\mathbf{D}_A^{-1/2}\mathbf{P}^T\bar{\mathbf{x}}_*(k)\| \geq \big|\|\mathbf{D}_A^{-1/2}\mathbf{P}^T\bar{\mathbf{x}}_0(k)\| - \|\mathbf{D}_A^{-1/2}\mathbf{P}^{\mathrm{T}}\Theta_j\mathbf{f}_j\|\big| \tag{8}$$

remembering that for fault detection the condition $T^2(k) \geq \delta_{T^2}$ must be fulfilled and that $\|\mathbf{D}_A^{-1/2}\mathbf{P}^{\mathrm{T}}\bar{\mathbf{x}}_0(k)\| \leq \delta_{T^2}$ by construction. The detectability condition using Hotelling statistic can be written as

$$\|\mathbf{D}_A^{-1/2}\mathbf{P}^{\mathrm{T}}\Theta_j\mathbf{f}_j\| \geq 2\delta_{T^2} \tag{9}$$

and the minimal fault magnitude (MFM) detectable with this statistic can be obtained with a most restrictive condition:

$$\|\mathbf{D}_A^{-1/2}\mathbf{P}^{\mathrm{T}}\Theta_j\|\|\mathbf{f}_j\| \geq \|\mathbf{D}_A^{-1/2}\mathbf{P}^{\mathrm{T}}\Theta_j\mathbf{f}_j\| \geq 2\delta_{T^2} \tag{10}$$

resulting in the following expression for the MFM estimate

$$\|\mathbf{f}_j\|_{\mathrm{MFM}}^{T^2} = \|\mathbf{D}_A^{-1/2}\mathbf{P}^{\mathrm{T}}\Theta_j\|^{-1}2\delta_{T^2} \tag{11}$$

Considering the fault detection with Q statistic:

$$Q(k) = \|\tilde{\mathbf{C}}\bar{\mathbf{x}}_*(k)\|^2 = \|\tilde{\mathbf{C}}\bar{\mathbf{x}}_0(k) + \tilde{\mathbf{C}}\Theta_j\mathbf{f}_j\|^2 \tag{12}$$

and again using the triangle inequality, it can be expressed as

$$\|\tilde{\mathbf{C}}\bar{\mathbf{x}}_*(k)\| \geq \big|\|\tilde{\mathbf{C}}\bar{\mathbf{x}}_0(k)\| - \|\tilde{\mathbf{C}}\Theta_j\mathbf{f}_j\|\big| \tag{13}$$

where $\|\mathbf{C}\bar{\mathbf{x}}_0(k)\| \leq \delta_Q$ by construction. Thus the fault detectability condition with this statistic is summarized as

$$\|\tilde{\mathbf{C}}\Theta_j\mathbf{f}_j\| \geq 2\delta_Q \tag{14}$$

and, in this case, the MFM estimation results

$$\|\mathbf{f}_j\|_{\mathrm{MFM}}^{Q} = \|\tilde{\mathbf{C}}\Theta_j\|^{-1}2\delta_Q \tag{15}$$

2.3. Genetic Algorithms. The genetic algorithms is a stochastic global search method that mimics the metaphor of natural biological evolution. Mainly, the GA operates on a population of potential solutions, N_i, applying the principle of survival of the fittest to produce better and better approximation to the solution. In each generation a new set of individuals, $\mathbf{I}_i$, is created on the basis of their level of fitness with respect to the specific functional cost for each problem, $FC(\mathbf{I}_i)$, and breeding them together using operators borrowed from natural genetics. Thus, the evolution of the population is developed toward the individuals that are better suited to their environment. Specifically, in this work a code version developed in ref 17 for Matlab environment is used.

The individuals, $\mathbf{I}_i = [g_1, \cdots, g_{N_c}]$, are encoded as strings (chromosomes) composed over some alphabet, so that the chromosome values, g_j, are uniquely mapped onto the decision variables domain. Decoding the chromosome representation, the decision variables (individuals) can be evaluated considering some performance or fitness function. This function establishes the basis for pairs selection of individuals that will be mated together during reproduction. In this phase, each individual is assigned to a fitness value derived from this objective function and the selection over the population is made with a determined probability according to their relative fitness. Thus the recombination process is carried out to produce the next generation.

Typically, the initial population is adjusted between 2−20% of the problem dimension for a small combinatorial size[17] and about $N_c \times 100$ to large scale problems.[5,11] Generally, the GA termination criteria is the maximum generation number, N_g, or a level of fitness. The individuals selection is performed by means of the roulette wheel method according to their fitness measure, thus the best $N_{sel} = N_i/2$ individuals are retained. The production of new chromosomes is developed by the crossover operator, in this case double-point method with a probability of P_{co}. Analogously to the natural evolution, the mutation produces a new genetic structure and basically is applied with a low probability, P_m. Mutation generally tends to inhibit the possibility of converging to a local optimum.

The following steps summarize the GA procedure to find the optimal set of individuals along generations:

A. Initialization. The random initial population, $\mathbf{P}_j = [\mathbf{I}_1, \cdots, \mathbf{I}_{N_i}]^{\mathrm{T}}$ of dimension $N_i \times N_c$, is defined with $j = 0$. In addition, N_i, N_g, N_{sel}, P_{co}, and P_m are selected.

B. Fitness Evaluation. The functional cost is evaluated for each individual, $\mathbf{I}_i$, from the actual population set, $\mathbf{P}_j$. In addition, the individual with best fitness value is stored in the best population set, $\mathbf{P}_b$.

C. While the termination criteria is false, do

C.1. Selection: the N_{sel} best individuals are selected using their relative fitness values and stored in $\mathbf{P}_s$. The $N_i - N_{sel}$ remaining individuals are discarded.

C.2. Recombination: the individuals in $\mathbf{P}_s$ are recombining by the crossover operator and stored in $\mathbf{P}_r$.

C.3. Mutation: the recombined individuals in $\mathbf{P}_r$ suffer the mutation process and new genetic structures are obtained and stored in $\mathbf{P}_m$.

C.4. Merging: both selected and mutated populations are merging together to give the next generation of individuals, $\mathbf{P}_j = [\mathbf{P}_s, \mathbf{P}_m]^{\mathrm{T}}$, with $j = j + 1$ and go to step **B**.

D. Else. If the termination criteria is true, the set of the best individuals for each generation is obtained from $\mathbf{P}_b$, with a dimension of $j \times N_c$, and the optimization procedure stops.

Thus, the better suited individual to their environment is the best individual in the latest generation.

3. Optimal Monitoring System Design

In this section the main steps of the proposed methodology are detailed on the corresponding subsections based on the theoretical background given before.

3.1. Fault Detectability Index Based on Combined Statistic. In this subsection the development of the fault detectability index with the combined statistic ($z(k)$) is given. It is done by performing an extension of the concepts given in the previous section. By grouping both Hotelling and square prediction error statistics can be obtained with improvements in the fault detection performance. In this case the combined statistic can be represented as

$$\begin{aligned} z(k) &= \frac{\bar{\mathbf{x}}_*^{\mathrm{T}}(k)\mathbf{P}\mathbf{D}_A^{-1}\mathbf{P}^{\mathrm{T}}\bar{\mathbf{x}}_*(k)}{\delta_{T^2}} + \frac{\bar{\mathbf{x}}_*^{\mathrm{T}}(k)\tilde{\mathbf{C}}^{\mathrm{T}}\tilde{\mathbf{C}}\bar{\mathbf{x}}_*(k)}{\delta_Q} \\ &= \bar{\mathbf{x}}_*^{\mathrm{T}}(k)\left[\frac{\mathbf{P}\mathbf{D}_A^{-1}\mathbf{P}^{\mathrm{T}}}{\delta_{T^2}} + \frac{\tilde{\mathbf{C}}^{\mathrm{T}}\tilde{\mathbf{C}}}{\delta_Q}\right]\bar{\mathbf{x}}_*(k) \\ &= \bar{\mathbf{x}}_*^{\mathrm{T}}(k)\mathbf{M}\bar{\mathbf{x}}_*(k) \end{aligned} \tag{16}$$

considering that $\mathbf{M}$ is a symmetric and definite positive matrix, its decomposition is possible by Cholesky factorization to obtain $\mathbf{M} = \mathbf{R}^{\mathrm{T}}\mathbf{R}$. Thus the combined statistic case presents a similar treatment as the Q statistic, giving the following fault detectability condition

$$\|\mathbf{R}\Theta_j\mathbf{f}_j\| \geq 2\delta_z \tag{17}$$

and its corresponding MFM estimate

$$\|\mathbf{f}_j\|_{\mathrm{MFM}}^z = \|\mathbf{R}\Theta_j\|^{-1}2\delta_z \tag{18}$$

The combined matrix $\mathbf{M}$ (and eventually $\mathbf{R}$) depends on the PCA model developed opportunely. The influences of factors such as sensors location in the process, signals selection to perform PCA model, variance retained, and confidence limits are crucial and limit the attainable MFM. Similarly, the fault subspace matrix Θ_j is directly influenced by the sensors network (potential sources of faults) and the selected control structure. In other words, a particular design of these matrices can incur losses or poor quality fault detections. An interesting result can be obtained through the comparison between a PCA model developed using the overall available measurement points and the optimal solution given here.

3.2. Optimal Signal Selection Based on Detectability Maximization. The proper signals selection for the PCA model development is performed by focusing on faults detectability maximization. It is based on the existing sensors network in the previously proposed optimal control structure[13] and the potential additional cost in case new measurement points would be required.

In this context, the problem can be defined as follows, considering N_c available signals including controlled as well as manipulated variables in the process and $\mathbf{I}_i = [g_1, g_2, \cdots, g_{N_c}]$ a particular signals selection, where $g_l = \{1;0\}$ with $l = 1, \cdots, N_c$ represents a binary alphabet indicating the utilization or not of the signal in the l location. Then, the PCA model construction depends on this particular selection, $\mathbf{P}(\mathbf{I}_i)$ and $\mathbf{D}_A(\mathbf{I}_i)$. In addition, the MFM calculation when combined statistics are used results

$$\|\mathbf{f}_j^i\|_{\mathrm{MFM}}^z = \|\mathbf{R}(\mathbf{I}_i)\Theta_j(\mathbf{I}_i)\|^{-1}2\delta_z \tag{19}$$

where i makes reference to the signals selection $\mathbf{I}_i$, with $i = 1, \cdots, 2^{N_c}$ and where $j = 1, \cdots, J$ is the type of considered abnormal events (disturbances, faulty elements, etc.).

Another important point to consider is a cost penalization $\mathbf{c} = [c_1, \cdots, c_{N_c}]$, greater than zero each time that new measurements are introduced. Note that it is not necessary that $\mathbf{c}$ must be in monetary units, in fact, its value can be selected by considering a trade-off between detectability and quantity of new sensors. Thus, lower values of $\mathbf{c}$ means that the minimization of the objective function in eq 20 prioritizes detectability without considering the cost of new sensors. On the contrary, when $\mathbf{c}$ has a considerable weight the cost of new sensors is penalized without considering the detectability index. The penalization coefficients must be normalized to contribute in the same order of magnitude as the term $\|\mathbf{R}(\mathbf{I}_i)\Theta_j(\mathbf{I}_i)\|^{-1}2\delta_z$ to solve the trade-off problem. Therefore, the complete problem to be solved can be stated as

$$\min_{\mathbf{I}_i}\left[\sum_{j=1}^{J}\|\mathbf{f}_j^i\|_{\mathrm{MFM}}^z + \mathbf{c}\mathbf{I}_i^{\mathrm{T}}\right] = \min_{\mathbf{I}_i}\left[\sum_{j=1}^{J}\|\mathbf{R}(\mathbf{I}_i)\Theta_j(\mathbf{I}_i)\|^{-1}2\delta_z + \mathbf{c}\mathbf{I}_i^{\mathrm{T}}\right] \tag{20}$$

This is a combinatorial problem with 2^{N_c} potential solutions. In eq 20 the minimization of the MFM with combined statistics drives to the maximization of the faults detectability. In other words, the search is oriented toward to find the optimal signals selection $\mathbf{I}_{\mathrm{op}}$ (solution of eq 20) that guarantees the best fault detection of the most expected abnormal events at a lowest investment cost.

3.3. Genetic Algorithm Solution. In this subsection, the use of GA is discussed for solving the problem displayed in eq 20 subject to the following restrictions

$$\begin{gathered} \sum_{l=1}^{N_c}\mathbf{I}_i(l) > A \\ \|\mathbf{R}(\mathbf{I}_i)\Theta_j(\mathbf{I}_i)\| > 0 \end{gathered} \tag{21}$$

where the first restriction in eq 21 avoids the selection $\mathbf{I}_i$ that does not present dimensional reduction. In addition, if the signal dimension is equal to the principal components retained A the combined statistic z is reduced to Hotelling statistic only ($Q = 0$). The second constraint avoids the individual selection $\mathbf{I}_i$ that produces $\|\mathbf{f}_j\|_{\mathrm{MFM}}^z = \infty$, which means that a specific fault can not be detected. In other words, the second constraint guarantees the fault detectability condition in eq 17.

4. Case Study: Tennessee Eastman Process

Essentially, the TE plant[18] generates two products from four reactants and has five principal units of operation: the reactor, the product condenser, a vapor−liquid separator, a compressor, and a product stripper. The stripper underflow contains the key components G and H. In this work, is assumed that the plant is operating under conditions considered as *base case*, that is G/H mass ratio of 50/50 and a production rate of 7038 kgG/h and 7038 kgH/h. The process has 41 potential controlled variables (CV) and 12 manipulated variables (MV). The plant dynamic model used here was taken from Ricker's control department of the Washington University.[19−21]

In Zumoffen and Basualdo[13] a systematic approach for solving simultaneously optimal sensor location and plant-wide control based on genetic algorithms accounting only steady-state process information is presented. This generalized procedure is schematically shown at Figure 1 with light gray background. Basically, the optimal sensor location problem accounting the control structure is solved by using tools as steady-state models, internal model control (IMC), sum of square errors (SSE), relative gain array (RGA), and genetic algorithms (GA). Thus, for a nonsquare process model $\mathbf{G}(s) = [\mathbf{G}(s)_s^T, \mathbf{G}(s)_r^T]^T$ with m outputs and n inputs, $m > n$ being the problem to solve,

$$\min_{\mathbf{I}_i}[\mathrm{SSE}(\mathbf{I}_i)], \quad \text{subject to} \quad \|\mathbf{I}_i\| = n \tag{22}$$

where

$$\mathrm{SSE}(\mathbf{I}_i) = \mathrm{tr}(\mathbf{S}(\mathbf{I}_i)^T\mathbf{S}(\mathbf{I}_i)), \quad \mathbf{S}(\mathbf{I}_i) = \left[\begin{bmatrix} \mathbf{I}_{n\times n} \\ \mathbf{0}_{(m-n)\times n} \end{bmatrix} - \mathbf{G}\mathbf{G}_s(\mathbf{I}_i)^{-1}\right] \tag{23}$$

where tr(·) is the trace function, $s = 0$, and $\mathbf{I}_i$ is the GA chromosome that selects the sensor network. Minimization at eq 22 generates the sensors locations with the best decoupled RGA without heuristic considerations.

The control structure obtained by this methodology is displayed at Figure 2 which involves a number of variables described in the Table 1 (indicated with boldface type). This table displays the 52 potential measurement points and signals available in the TE process. A comparison with other control structures was presented in Zumoffen and Basualdo[13] considering both the dynamic behavior (set point changes and disturbances) and hardware requirements.

The optimal monitoring system design is based on analyzing the abnormal event detectability on the proposed control structure and the optimal signal selection mentioned above. In this framework, typical problems in process elements[1−3,22] are proposed and summarized at Table 2. Fourteen potential sensor faults, all of them consisting on shifts, are accounted within a specific range of magnitude. The disturbances analyzed here are the most critical events for the TE process, as it has been defined in the work of Downs and Vogel,[18] according to their deep knowledge of this process.

5. Results

In this section the chosen parameters for the implementation on the TE case study are included. According to the data given at Table 1 the combinatorial problem has dimension $N_c = 52$,

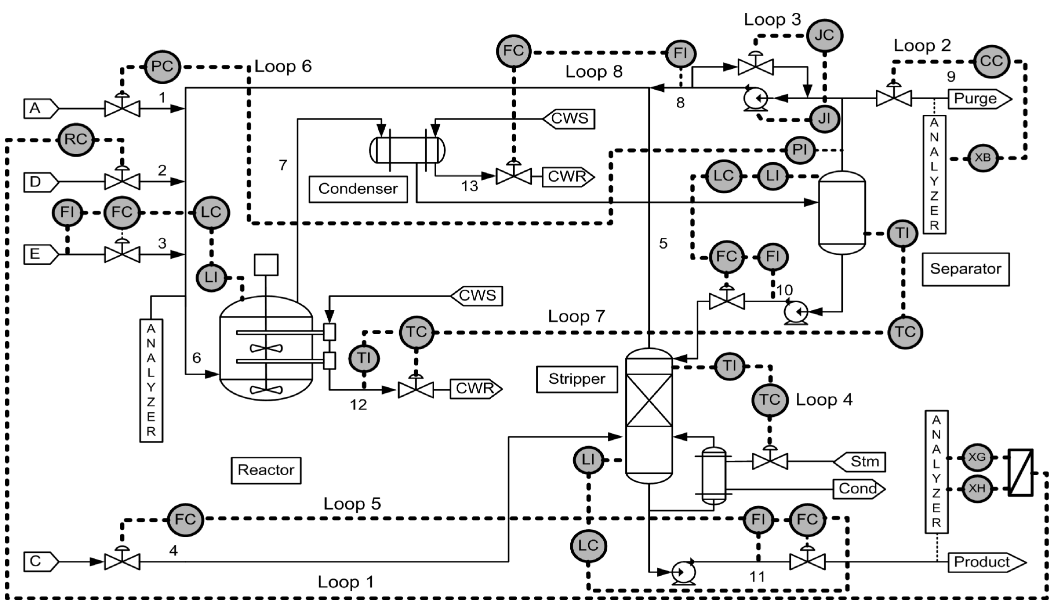


Figure 2. The proposed control structure.

Table 1. Variables for TE

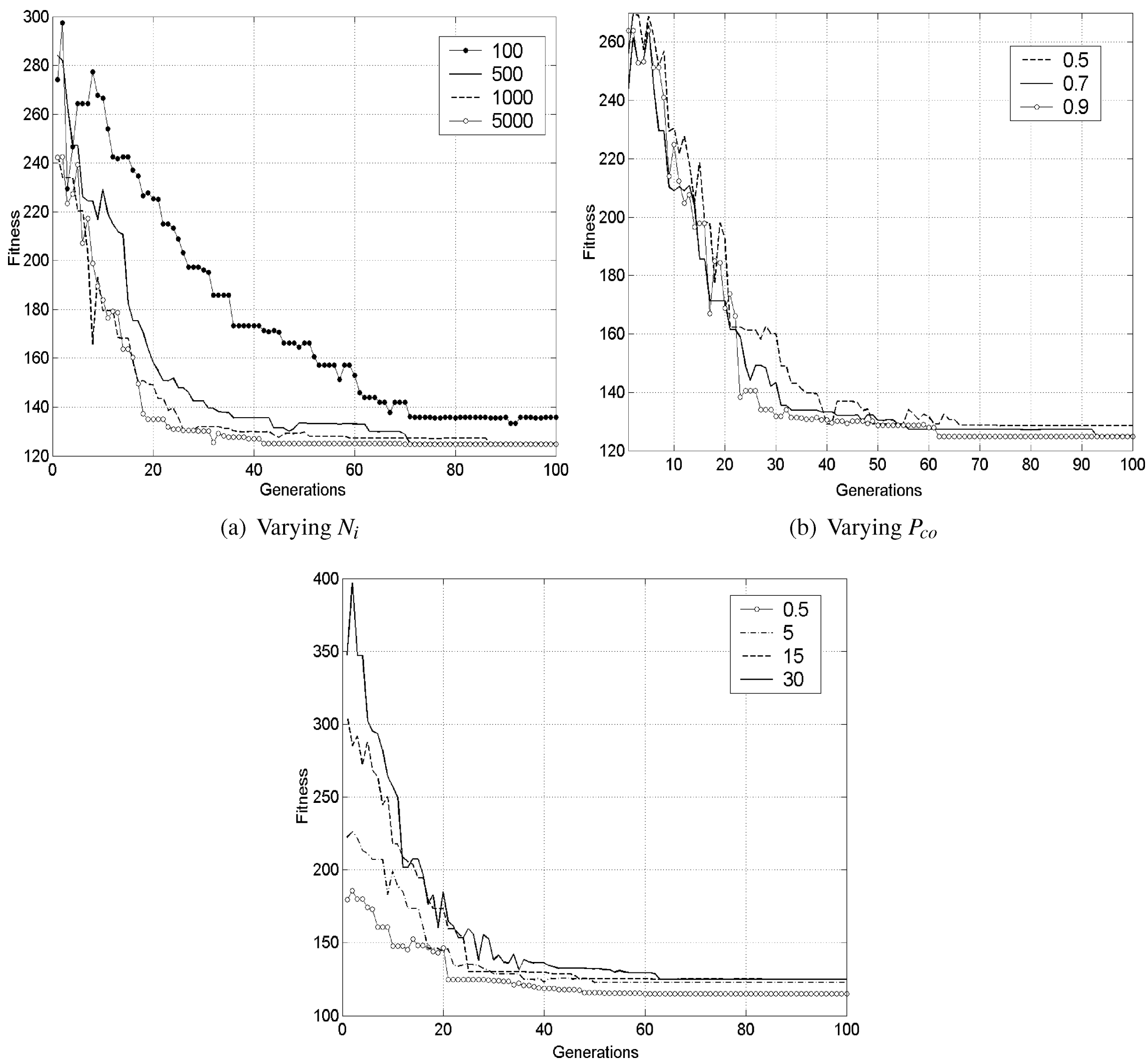
potential measurements				manipulated variables	
variable	description	variable	description	variable	description
$\boldsymbol{y}_{1}$	E feed flow	y_{20}	A comp. feed	$\boldsymbol{u}_{1}$	D feed valve
$\boldsymbol{y}_{2}$	recycle flow	y_{21}	B comp. feed	$\boldsymbol{u}_{2}$	E feed valve
y_{3}	reactor feed	y_{22}	C comp. feed	$\boldsymbol{u}_{3}$	A feed valve
y_{4}	reactor pressure	y_{23}	D comp. feed	$\boldsymbol{u}_{4}$	AC feed valve
$\boldsymbol{y}_{5}$	reactor level	y_{24}	E comp. feed	$\boldsymbol{u}_{5}$	comp. recycle valve
y_{6}	reactor temp	y_{25}	F comp. feed	$\boldsymbol{u}_{6}$	purge valve
y_{7}	purge rate	y_{26}	A comp. purge	$\boldsymbol{u}_{7}$	sep. liq. valve
$\boldsymbol{y}_{8}$	separator temp	$\boldsymbol{y}_{27}$	B comp. purge	$\boldsymbol{u}_{8}$	str. prod. valve
$\boldsymbol{y}_{9}$	separator level	y_{28}	C comp. purge	$\boldsymbol{u}_{9}$	str. steam valve
$\boldsymbol{y}_{10}$	separator pressure	y_{29}	D comp. purge	$\boldsymbol{u}_{10}$	RCW valve
$\boldsymbol{y}_{11}$	separator underflow	y_{30}	E comp. purge	$\boldsymbol{u}_{11}$	CCW valve
$\boldsymbol{y}_{12}$	stripper level	y_{31}	F comp. purge	$\boldsymbol{u}_{12}$	E feed set point
y_{13}	stripper pressure	y_{32}	G comp. purge	$\boldsymbol{u}_{13}$	Psuf set point
$\boldsymbol{y}_{14}$	production	y_{33}	H comp. purge	$\boldsymbol{u}_{14}$	suf set point
$\boldsymbol{y}_{15}$	stripper temp	y_{34}	D comp. product	$\boldsymbol{u}_{15}$	RCW flow set point
y_{16}	stripper steam flow	y_{35}	E comp. product		
$\boldsymbol{y}_{17}$	compressor work	y_{36}	F comp. product		
$\boldsymbol{y}_{18}$	reactor CWO temp	$\boldsymbol{y}_{37}$	G/H product		
y_{19}	separator CWO temp				

Table 2. Abnormal Events Proposed

faults	variable	range [%]	type	disturbances	description	type
F_1	E feed flow	±15	step	IDV1	A/C feed ratio	step
F_2	recycle flow	±10	step		B comp. constant	step
F_3	reactor level	±10	step	IDV4	RCW inlet temp	step
F_4	separator temp	±5	step	IDV8	A,B,C feed comp.	random
F_5	separator level	±10	step	IDV12	CCW inlet temp	random
F_6	separator pressure	±2	step	IDV15	CCW valve	sticking
F_7	separator underflow	±20	step			
F_8	stripper level	±10	step			
F_9	stripper underflow	±10	step			
F_{10}	stripper temp	±5	step			
F_{11}	compressor work	±5	step			
F_{12}	RCWO temp	±5	step			
F_{13}	B comp. purge	±10	step			
F_{14}	G/H ratio product	±10	step			

and $2^{52} \approx 4.5 \times 10^{15}$ possible solutions. Taking into account that a single iteration demands $\approx$ 0.156 s, an exhaustive evaluation of this problem is unpractical.

Figure 3 shows the functional cost profiles when different settings are used. Figure 3 shows the GA behavior with $P_{co} = 0.7$, $P_m = P_{co}/N_c$, and the weights for penalizing extra measurement selection being equal to 10. For this case different initial populations were used, N_i = 100, 500, 1000, 5000. Thus N_i affects the convergence time directly. Generally, for small dimension problems, a good choice is to select the initial population equal to 10−20% of the problem dimension. For large combinatorial problems, as in this case, an adequate selection may be $N_i \approx N_c \times 100$. Obviously, the dimension of the initial population is limited by the computational resources. Figure 3b summarizes the fitness profiles when different crossover probabilities, P_{co} = 0.5, 0.7, 0.9, are selected. The mutation probability is adjusted as $P_m = P_{co}/N_c$, the initial population is N_i = 500. In this case a low P_{cm} value, and therefore P_m, limit the recombination and mutation possibilities of the individuals. This behavior may give locally optimal



(a) Varying N_i

(b) Varying P_{co}

(c) Varying the weights for measurement selection

Figure 3. Fitness function for different settings.

Table 3. GA Setting Parameters

N_i	N_c	J	N_g	P_m	P_{co}	selection	crossover	$\mathbf{c}(i)$	optimization time
5000	52	19	100	$0.7/N_c$	0.7	roulette wheel	double-point	10	1.37 h

solutions.[17] On the other hand, an excessive value (on these probabilities) generates an erratic behavior at the beginning of the search. The influence of the sensors penalization weights in eq 20 can be observed in Figure 3c. In this case the weights are adjusted to 0.5, 5, 15, and 30, with the remaining parameters fixed to $P_{co} = 0.7$, $P_m = P_{co}/N_c$ and $N_i = 1000$. Clearly, by imposing a low penalty, eq 20 focuses on minimizing the minimal fault magnitude and new hardware may be included. By increasing the penalization weight the inclusion of new sensors is avoided and, obviously, the achievable MFM is limited to the existing measurement devices. For the cases shown in Figure 3c, and taking into account Table 1, the sensor selection results [1, 7, 11, 16] and [1, 7, 11, 37] for a weight of 0.5 and 5, respectively. The first setting includes two new sensors at locations 7 and 16, and the second one includes only one new sensor at location 7. For weights of 15 and 30 the optimal location is the same [1, 7, 11, 27, 37] and does not include new measurement devices.

Table 3 summarizes the adopted parameters for the GA in this case study. The simulations were performed in an Intel core duo, 2.53 GHz and 2 GB RAM computer. Considering Table 2, $J = 19$ are the potential abnormal events in the TE process. The initial population of individuals N_i provides the starting point for the search and evolution space (initial potential solutions). The GA termination criteria is the maximum generation number, N_g. The individuals selection is carried out by means of the roulette wheel method according to their relative fitness measure. The production of new chromosomes is developed by the crossover operator, in this case double-point method with probability of P_{co} is adopted. Analogously as it happens in the natural evolution, the mutation produces a new genetic structure and basically is applied with a lower probability, P_m. All of these parameter are adjusted by accounting the analysis shown in Figure 3. Finally, $\mathbf{c}(i) = 10$, where $\mathbf{c}$ represents the penalty vector of dimension $[1 \times (52 - 14)]$ related to the measurements selection.

Figure 4 shows the GA results. The fitness profile to the best individual selected along the generations is displayed in the

Table 4. Different Signals Selection

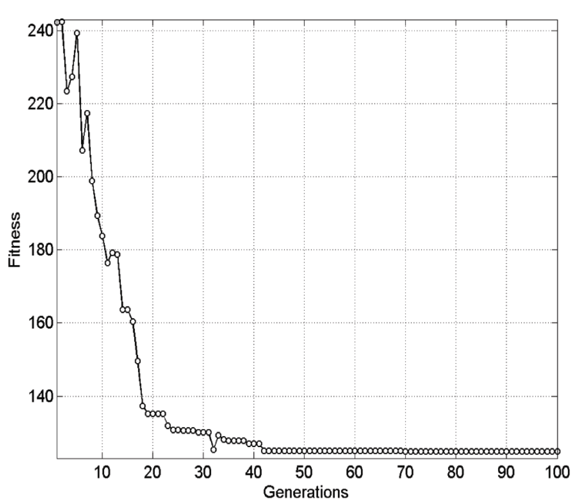
individual	signals selection
$\mathbf{I}_{op}$	$[y_1, y_{10}, y_{11}, y_{27}, y_{37}, u_3, u_9, u_{10}, u_{11}, u_{12}, u_{13}]$
$\mathbf{I}_{unf}$	$[y_1, y_5, y_9, y_{12}, y_{14}, y_{25}, y_{26}, y_{27}, y_{30}, y_{33}, y_{34}, y_{35}, y_{36}, u_1, u_2, u_4, u_7, u_8, u_{14}]$
$\mathbf{I}_{full}$	$[y_1, \cdots, y_{37}, u_1, \cdots, u_{15}]$

Table 5. PCA and Fault Subspace Parameters Setting

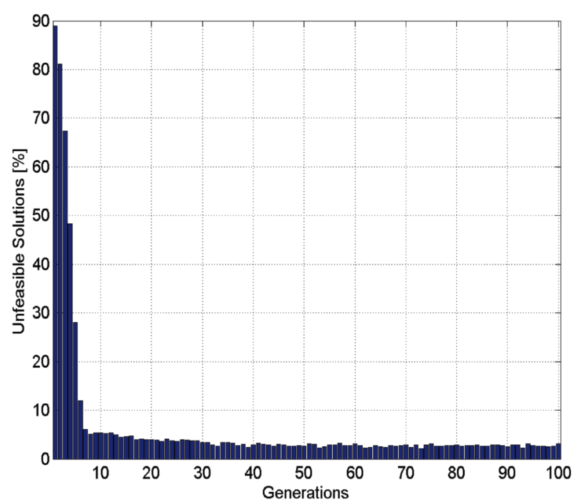
$m \times n$	δ_{cpv}[%]	ν	A_{op}	A_{unf}	A_{full}	δ_z	N_* [samp.]	N [samp.]
841×52	90	3	7	14	29	2	60	20

Figure 4a. Figure 4b shows the unfeasible solutions percentage for each generation. The optimal solution $\mathbf{I}_{op}$, corresponding to the last generation, is presented in Table 4. As can be observed the optimal solution involves only five sensors (y_1, y_{10}, y_{11}, y_{27}, y_{37}) and six signals (u_3, u_9, u_{10}, u_{11}, u_{12}, u_{13}); fortunately all of them already exist in the process control structure. *This means zero additional investment cost because the existing hardware resources are enough for monitoring purposes*. In Table 4 two other alternative signals selection are included for comparison. The solution $\mathbf{I}_{unf}$ does not fulfill the second constraint in eq 21 because it belongs to the unfeasible group. This selection presents two serious problems: losses in fault detection and an increase in the investment costs since it includes new sensors. Another possibility is considered for discussion purposes, that is the achievable performance when the full signal selection case $\mathbf{I}_{full}$ is done. It means that all the signals that appear in the Table 1 are selected without previous considerations. It clearly presents an increase in the cost due to the new measurement points and this sensor network does not guarantee an optimal performance from the monitoring point of view. It seems to indicate that the amount of measurements is not directly related to the quality of fault detection as can be observed in the following paragraphs.

The PCA model parameters corresponding to the TE process are presented in Table 5. In this case the normal database has a dimension of 841 samples and 52 measured process variables. The cumulative percent variance limit has been selected to 90% and the confidence limit for the T^2 and Q is chosen about 99% ($\nu = 3$) in all cases. In addition, the principal component retained for each case and the confidence limit adopted for the combined statistic are included. The setting value for the moving window N in the smoothed moving average (SMA) algorithm and the



(a) Fitness function

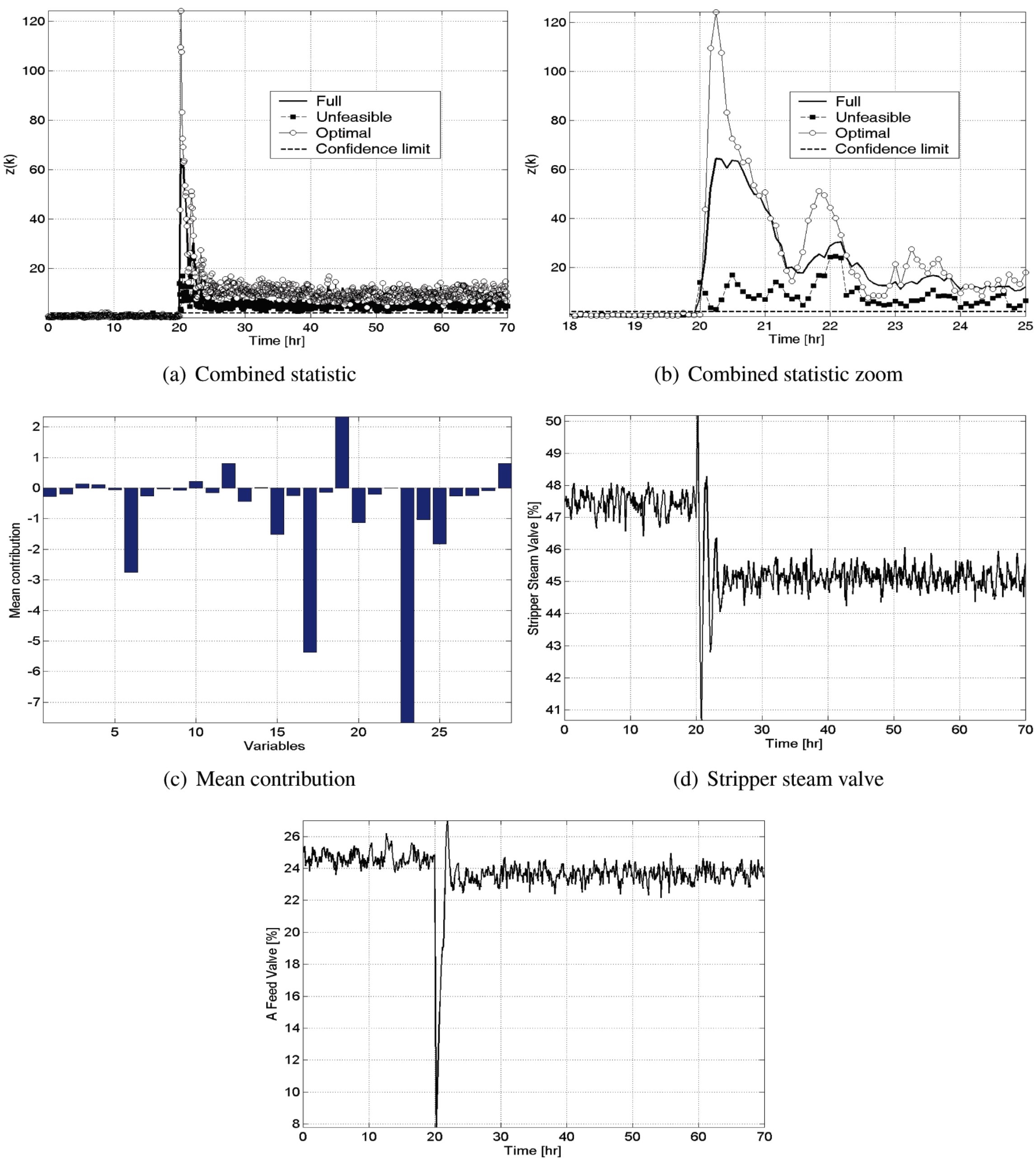


(b) Unfeasible solutions

Figure 4. GA performance.

samples N_* to evaluate the filtered version of the process variables are also presented (see the Appendix).

The first simulation case corresponds to the reactor level sensor offset, recognized as F_3 fault type, and can be observed in Figure 5. The fault magnitude is +10%, and the occurrence time is $T_f = 20$ h. Figure 5 displays the temporal evolution of the combined statistic when different sets of measurements and signals are selected according to Table 4. In Figure 5b a temporal zoom is presented for more clarity. The optimal signals selection, $\mathbf{I}_{op}$, presents better fault detection conditions compared with the full and unfeasible cases. When analyzing which signal presents more clearly the deviation from the mean contribution magnitude from the confident limit, it is evident that the optimal one does. In addition, if an expert system is designed for classification purposes, as it was done by Zumoffen and co-workers,[1-3] the optimal solution presented here is the most suitable. In Figure 5c can be observed this behavior computed over a zone analysis from 20 to 30 h. The used variables correspond to those highlighted in Table 1 ([y_{1-2}, y_5, y_{8-12}, y_{14-15}, y_{17-18}, y_{27}, y_{37}, u_{1-15}] in their autoscaling versions with respect to the normal case. Figure 5 panels d and e summarize the temporal evolutions of the two variables with major contribution in Figure 5c. This fault type modifies the real level of the reactor; in this case of positive offset the real reactor level decreases from its nominal value. Consequently, it can be observed that the transient effects in the pressure and temperature are compensated by the existing control structure. However, an increase in the magnitude of this fault type produces a performance degradation and/or possible plant shutdown because the high pressure violates its allowed limit. This behavior clearly evidences the necessity of including any fault tolerant characteristics for the base case control structure, which will guarantee a profitable and safe process operation.



(a) Combined statistic

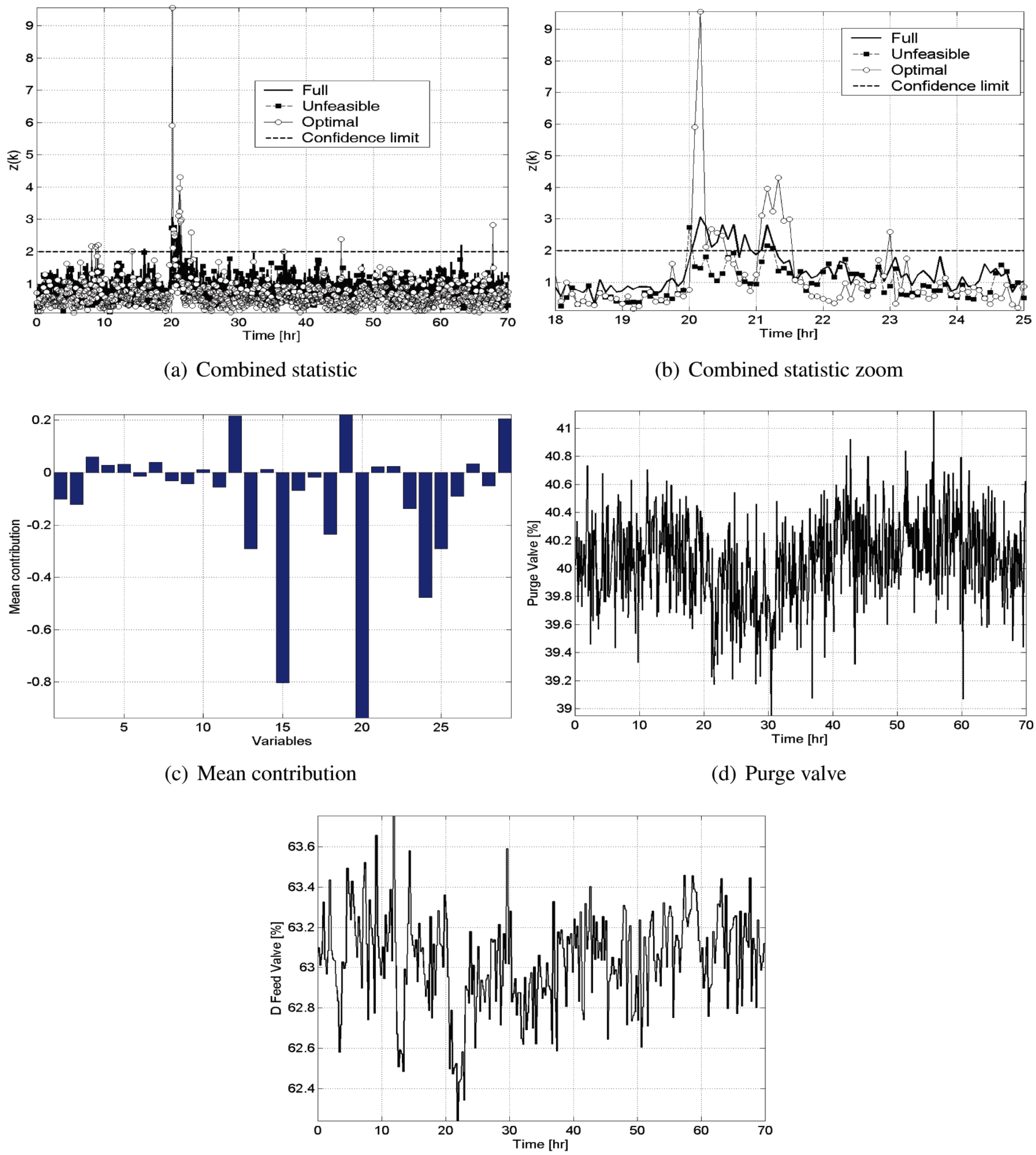
(b) Combined statistic zoom

(c) Mean contribution

(d) Stripper steam valve

(e) A feed valve

Figure 5. Fault F_3 (reactor level offset of 10%).



(a) Combined statistic

(b) Combined statistic zoom

(c) Mean contribution

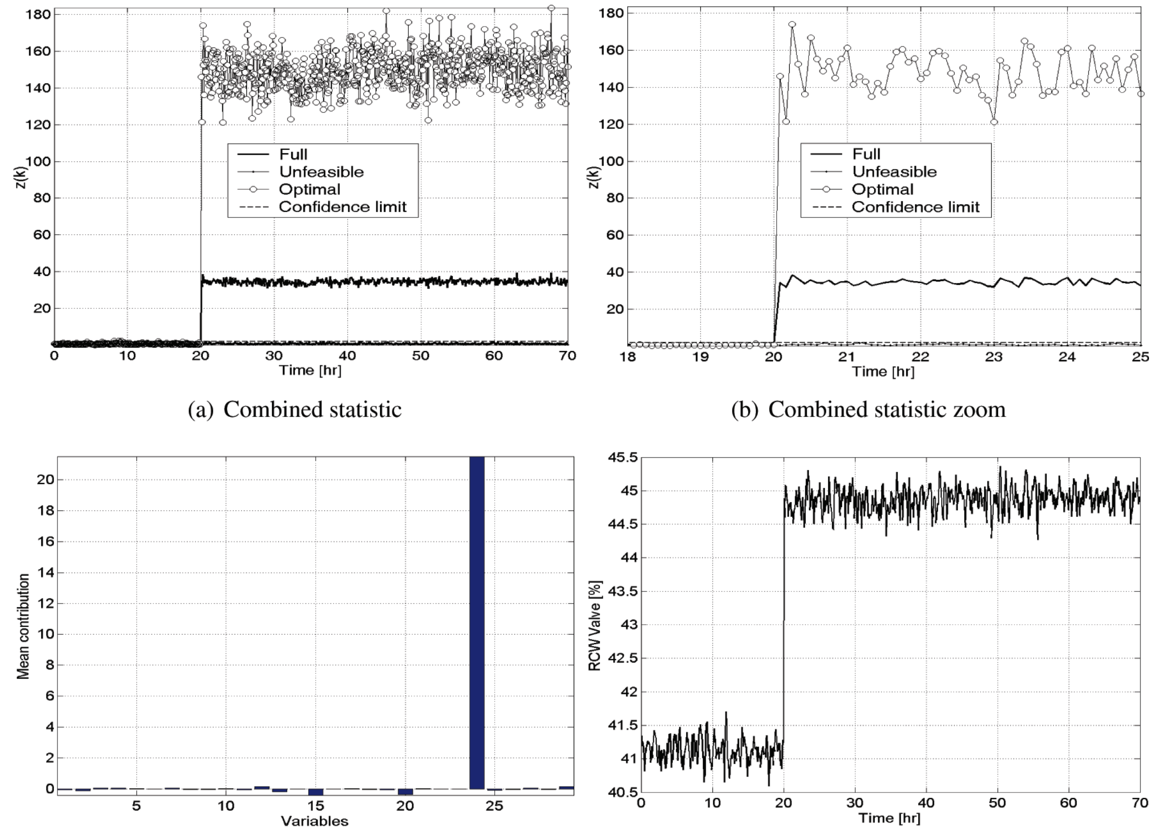
(d) Purge valve

(e) D feed valve

Figure 6. Fault F_5 (separator level offset of 10%).

The second simulation case presented in Figure 5c shows the separator level offset fault, occurring at $T_f = 20$ h, with a value of about +10%. This fault type affects essentially the purge transient flow and the D feed flow. When the fault magnitude is within ±10%, it does not have direct influences on the cost. However, an important increase in this magnitude can produce unacceptable transient behaviors in the purge and D feed flow affecting the quality performances in both B composition in purge and G/H ratio in product. Figure 6a displays the fault detection scenario with combined statistic when different variables and signals are selected (according to Table 4). In Figure 6b a temporal zoom is presented for more clarity. The unfeasible signals selection presents a loss of detection, and the full signal case presents poor detectability conditions. On the other hand the optimal case $\mathbf{I}_{op}$ clearly allows an adequate fault detection. Figure 6c shows the variables contribution for the period between 20 and 30 h, which are autoscaled with respect to the normal case. The two variables with major contribution are displayed in the Figure 6 panels d and e, and they are the purge and D feed valve, respectively.

The Figure 7 presents the process behavior when the IDV4 disturbance appears at 20 h. This abnormal event produces an abrupt variation in the reactor cooling water inlet temperature and its effect impacts the inner control loop in stream 12 only (see Figure 2). Because of this, only a few variables are able to detect these changes. Figure 7a displays the fault detection conditions with combined statistic when different variables and signals are selected (according to Table 4). In Figure 7b a temporal zoom is presented. Clearly the unfeasible signals selection presents a complete loss of detection of this event. The optimal and full selections present comparable behavior. It is a remarkable issue since the optimal structure only uses a few number of signals from the existing sensor network. The



(a) Combined statistic

(b) Combined statistic zoom

(c) Mean contribution

(d) RCW valve

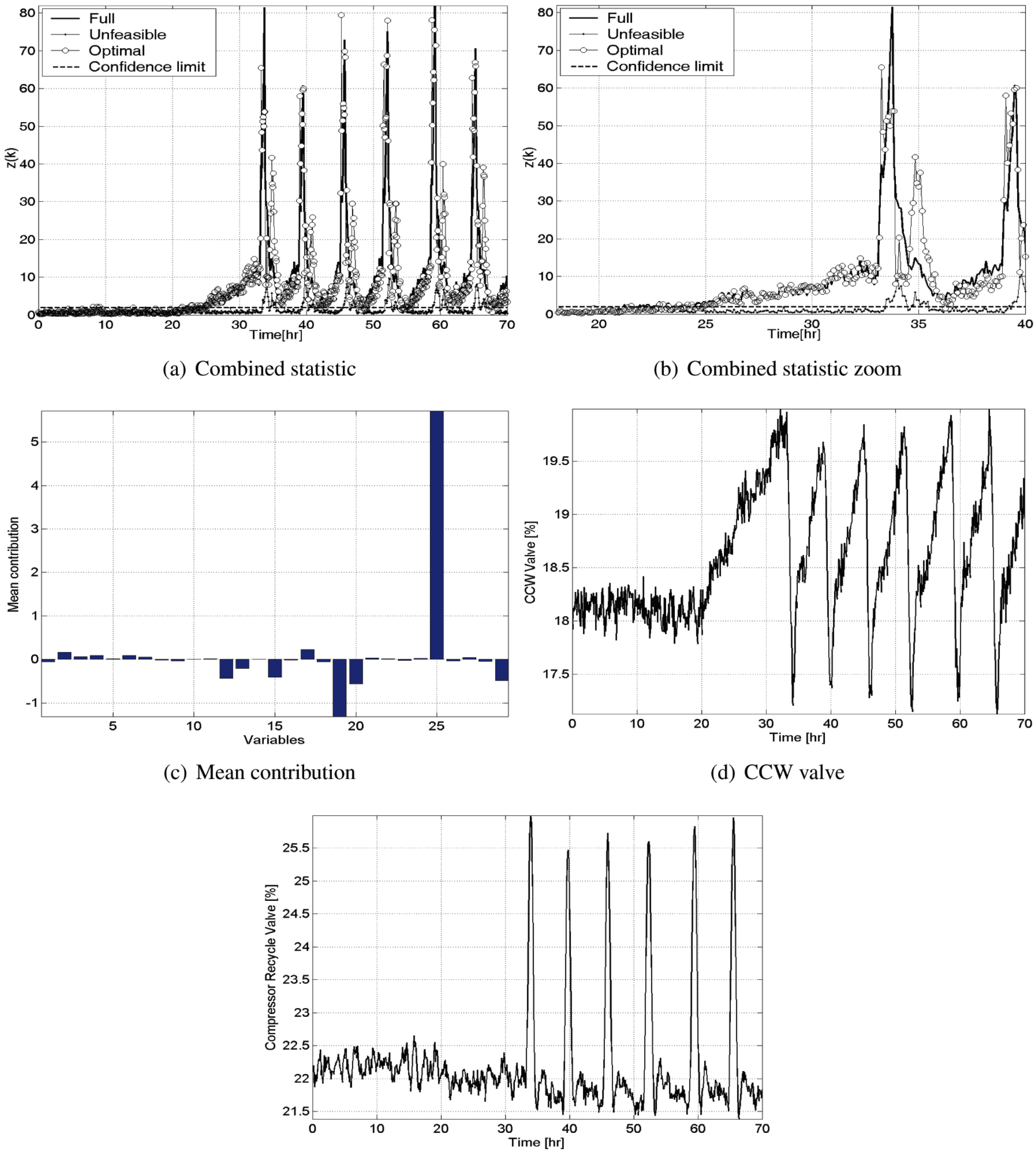
Figure 7. Disturbance IDV4 (RCW inlet temperature).

mean contribution of the variables in the control structure can be observed at Figure 7c where clearly the principal effect comes from the RCW valve variable. Its temporal profile is displayed in Figure 7d.

The final simulation case presented graphically is the effect of the IDV15 disturbance which can be observed in Figure 8. This abnormal event is characterized by a valve sticking in the condenser cooling water stream. The result is that the recycle flow cannot be effectively controlled in its desired operational point affecting several process units and eventually the operational cost. Figure 8a summarizes the combined statistic profiles when different signals selection are used, according to Table 4. In Figure 8b a temporal zoom is presented. The optimal and full cases present comparable performance in the fault detection conditions. On the other hand, the unfeasible selection case presents a considerable delay (≈14 h) for effectively detecting this fault with poor performance. Figure 8c summarizes the mean contribution in the zone analysis from the existing control structure signals. The variables with major contribution are displayed in Figure 8 panels d and e coerresponding to the CCW and compressor recycle valves, respectively.

The remaining simulation cases from Table 2 are summarized in Table 6 using the reliability index. This indicator considers the amount of samples that are over the confidence limit in percentage magnitude. Note that in the normal operation case the process presents 1% of the samples over this limit according to the 99% of confidence. Thus, this index is computed and displayed in Table 6 for each abnormal event given in Table 2 considering different signal selection according to Table 4 information. The unfeasible selection clearly presents loss of detection for faults F_5 and F_8 and the disturbance IDV4 characterized by a reliability index very close to zero. The full selection case also presents difficulties to perform a suitable fault detection. For the faults F_5 and F_8 this selection presents a lower reliability index and poor performance than the optimal one. As can be deduced, the optimal signals selection case uses lower amount of variables with a suitable fault detection performance for the overall abnormal events analyzed here. In all cases the reliability index is comparable with the full case and even better for some of them (F_5 and F_8).

Finally, an important comparison can be done from previous works[11,23] in this area. It is important to note some basic differences from the design and testing point of view. In those works where separated statistics (T^2 and Q) were used, the methodology is tested only under typical disturbances and the signal selection for the monitoring system is performed on a plant with a classical control structure and sensor network design. In contrast, the methodology presented here, uses combined statistic which have better performance, the proposed strategy is rigorously tested under typical disturbances and faults in sensors, and the overall approach is based on a controlled plant that has an optimal sensor network design as well as an optimal control structure.[13] Remembering the process conditions stated by Downs and Vogel,[18] the principal disturbance scenarios are IDV1, IDV4, IDV8, IDV12, and IDV15. Musulin et al[11] presented their results about the reliability index in contrast with the sensor network used by Chen and McAvoy[25] on the TE plant with the base case control strategy. In Table 7 these results are summarized to compare with the obtained ones in Table 6 with combined statistic which gives real improvements. These previous works[11,25] did not display results about the faulty elements in the process. It was limited to test for some disturbances suggested by Downs and Vogel.



(a) Combined statistic (b) Combined statistic zoom

(c) Mean contribution (d) CCW valve

(e) Compressor recycle valve

Figure 8. Disturbance IDV15 (CCW valve sticking).

Table 6. Reliability Index

set	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
full	17.2	100	100	100	2.3	100	98.7	0.3	100	100
unfeasible	18.2	99.3	100	99.7	0.3	97.7	1.2	0	100	98.7
optimal	11.2	100	100	100	2.5	100	39.5	1.7	100	100

set	F_{11}	F_{12}	F_{13}	F_{14}	IDV1	IDV4	IDV8	IDV12	IDV15
full	100	100	100	100	100	100	99.3	99.2	95.2
unfeasible	99.5	6.3	92	99.7	99.2	0	93.5	96.5	21.2
optimal	100	11	68	100	100	100	99.2	99.2	88.5

Table 7. Reliability Indexes and Amount of Sensors in Previous Works

	Statistic	IDV1	IDV4	IDV8	IDV12	sensors
Chen and McAvoy[25]	T^2	99.72	1.97	98.89	35.33	16
	Q	99.86	3.68	99.17	70.91	
Musulin et al[11]	T^2	100.00	2.38	99.29	55.56	6
	Q	93.81	1.07	80.01	35.40	
proposed here	z	100.00	100.00	99.20	99.20	5

6. Conclusions

This work presents a promising alternative of optimal PCA-based monitoring system design for large scale processes. The results obtained here show that the use of GA offers an important opportunity to do a rigorous search among a great number of combinations which is a common task for this kind of multivariable complex systems. Generally, these types of problems are solved with the help of a considerable amount of heuristic considerations. The optimal supervisory system is developed using multiple-integrated tools as PCA, combined statistic, detectability analysis, fault subspace extraction, and associated investment costs. It was found how to include these elements in the objective function so as to promote the use of the existing control structure. In particular, it must be remarked that the detectability function is extended to be calculated on the basis of the combined statistic which represents a new result in the context of expert fault diagnosis system design. In fact, the optimal solution, obtained by this approach, with lower investment costs presents comparable and even better detection performance than that obtained by the full selection case. It is done by using only the already existing hardware from the optimal control structure. This approach integrated with the previous work of the authors[13] impels a generalized and systematic strategy to solve typical problems in process industries as sensors location, control structure selection, and monitoring system design. Additionally, this methodology can be tied with other open problems such as, process synthesis, investment cost analysis, observability, fault diagnosis, and fault tolerant control,[1,2] which is for future work.

Acknowledgment

The authors want to acknowledge the financial support from CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), ANPCYT (Agencia Nacional de Promoción Científica y Técnica) from Argentina. The authors also acknowledge professor Ricker's control group of the Washington University for the rigorous model of the Tennessee Eastman Process in the MatLab environment. Finally, we are grateful for support from the Technological National University-Regional Faculty of Rosario-Argentina.

Appendix

A.1. Fault Subspace Extraction

The fault subspace, Θ, can be extracted from the abnormal database by processing the 52 potential measurements and signals from Table 1. Initially, due to the typical noise present in the process measurements a smoothed moving average (SMA) filter is applied for consistency. It performs the average (mean) of the original signal $x(k)$ over a specified moving window of dimension $N + 1$ samples, as can be observed in eq 24, resulting in the filtered version $x_f(k)$.

$$x_f(k) = (N + 1)^{-1}\sum_{i=0}^{N} x(k - i) \qquad (24)$$

An improved algorithm exists to avoid problems with lagged samples called exponential weighted moving average (EWMA) for online applications.[23,15] Thus, before the faults subspace extraction occurs, the autoscaled database is filtered using eq 24.

The faults direction are computed using the well-known "3δ edit rule"[24] which suggests that if $|x_f(k)| > 3$ the variable is considered to be deviated from its normal state. Analyzing the overall database for each abnormal case present the fault directions can be computed as

$$F(j,i) = \begin{cases} 1, & \text{if } \bar{x}_f^{ji}(k_*) > 3 \\ 0, & \text{if } -3 \le \bar{x}_f^{ji}(k_*) \le 3 \\ -1, & \text{if } \bar{x}_f^{ji}(k_*) < -3 \end{cases} \qquad (25)$$

$$\text{with} \quad F(j,i) = \begin{matrix} j = 1, ..., J \\ i = 1, ..., N_c \end{matrix}$$

where $\bar{x}_f^{ji}(k_*)$ is the autoscaled and filtered version of the variable number i from the abnormal database (event) j evaluated in the sampled instant k_*. This temporal instant is specified according to the dynamic response of the process trying to avoid the transient behavior due to the fault occurrence $k_* = t_f + N_*$. Where t_f is the fault occurrence sample. Thus each row of F represented by $F(j,1 = N_c)$ corresponds to the j fault propagation over the N_c variables analyzed from the abnormal database. The faults subspace Θ can be computed directly from the fault direction matrix $\mathbf{F}$ as can be observed in eq 26

$$\Theta_j = \mathrm{nzr}\{\mathrm{diag}[\mathbf{F}(j, 1 = N_c)]\}^T \qquad (26)$$

where the function diag(•) takes the vector $\mathbf{F}(j,1 = N_c)$ as input argument and gives back a diagonal matrix of $N_c \times N_c$ with $\mathbf{F}(j,1 = N_c)$ in its diagonal. On the other side, the function nzr{•} takes as input argument the diagonal matrix constructed previously and gives back another matrix Θ_j. This fault subspace matrix contains only the nonzero rows from the original diagonal matrix. Finally, this fault propagation matrix can be applied in the optimization algorithm of eq 20 and eq 21.

An alternative approach to obtain the fault direction matrix $\mathbf{F}$ exists.[1-3] In these works a strategy based on fuzzy logic tools is applied and the resultant matrix rules represent the fault propagations over the process variables. In this case the matrix rules evaluation accounts the mean contribution of the variables within the specified zone of analysis.

A.2. Nomenclature

Variables

A = principal component retained
$\mathbf{b}$ = mean vector
$\mathbf{c}$ = cost vector
CPV(A) = CPV to A
$\mathbf{D}_A$ = first A eigenvalues matrix
$\mathbf{D}_\lambda$ = full eigenvalues matrix
$\mathbf{E}$ = residual matrix
$\mathbf{f}_i$ = fault components vector
F_i = type i fault
$\|\mathbf{f}_j\|^i_{\mathrm{MFM}}$ = MFM for i statistic and fault j
g_i = gen parameter i
$\mathbf{I}_i$ = chromosome i
$\mathbf{I}_{\mathrm{op}}$ = optimal selection
$\mathbf{I}_{\mathrm{unf}}$ = unfeasible selection
$\mathbf{I}_{\mathrm{full}}$ = full selection
IDVi = disturbance i
J = abnormal events
k_* = evaluation sample
m = data matrix samples
$\mathbf{M}$ = combined statistic matrix
n = data matrix variables
N = moving window dimension
N_* = evaluation window
N_c = chromosome length
N_g = total generations
N_i = initial population
$\mathbf{p}_i$ = eigenvector i
P_{co} = crossover probability
P_m = mutation probability
$\mathbf{P}$ = PCA model
$\mathbf{P}_b$ = best population set
$\mathbf{P}_j$ = actual population set
$\mathbf{P}_m$ = mutated population set
$\mathbf{P}_r$ = recombined population set
$\mathbf{P}_s$ = selected population set
Q = SPE statistic
$\mathbf{R}_c$ = correlation matrix

s = variance vector
t = latent variable
T^2 = Hotelling statistic
T = latent matrix
u_i = input i
U = eigenvector matrix
x = signal
x_f = filtered signal
x = process sample
$\bar{\mathbf{x}}$ = normalized process sample
$\mathbf{x}_*$ = normalized process sample model
$\bar{\mathbf{x}}_0$ = normal behavior case
X = process data matrix
$\bar{\mathbf{X}}$ = normalized process data matrix
y_i = output i
z = combined statistic

Greek Symbols

δ_i = confidence limit i
λ_i = eigenvalue i
μ_i = mean value for i signal
σ_i = variance value for i signal
Θ_j = j fault subspace
F = Fault directions matrix

Abbreviations

CPV = cumulative percent variance
CV = controlled variable
CCW = compressor cooling water
EWMA = exponential weighted moving average
FD = fault detection, GA = genetic algorithm
IMC = internal model control
MFM = minimal fault magnitude
MV = manipulated variable
PCA = principal component analysis
RCW = recycle cooling water
RCWO = RCW outlet
RGA = relative gain array,
SMA = smoothed moving average
SPE = square predictive error
SSE = sum of square errors
SVD = singular value decomposition
TE = Tennessee Eastman

Literature Cited

(1) Zumoffen, D.; Basualdo, M. From large chemical plant data to fault diagnosis integrated to decentralized fault-tolerant control: Pulp mill process application. *Ind. Eng. Chem. Res.* **2008**, *47*, 1201–1220.

(2) Zumoffen, D.; Basualdo, M. Improvements in fault tolerance characteristics for large chemical plants. Part I: Waste water treatment plant with decentralized control. *Ind. Eng. Chem. Res.* **2008**, *47*, 5464–5481.

(3) Zumoffen, D.; Basualdo, M.; Molina, G. Improvements in fault tolerance characteristics for large chemical plants. Part II: Pulp mill process with model predictive control. *Ind. Eng. Chem. Res.* **2008**, *47*, 5482–5500.

(4) Kourti, T.; MacGregor, J. Process anlysis, monitoring and diagnosis, using multivariable projection methods. *Chemom. Intell. Lab. Syst.* **1995**, *28*, 3–21.

(5) Martin, E.; Morris, J.; Lane, S. Monitoring process manufacturing performance. *IEEE Control Syst. Mag.* **2002**, *22*, 26–39.

(6) Lane, S.; Martin, E.; Morris, J.; Gpwer, P. Application of exponentially weighted principal component analysis for the monitoring of a polymer fil manufacturing process. *Trans. Inst. Measure. Control* **2003**, *25*, 17–35.

(7) Yue, H. H.; Qin, S. J. Reconstruction-based fault identification using a combined index. *Ind. Eng. Chem. Res.* **2001**, *40*, 4403–4414.

(8) Kadu, S.; Bhushan, M.; Gudi, R. Optimal sensor network design for multirate system. *J. Process Control* **2008**, *18*, 594–609.

(9) Singh, A.; Hahn, J. Determining optimal sensor location for state and parameter estimation for stable nonlinear systems. *Ind. Eng. Chem. Res.* **2005**, *44*, 5645–5659.

(10) Bhushan, M.; Narasimhan, S.; Rengaswamy, R. Robust sensor network design for fault diagnosis. *Comput. Chem. Eng.* **2008**, *32*, 1067–1084.

(11) Musulin, E.; Bagajewicz, M.; Nougués, J. M.; Puigjaner, L. Instrumentation design and upgrade for principal components analysis monitoring. *Ind. Eng. Chem. Res.* **2004**, *43*, 2150–2159.

(12) Musulin, E.; Yelamos, I.; Puigjaner, L. Integration of principal component analysis and fuzzy logic systems for comprehensive process fault detection and diagnosis. *Ind. Eng. Chem. Res.* **2006**, *45*, 1739–1750.

(13) Zumoffen, D.; Basualdo, M. In *Optimal sensor location for chemical process accounting the best control configuration*; Alves, R. D., Nascimento, C. O., Biscaia, E., Eds.; Computer Aided Chemical Engineering, Vol. 27; Elsevier, 2009; pp 1593−1598.

(14) Li, W.; Yue, H. H.; Valle-Cervantes, S.; Qin, S. J. Recursive PCA for adaptive process monitoring. *J. Process Control.* **2000**, *10*, 471–486.

(15) Wold, S. Exponentially weighted moving principal components analysis and projection to latent structures. *Chemom. Intell. Lab. Syst.* **1994**, *23*, 149–161.

(16) Ruiz, D. Fault diagnosis in chemical plants integrated to the information system. Ph.D. Thesis. Departament d'Enginyeria Quimica, Escola Tècnica Superior d'Enginyers Industrials de Barcelona, Universitat Politècnica de Catalunya, Spain, 2001.

(17) Chipperfield, A.; Fleming, P.; Pohlheim, H.; Fonseca, C. *Genetic algorithm toolbox for use with Matlab*. University of Sheffield, Department of Automatic Control and Systems Engineering: Sheffield, U.K., 1994; http://www.shef.ac.uk/acse/research/ecrg.

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

(19) Ricker, N. Optimal steady-state operation of the Tennessee-Eastman challenge process. *Comput. Chem. Eng.* **1995**, *19*, 949–959.

(20) Ricker, N. Decentralized control of the Tennessee-Eastman challenge process. *J. Process Control* **1996**, *6*, 205–221.

(21) Ricker's Control Department of the Washington University. http://depts.washington.edu/control (Accessed 2007).

(22) Abnormal situation management (ASM) consortium. http://www.asmconsortium.com (Accessed 2006).

(23) Musulin, E. Process monitoring and abnormal situation management in chemical processes. Ph.D. Thesis. Departament d'Enginyeria Quimica, Escola Tècnica Superior d'Enginyers Industrials de Barcelona, Universitat Politècnica de Catalunya, Spain, 2005.

(24) Chiang, L. H.; Pell, R. J.; Seasholtz, M. B. Exploring process data with the use of robust outlier detection algorithms. *J. Process Control* **2003**, *13*, 437–449.

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

Received for review March 9, 2009
Revised manuscript received December 8, 2009
Accepted December 20, 2009

IE9017836