I&EC research
Industrial & Engineering Chemistry Research

Article

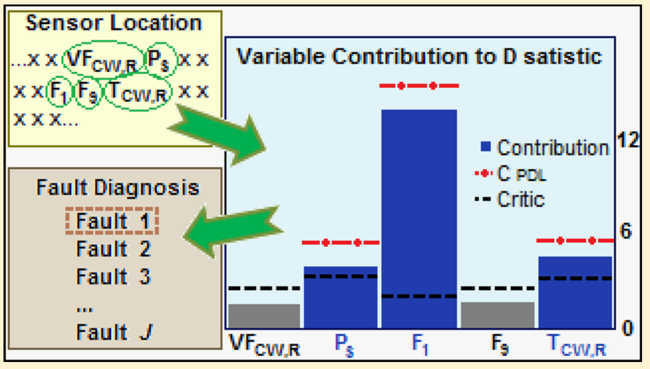
pubs.acs.org/IECR

Sensor Location for Enhancing Fault Diagnosis

Leandro P. F. Rodriguez, Marco V. Cedeño, and Mabel C. Sánchez*

Planta Piloto de Ingeniería Química, Universidad Nacional del Sur - CONICET, Camino La Carrindanga km 7, 8000, Bahía Blanca, Argentina

ABSTRACT: Multivariate statistical control techniques have been successfully applied to the detection and isolation of process faults. Because those strategies evaluate the current process state using the measurement values and the normal operation model, their performance is strongly influenced by the sensor network installed in the plant. Nevertheless, very few sensor location approaches have been presented to enhance faults isolation, and they do not guarantee that a fault can be distinguished from the other ones before its magnitude reaches a certain critical value. In this work a strategy for updating process instrumentation is presented that aims at detecting and isolating a given set of process failures using statistical monitoring procedures before the fault magnitudes exceed predefined threshold values. In this sense, fault isolation constraints are formulated and incorporated to the instrumentation update optimization problem. The proposed restrictions are expressed in terms of the variable contributions to the inflated statistics. These are used on line to determine the set of observations by which a fault is revealed, but they have not been incorporated into the sensor location problem for fault diagnosis until now. That problem is solved using an enhanced level traversal tree search, which takes advantage of the fact that the structural determinability of a fault is a necessary condition for its isolability. Application results of the methodology to the Tennessee Eastman Process are presented.



1. INTRODUCTION

Multivariate statistical approaches such as principal component analysis (PCA) have been successfully applied to the detection and isolation of process faults.[1−3] Those methodologies compare the current process state with the predictions given by the normal operation model to decide if a fault has occurred. The result of this statistical inference depends on the measurement values, in consequence, fault detection and isolation (FDI) strategies provide better conclusions if the process sensor network (SN) has been properly selected. This is achieved by solving an optimization problem, which is commonly known as the SN design or upgrade problem.[4]

Comprehensive reviews about the variety of methods proposed to locate instruments with fault diagnosis purposes have been presented.[5−7] The instrumentation selection strategies for processes to be monitored using PCA satisfy at least the observability and resolution of a predetermined set of faults. In this way, the detection and isolation of those failures are partially ensured.[5,8] It has been suggested to select sensors at the design stage by applying directed-graph-based procedures[9] or a technique that assumes the fault probability values for the instruments and the process.[10] It is well-known that the information provided by that SN may not be sufficient to distinguish which fault has occurred.[5] Therefore, it has proposed to obtain a residual matrix for each fault by simulation and to consider its first eigenvector as the direction of the failure. During plant operation, if the statistical technique has been able to detect the occurrence of a fault, then the residual vector of the PCA model is projected on the matrix of fault directions, and the maximum projection value is associated with the current fault. One drawback of this technique is that the first eigenvector of the residual matrix may not correctly retain the fault direction.[5]

Other strategies select instruments to satisfy the detection of certain process faults before their magnitudes exceed critical limits. Those methodologies assume that a simulation procedure which sensibly represents the process dynamic response is available.[11,12] Some researchers have formulated an optimization problem whose objective function is a global penalty index.[11] It is made up of the total instrumentation cost and a penalization term, which is calculated as a function of the minimum fault magnitude[13] that PCA's statistics can detect. Its calculus is conservative; therefore, the strategy may neglect lower-cost feasible solutions. This inconvenience is avoided using fault detection constraints straightaway defined in terms of the PCA's statistics, which have been used to solve instrumentation upgrade problems by applying an enhanced traversal tree search.[12] Furthermore, the robustness of the SN in the presence of sensor failures or outliers has been ensured by introducing the Key Fault Detection Degree concept as constraint of the optimization problem.[12]

Received: June 13, 2016
Revised: July 20, 2016
Accepted: July 23, 2016

DOI: 10.1021/acs.iecr.6b02288
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

The isolation of process failures before their magnitudes exceed critical limits provides valuable information for implementing corrective actions while the plant is still operating in a controllable region. Nevertheless, the sensor location strategies for fault diagnosis have not addressed that issue until now. This work aims at presenting a methodology to optimally locate new sensors in an existing process to guarantee the detection and isolation of a predetermined set of faults while the process is still operating in a safe mode. With this purpose, fault isolation constraints are formulated as a function of the variable contributions (VCs) to the inflated statistics. These are used when statistical monitoring techniques are applied on line to determine the set of observations by which a fault is revealed, but they have not been previously incorporated into the sensor location problem for fault diagnosis. Furthermore, an exact solution procedure of the instrumentation update problems is presented.

The rest of the paper is organized as follows. In section 2, the classic PCA procedure is briefly revised. A sensor network upgrade problem (SNUP) which guarantees FDI is presented in section 3, which also describes a methodology to solve the optimization problem. Next, the application results of the proposed strategy to a benchmark are analyzed. The last section contains the conclusions of this research work.

2. FAULT DIAGNOSIS USING PCA

PCA is a statistical method used to reduce the dimensions of a data set. It applies an orthogonal transformation to convert a set of measurements of correlated variables into a set of values of linearly uncorrelated variables known as principal components (PCs). The transformation is stated in such a way that the first PC has the largest possible variance, and each subsequent component in turn has the highest variance possible under the restriction that it is orthogonal to the preceding ones. Different techniques are applied to determine the number of retained PCs. The nonretained ones are discarded to reduce the system dimensions.

When PCA-based techniques are used for process monitoring, the calculation of the hotelling statistic (D) and the squared prediction error (SPE) are performed at each time interval to find out if a fault has occurred. The statistic values are evaluated using the following expressions

$$D = \| \mathbf{\Lambda}^{-1/2}\mathbf{P}^{\mathrm{T}}\mathbf{x} \|^2 \quad (1)$$

$$SPE = \|(\mathrm{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})\mathbf{x}\|^2 \quad (2)$$

where $\mathbf{P}$ and $\mathbf{\Lambda}$ are matrices obtained by the eigendecomposition of the correlation matrix $\mathbf{R}$ and comprise the retained PCs and their eigenvalues, respectively. The matrix $\mathbf{R}$ is calculated using the normal operating data included in the matrix $\mathbf{X}(M, N)$, where M is the number of samples and N stands for the number of measured variables. Furthermore, $\mathbf{x}$ represents the standardized measurement vector.

Both statistics are compared with their respective critical values, and if one of them at least exceeds its limit for three consecutive measurements,[14] then it is assumed that the process is out of control. It is a common industrial practice to calculate the VCs to the statistics at this point and to determine which ones exceed their control limits. This information is useful to diagnose the fault cause and to implement appropriate corrective actions.

Several methodologies have been proposed to evaluate the contribution of the nth variable to D's value (c_n^D) when the fault is detected. The most used technique[15] evaluates this contribution as follows

$$c_n^D = \mathbf{t}^{\mathrm{T}}\mathbf{\Lambda}^{-1}\mathbf{P}_{(:,n)}x_n \quad (3)$$

where $\mathbf{t}$ is the score vector obtained by projecting $\mathbf{x}$ onto the space defined by $\mathbf{P}$. The control limit of c_n^D is calculated by means of the following expression:

$$c_n^D|_C = \overline{c_n^D} + \kappa\sigma_{c_n^D} \quad (4)$$

where $c_n^D|_C$ is the control limit for the nth VC to D and $\overline{c_n^D}$ and $\sigma_{c_n^D}$ are the mean value and standard deviation of c_n^D, respectively. These are calculated using the data contained in matrix $\mathbf{X}$. A suggested value for the parameter κ is 3.[15]

The definition of the nth VC to SPE (c_n^{SPE}) directly arises from the reformulation of eq 2:

$$SPE = \|\mathbf{x} - \mathbf{P}\mathbf{P}^{\mathrm{T}}\mathbf{x}\|^2 = \sum_{n=1}^{N} (x_n - (\mathbf{P}\mathbf{P}^{\mathrm{T}}\mathbf{x})_n)^2 = \sum_{n=1}^{N} c_n^{\mathrm{SPE}} \quad (5)$$

where $(\mathbf{P}\mathbf{P}^{\mathrm{T}}\mathbf{X})$ is the vector of measurement values predicted by the model. The control limit for c_n^{SPE}, $c_n^{SPE}|_C$, is set in terms of the χ^2 distribution[16]

$$c_n^{SPE}|_C = \frac{\sigma^2_{c_n^{SPE}}}{2\overline{c_n^{SPE}}}\chi^2_{(2(\overline{c_n^{SPE}})^2/\sigma^2_{c_n^{SPE}},\alpha)} \quad (6)$$

where $\overline{c_n^{SPE}}$ and $\sigma^2_{c_n^{SPE}}$ are the mean value and standard deviation of c_n^{SPE}, respectively, for the samples included in $\mathbf{X}$ and α is the level of significance of the hypothesis test.

Another technique makes use of the combined index[17] for the detection of faults when PCA is applied. The index is based on the values of D, SPE, and their critical limits. That methodology applies the reconstruction-based contributions and their control limits to identify the variables by which the fault is revealed. In this work, it is assumed that the process is monitored using the classic PCA procedure, but the instrumentation update problems proposed in the next section can be directly extended to processes monitored using the combined index.

3. SENSOR NETWORK UPGRADE FOR FAULT DETECTION AND ISOLATION

In this section a new methodology for the optimal upgrade of a plant SN is presented. The location of new instruments intends to satisfy the isolation of J process faults. This set of failures has been previously defined by process engineers taken into account process operation and safety objectives. At first, the formulation of new fault isolation restrictions for the minimum cost update problem is addressed. Then, this problem is defined and a solution procedure is briefly described.

3.1. Fault Isolation Constraints. Let us assume that engineers have set process deviation limits (PDLs)[11] for some process variables taking into account operation and safety issues and that a simulation procedure which sensibly represents the process dynamic response is available.[5,11,12] If the simulation of the jth fault (j = 1, ..., J) is performed until the time for which one or more variables reach their PDLs, then the vector of standardized variable values for that time, $\mathbf{x}_j$, is obtained. A set of simulated data that represents the normal process variability can be employed for the standardization. It has been proposed to use the information contained in $\mathbf{x}_j$ (j = 1, ..., J) with two purposes.

DOI: 10.1021/acs.iecr.6b02288
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

The first aim is to define the cause-effect matrix $\mathbf{A}$. It is considered that the elements of $\mathbf{x}_j$ greater than 3 signal the variables affected by the occurrence of the jth failure.[11] Therefore, if the whole set of analyzed faults have influence on a total number of process variables equal to I, then $\mathbf{A}(I, J)$ is composed by a_{ij} elements such that $a_{ij} = 1$ if the ith variable value is significantly modified when the jth failure occurs and otherwise $a_{ij} = 0$ $(i = 1, ..., I)$. The jth column of $\mathbf{A}$, $\mathbf{a}_j$, is a binary vector that contains the same information for that particular fault.[12] The second purpose is to determine a vector $\mathbf{x}_j^{\mathbf{PDL}}(j = 1, ..., J)$ such that it only includes the elements of $\mathbf{x}_j$ corresponding to the I variables involved in $\mathbf{A}$.

For an operating plant, some of the variables affected by the occurrence of the J faults may be already measured, but if the existing SN does not satisfy the isolation of all the failures when PCA is applied for process monitoring, then an instrumentation upgrade problem should be formulated and solved.

A current solution of the aforementioned problem can be represented by a binary vector $\mathbf{q}$, of dimension I, such that $q_i = 1$ if the ith variable is considered as a measured one. The sum of the nonzero elements of $\mathbf{q}$, i.e., the total number of observations, is represented by N. Therefore, $(I - N)$ stands for the number of unmeasured variables. The N measurements are related to the existing sensors and the new instruments involved in the current solution. For this vector $\mathbf{q}$, the jth fault subspace vector $\mathbf{sf}_j$ $(j = 1, ..., J)$ of dimension N is defined such that $sf_{n,j} = 1$ if the nth variable value is modified when the jth fault occurs, and in contrast, $sf_{n,j} = 1$. That is, the information contained in $\mathbf{q}$ and $\mathbf{a}_j$ is used to build $\mathbf{sf}_j$. If the same fault subspace vector arises for two failures, then they cannot be isolated by the current solution of the upgrade problem.

In general, the nth observation of an instrumentation system is identified as a suspicious variable if its contribution to the value of a certain statistic, S, is greater than its control limit, i.e., $c_n^S(\mathbf{q}) > c_n^S|_C(\mathbf{q})$. This means that the occurrence of a certain fault has produced a change in the nth measurement value with respect to the normal one. Therefore, to distinguish the jth fault from the other ones, a prerequisite is that at least one of the measurements related to $\mathbf{sf}_j$ is signaled as suspicious. This points out the possible occurrence of the fault.

Let us suppose that the jth fault has occurred and that the faulty state has been declared at the time when one or more variables reach their PDLs. At this moment, the VCs to the PCA's statistics are the following

$$c_{n,j}^{D,s} = \mathbf{t}^{\mathbf{T}}\mathbf{\Lambda}^{-1}\mathbf{P}_{(:,n)}x_{n,j}^{\mathrm{PDL}*} \quad n = 1, ..., N \tag{7}$$

$$c_{n,j}^{SPE,s} = (x_{n,j}^{\mathrm{PDL}*} - (\mathbf{P}\mathbf{P}^{\mathbf{T}}\mathbf{x}_j^{\mathbf{PDL}*})_n)^2 \quad n = 1, ..., N \tag{8}$$

where $c_{n,j}^{D,s}$ and $c_{n,j}^{SPE,s}$ are the nth VCs to D and SPE, respectively, $\mathbf{t} = \mathbf{P}^{\mathbf{T}}\mathbf{x}_j^{\mathbf{PDL}*}$ and $x_{n,j}^{\mathrm{PDL}*}$ is the nth element of $\mathbf{x}_j^{\mathbf{PDL}*}$. This vector contains the N elements of $\mathbf{x}_j^{\mathbf{PDL}}$ corresponding to the nonzero elements of $\mathbf{q}$. If VCs control limits are greater than VCs values, i.e., $c_n^S|_C(\mathbf{q}) > c_{n,j}^{D,s}(\mathbf{q})$ and $c_n^{SPE}|_C(\mathbf{q}) > c_{n,j}^{SPE,s}(\mathbf{q})$, then any variable is identified as suspicious; nevertheless, some of them have reached their PDLs at that time.

The situation described previously is an undesirable event. When the jth fault occurs, the SN should be able to point out the suspicious variables while the process is still operating in a safe mode, i.e., before the PDL of at least one measured variable contained in $\mathbf{sf}_j$ is exceeded. To achieve this objective, VCs values at that time should be greater than VCs control limits. Because the classic PCA monitoring technique uses two statistics, the nth observation is declared as a suspicious variable if one of the following logic statements is satisfied

$$c_{n,j}^{SPE,s}(\mathbf{q}) \geq c_n^{SPE}|_C(\mathbf{q}) \vee c_{n,j}^{D,s}(\mathbf{q}) \geq c_n^{D}|_C(\mathbf{q}) \tag{9}$$

$$c_{n,j}^{SPE,s}(\mathbf{q}) \geq c_n^{SPE}|_C(\mathbf{q}) \wedge c_{n,j}^{D,s}(\mathbf{q}) \geq c_n^{D}|_C(\mathbf{q}) \tag{10}$$

The first one is verified if either $c_{n,j}^{D,s}(\mathbf{q})$ or $c_{n,j}^{SPE,s}(\mathbf{q})$ is greater than its control limit. The second statement is confirmed if both contributions exceed their control thresholds. The result of the identification task is represented by the binary vector $\mathbf{iv}_j$ of dimension N. If $iv_{n,j} = 1$, then the nth measurement has been declared as a suspicious variable (i.e., inequalities 9 or 10 are satisfied), and $iv_{n,j} = 0$ otherwise.

If $\mathbf{iv}_j$ is a non-null vector, then the prerequisite for the isolation of the jth fault is fulfilled, but to distinguish this fault from the other ones, the following two conditions should be satisfied.

Condition 1: The nth element of $\mathbf{iv}_j$ is equal to the nth element of $\mathbf{sf}_j$:

$$iv_{n,j}(\mathbf{q}) = sf_{n,j}(\mathbf{q}) \quad n = 1, ..., N \quad j = 1, ..., J \tag{11}$$

This ensures that all the measured variables related to the occurrence of the jth failure are pointed out as suspicious variables.

Condition 2: The set of variables identified as suspicious when the jth fault occurs is different from the set pointed out if the kth fault happens, $\forall\, k \neq j$.

$$\mathbf{iv}_j(\mathbf{q}) \neq \mathbf{iv}_k(\mathbf{q}) \;\forall\, j \neq k \tag{12}$$

This guarantees the jth fault is distinguished from the kth one.

To test condition 2, three vectors are defined in terms of $\mathbf{iv}_j$ and $\mathbf{iv}_k$ as follows

$$\mathbf{v1} = \mathbf{iv}_j \wedge \neg\mathbf{iv}_k \;\forall\, j \neq k \tag{13}$$

$$\mathbf{v2} = \neg\mathbf{iv}_j \wedge \mathbf{iv}_k \;\forall\, j \neq k \tag{14}$$

$$\mathbf{v3} = \mathbf{iv}_j \wedge \mathbf{iv}_k \;\forall\, j \neq k \tag{15}$$

where $\mathbf{v}_1$ contains all the measured variables of $\mathbf{iv}_j$ not included in $\mathbf{iv}_k$, in contrast $\mathbf{v}_2$ involves the set of observations of $\mathbf{iv}_k$ not contained in $\mathbf{iv}_j$, and $\mathbf{v}_3$ has the common elements. It has been demonstrated[6] that faults j and k are isolatable if at least two of the three vectors $(\mathbf{v1}, \mathbf{v2}, \text{and } \mathbf{v3})$ are non-null, that is, if the number of non-null vectors that arise from the conjunction between $\mathbf{iv}_j$ and $\mathbf{iv}_k$, denoted as $NNV_{jk}(\mathbf{q})$, is greater than 2.

$$NNV_{jk}(\mathbf{q}) \geq 2 \;\forall\, j \neq k \tag{16}$$

It has been demonstrated that the previous inequality can be replaced by the following set of constraints[6]

$$\sum_{n=1}^{N} (\mathbf{iv}_j)_n \geq 1 \quad j = 1, ..., J$$

$$\sum_{n=1}^{N} (\mathbf{iv}_j \wedge \neg\mathbf{iv}_k)_n + \sum_{n=1}^{N} (\mathbf{iv}_k \wedge \neg\mathbf{iv}_j)_n \geq 1$$

$$j = 1, ..., J,\; k = 1, ..., J,\; j \neq k \tag{17}$$

3.2. Optimization Problem Formulation and Resolution.

At first, the formulation of the upgrade problem of SNs that satisfy FDI constraints is addressed. It is assumed that there is only one potential measuring device for each variable and that no instruments' localization restrictions are imposed. Next, the problem resolution procedure is described.

DOI: 10.1021/acs.iecr.6b02288
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

To ensure that the updated SN is able to detect the predetermined set of J faults before one or more process variables reach their PDLs, fault detection restrictions straightaway formulated as a function of the PCA's statistics are incorporated to the SNUP. These constraints are represented as follows[12]

$$[(D_j^s(\mathbf{q}) \geq \delta_{D,\alpha}^2(\mathbf{q})) \vee (SPE_j^s(\mathbf{q}) \geq \delta_{SPE,\alpha}^2(\mathbf{q}))] \quad j = 1, ..., J \tag{18}$$

where D_j^s and SPE_j^s are the values of both statistics calculated using $\mathbf{x}_j^{\mathbf{PDL}*}$ and $\delta_{D,\alpha}^2$ and $\delta_{SPE,\alpha}^2$ represent the statistics' critical limits for a significance level equal to α. Regarding the fault isolation constraints, conditions 1 and 2 are incorporated to the SNUP by means of eqs 11 and 17, respectively.

The minimum cost SN that verifies the FDI restrictions is obtained by solving the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{q}} \mathbf{c}^{\mathbf{T}}\mathbf{q} \\ &\text{s.t.} \\ &\quad [(D_j^s \geq \delta_{D,\alpha}^2) \vee (SPE_j^s \geq \delta_{SPE,\alpha}^2)] \quad j = 1, ..., J \\ &\quad iv_{n,j} = sf_{n,j} \quad n = 1, ..., Nj = 1, ..., J \\ &\quad \sum_{n=1}^{N} (\mathbf{iv}_j)_n \geq 1 \quad j = 1, ..., J \\ &\quad \sum_{n=1}^{N} (\mathbf{iv}_j \wedge \neg\mathbf{iv}_k)_n + \sum_{n=1}^{N} (\mathbf{iv}_k \wedge \neg\mathbf{iv}_j)_n \geq 1 \\ &\qquad j = 1, ..., J, k = 1, ..., J, j \neq k \end{aligned} \tag{19}$$

where $\mathbf{c}$ is the vector of sensor costs and $c_i = 0$ if the instrument is already installed in the plant. A mixed integer nonlinear optimization problem results, which comprises J constraints related to the detection of the predetermined set of faults and $[2J + J(J-1)/2]$ restrictions used to verify the isolation of the failures.

For solving that problem, an improved Level Traversal Tree Search algorithm (LTTS)[12,18] is developed. Because faults observability and resolution (FOR) are necessary conditions to achieve their detection and isolation applying monitoring approaches,[8] eq 19 is extended by incorporating those structural restrictions to reduce the computational time. First, let us recall FOR definitions: (a) A failure is observable if at least one sensor points out the existence of the anomalous event. (b) A fault is resolvable if its cause can be distinguished from the source of another fault.

It has proved that FOR are satisfied if the subsequent conditions are fulfilled[6]

$$\begin{aligned} &\sum_{i=1}^{I} (\mathbf{rv}_j)_i \geq 1 \quad j = 1, ..., J \\ &\sum_{i=1}^{I} (\mathbf{rv}_j \wedge \neg\mathbf{rv}_k)_i + \sum_{i=1}^{I} (\mathbf{rv}_k \wedge \neg\mathbf{rv}_j)_i \geq 1 \\ &\quad j = 1, ..., J, j \neq k \end{aligned} \tag{20}$$

where $\mathbf{rv}_j = (\mathbf{a}_j \wedge \mathbf{q})$ is the resolution vector ($\mathbf{rv}$) of the jth failure, and $rv_{i,j} = 1$ means that the ith variable is measured and its value changes when this fault takes place. It should be remarked that FOR conditions can be easily checked using a set of $[J + J(J-1)/2]$ linear inequalities.

If the aforementioned structural constraints are added to eq 19, then the following SNUP results

$$\begin{aligned} &\min_{\mathbf{q}} \mathbf{c}^{\mathbf{T}}\mathbf{q} \\ &\text{s.t.} \\ &\quad \sum_{i=1}^{I} (\mathbf{rv}_j)_i \geq 1 \quad j = 1, ..., J \\ &\quad \sum_{i=1}^{I} (\mathbf{rv}_j \wedge \neg\mathbf{rv}_k)_i + \sum_{i=1}^{I} (\mathbf{rv}_k \wedge \neg\mathbf{rv}_j)_i \geq 1 \\ &\qquad j = 1, ..., J, k = 1, ..., J, j \neq k \\ &\quad [(D_j \geq \delta_{D,\alpha}^2) \vee (SPE_j \geq \delta_{SPE,\alpha}^2)] \quad j = 1, ..., J \\ &\quad iv_{n,j} = sf_{n,j} \quad n = 1, ..., N, j = 1, ..., J \\ &\quad \sum_{n=1}^{N} (\mathbf{iv}_j)_n \geq 1 \quad j = 1, ..., J \\ &\quad \sum_{n=1}^{N} (\mathbf{iv}_j \wedge \neg\mathbf{iv}_k)_n + \sum_{n=1}^{N} (\mathbf{iv}_k \wedge \neg\mathbf{iv}_j)_n \geq 1 \\ &\qquad j = 1, ..., J, k = 1, ..., J, j \neq k \end{aligned} \tag{21}$$

For solving the previous problem, a two-step algorithm is applied. In step 1, mixed integer linear programming is used to obtain a SN, denoted as $\mathbf{q}_1$, which satisfies the FOR restrictions at minimum cost. The initial level of the LTTS for solving the expanded problem and a cost-instrumentation lower limit are fixed using $\mathbf{q}_1$. In step 2, the search starts at the previously set initial level. If the cost of a node is greater than the cost lower limit, then FOR restrictions are evaluated because they are necessary conditions for FDI. If those requirements are verified, then fault detection constraints are tested. If these are satisfied, then the capability of the SN to isolate the J faults is analyzed. This computation sequence allows reducing the total computation time because testing linear constraints is faster than validating the nonlinear ones.

4. APPLICATION RESULTS

Application results of the SNU methodology to the Tennessee Eastman Process (TEP) are discussed in this section. It has been considered that the normal operation and certain faulty states of the process can be sensibly described by the available steady state and dynamic simulation procedures.[19]

In a previous work the TEP flowsheet, the cause effect matrix for a predetermined set of eight faults, the list of variables involved in $\mathbf{A}$ and the sensor costs have been presented. Matrix $\mathbf{A}$ and $\mathbf{x}_j^{\mathbf{PDL}}$ vectors have been defined by applying the procedure outlined in the previous section. The PDL of each variable has been set at twice its maximum deviation in the normal operating data set.[12] Only the first seven failures are involved in this example because the third and eight ones are not structurally resolvable, i.e., $\mathbf{a}_1$ and $\mathbf{a}_8$ are equal.

The first six failures are step changes in composition ratio of the reactants A and C in the stripper feed stream (stream 4), composition of the inert compound B in stream 4, inlet temperature of the reactor cooling water, inlet temperature of the condenser cooling water, loss in the reactor feed stream, and header pressure loss in stream 4 (reduced availability of reactant C) .The next fault is a slow drift in the reaction kinetics. Because of space reasons, the nomenclature of the variables that

DOI: 10.1021/acs.iecr.6b02288
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

Table 1. Results for TEP

constraints	solution for α = 0.05	cost	solution for α = 0.03	cost
O+D	$VF_9VF_{CW,R}T_{CW,R}$	50	$VF_9VF_{CW,R}T_{CW,R}$	50
O+R+D+SVI+I	$VF_{CW,R}P_SF_1F_9T_{CW,R}$	120	$VF_{CW,R}P_SF_1F_9T_{CW,R}$	120

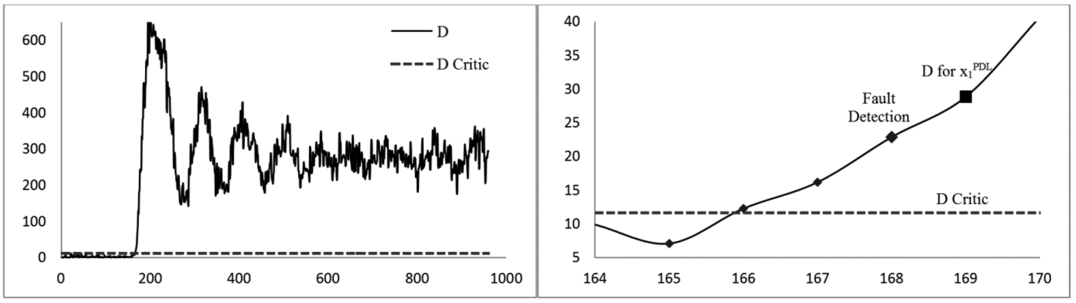


Figure 1. *D* chart: fault 1.

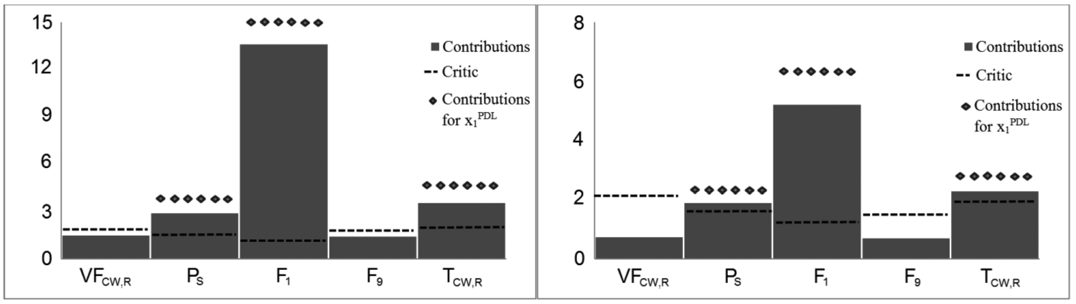


Figure 2. Variable contributions to *D* and SPE: fault 1.

take part in the updated SNs is presented in the Nomenclature section. It is assumed that the set of sensors already installed in the process is the following: VF_1, VF_2, VF_3, VF_4, VF_9, $VF_{R,C}$, and $VF_{CW,R}$; consequently, the costs of these instruments are set equal to zero.

An enhanced LTTS programmed in Matlab is employed to solve different SNUPs. The number of retained PCs is calculated by setting the percentage of the total variance that the PCA model explains at 80%. Also two values of α are used (0.05; 0.03). Problem solutions are displayed in Table 1. The first one satisfies faults observability and detection constraints, denoted as O+D.[12] The second one guarantees faults observability, resolution, detection, suspicious variable identification, and isolation restrictions, indicated as O+R+D+SVI+I, i.e., all those involved in eq 21.

The optimization problem with O+D requirements includes 14 constraints, such that half of them are nonlinear ones. With respect to the upgrade problem that satisfies O+R+D+SVI+I requisites, it comprises 28 linear restrictions used to verify FOR conditions and 63 nonlinear constraints related to D+SVI+I requirements. Both problems involve 34 binary variables.

The analysis of the results presented in Table 1 points out that the solutions are not sensitive to the value of α. Furthermore, it can be seen that the SN which satisfies O+D requisites does not guarantee the isolation of the seven faults. In consequence, the solution of the second problem comprises more instruments, and it is more expensive than the corresponding one to the first location problem.

The results are validated to prove that the obtained SN is able to detect and isolate the proposed failures while the process is still operating in a safe mode. For illustrative purposes, the time series of *D* values, which arise if fault 1 takes place and the measured variables are $VF_{CW,R}P_SF_1F_9T_{CW,R}$, is displayed on the left-hand part of Figure 1. In this case *D* detects the failure before *SPE* does; consequently, only the *D* graph is shown. The previous diagram is amplified around the fault-detection time (right-hand part of Figure 1), and it is observed that the solution of the SNUP subject to O+R+D+SVI+I requirements satisfies fault 1 detection before one or more variables reach their PDLs.

Table 2. Fault Subspace and Identification Vectors

fault	$\mathbf{sf}_j$	$\mathbf{iv}_j$
F_1	$P_SF_1T_{CW,R}$	$P_SF_1T_{CW,R}$
F_2	F_9	F_9
F_3	$VF_{CW,R}$	$VF_{CW,R}$
F_4	$T_{CW,R}$	$T_{CW,R}$
F_5	$F_1T_{CW,R}$	$F_1T_{CW,R}$
F_6	$VF_{CW,R}P_ST_{CW,R}$	$VF_{CW,R}P_ST_{CW,R}$
F_7	P_SF_9	P_SF_9

In Figure 2, the VCs to *D* (left-hand side) and *SPE* (right-hand side), which are obtained if fault 1 has been detected, are shown. The sequence of variables is the same used for the solution presented in Table 1.

It can be seen that the contributions of variables $P_SF_1T_{CW,R}$ to the statistics exceed their control limits. Therefore, they are identified as suspicious variables, and vector $\mathbf{iv}_1 = [P_SF_1T_{CW,R}]$. Also it should be noticed that those contributions are lower than the VCs values corresponding to $\mathbf{x}_1^{\mathbf{PDL}*}$.

Table 2 includes the subspace and identification vectors for all the faults when the variables $VF_{CW,R}P_SF_1F_9T_{CW,R}$ are measured. It can be seen that $\mathbf{iv}_1$ is equal to $\mathbf{sf}_1$, that is, condition 1 is satisfied for fault 1. Furthermore, all the faults are distinguishable because all the $\mathbf{iv}$'s are different, that is, the solution of the SNUP verifies condition 2.

DOI: 10.1021/acs.iecr.6b02288
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

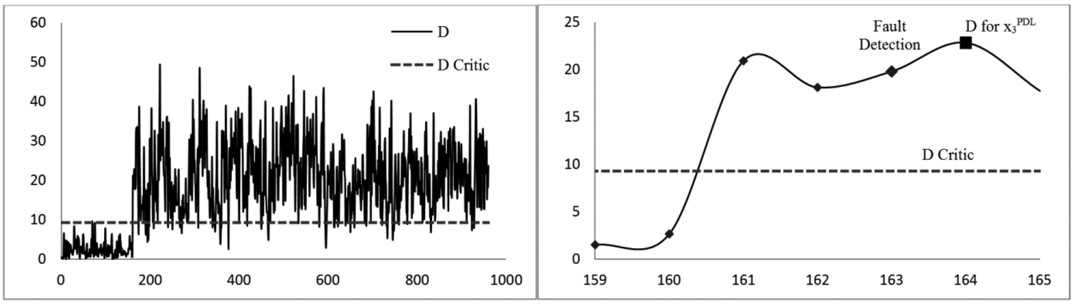


Figure 3. *D* chart: fault 3.

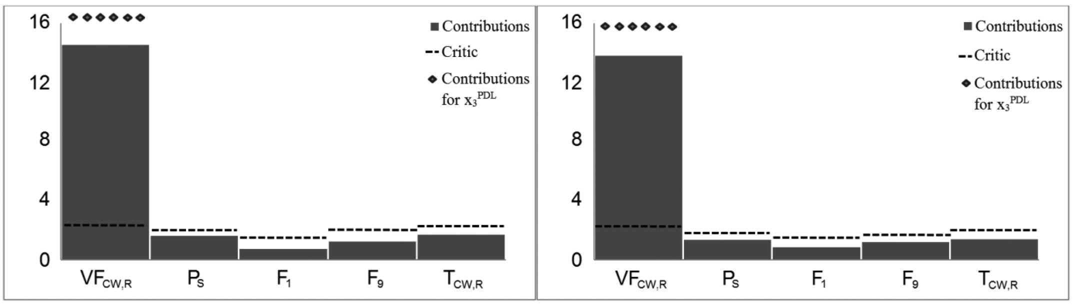


Figure 4. Variable contributions to *D* and SPE: fault 3.

Figures 3 and 4 represent the time series of *D* and the VCs values to both statistics when fault 3 happens. Based on the previous information, results show that fault 3 can be correctly isolated.

5. CONCLUSIONS

In this paper a new strategy devoted to optimally select the set of instruments to be installed in an existing process with fault diagnosis purposes is presented. Problem solution guarantees the detection and isolation of a predetermined set of failures by applying the classic PCA before process variables exceed certain deviation limits. Fault detection is satisfied using restrictions straightaway defined in terms of the PCA's statistics. Regarding failures isolation, constraints are formulated as a function of the VCs to the inflated statistics. These are commonly used to determine the subset of measurements that reveal a fault when statistical monitoring techniques are applied online, but they have not been incorporated to the SNUP to enhance fault diagnosis until now.

Two conditions are stated to guarantee that the updated SN is capable to isolate the predetermined set of faults. The first one considers that the set of measurements whose contributions exceed their control limits ($\mathbf{iv}_j$) should be equal to the group of observations affected by the fault occurrence ($\mathbf{sf}_j$). The second one takes into account that all the **iv**'s should be different. This condition is formulated using the concept of the NNV.[6]

The solution of the proposed optimization problems is carried out using an improved LTTS. Given that fault observability is a necessary condition for fault detection and that fault resolution has the same role for its isolation, FOR constraints are incorporated to the SNUP because they can be posed as linear inequalities. Therefore, the computational load of the expanded problem is lower than the corresponding one to the original formulation.

Application results of the strategy to the TEP are presented and also validated to prove the ability of the obtained SNs to detect and isolate the proposed failures while the process is still operating in a safe mode. Even though this research is focused on the classic PCA procedure, the proposed upgrade methodology can be straightway employed for processes monitored using other statistical strategies such that the CVs to the inflated statistic are explicitly calculated as a function of the measurement vector.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +54 291 486 1700. E-mail: msanchez@plapiqui.edu.ar.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the financial support of CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), and UNS (Universidad Nacional del Sur, Bahía Blanca, Argentina).

■ NOMENCLATURE

VF_1 = A feed flow, manipulated variable (stream 1)
VF_2 = D feed flow, manipulated variable (stream 2)
VF_3 = E feed flow, manipulated variable (stream 3)
VF_4 = total feed flow, manipulated variable (stream 4)
VF_9 = purge valve, manipulated variable (stream 9)
$VF_{R,C}$ = compressor recycle valve, manipulated variable
$VF_{CW,R}$ = reactor cooling water flow, manipulated variable
P_S = stripper pressure
F_1 = A feed (stream 1)
F_9 = purge rate (stream 9)
$T_{CW,R}$ = reactor cooling water outlet temperature
$y_{A,4}$ = molar fraction of A in stream 4
$y_{B,4}$ = molar fraction of B in stream 4
$y_{C,4}$ = molar fraction of C in stream 4
$T_{CW,C}$ = condenser cooling water outlet temperature

DOI: 10.1021/acs.iecr.6b02288
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX

REFERENCES

(1) Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S.; Yin, K. A review of process fault detection and diagnosis Part III: Process History Based Methods. *Comput. Chem. Eng.* **2003**, *27*, 327.

(2) MacGregor, J.; Yu, H.; García Muñoz, S.; Flores-Cerrillo, J. Data-based latent variable methods for process analysis, monitoring and control. *Comput. Chem. Eng.* **2005**, *29*, 1217.

(3) Ge, Z.; Song, Z.; Gao, F. Review of Recent Research on Data-Based Process Monitoring. *Ind. Eng. Chem. Res.* **2013**, *52*, 3543.

(4) Bagajewicz, M. *Design and Upgrade of Process Plant Instrumentation*; Technomic Publishing Company: Lancaster, PA, 2000.

(5) Li, F. Dynamic Modeling, Sensor Placement Design, and Fault Diagnosis of Nuclear Desalination Systems. Ph.D. Thesis, University of Tennessee, Knoxville, TN, May 2011.

(6) Rodriguez, L.; Cedeño, M.; Sánchez, M. A Structural Approach to Design Sensor Networks for Fault Diagnosis. *Ind. Eng. Chem. Res.* **2013**, *52*, 17941.

(7) Mobed, P.; Maddala, J.; Pednekar, P.; Bhattacharyya, D.; Rengaswamy, R. Optimal Sensor Placement for Fault Diagnosis Using Magnitude Ratio. *Ind. Eng. Chem. Res.* **2015**, *54*, 9369.

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

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

(10) Bhushan, M.; Rengaswamy, R. Comprehensive design of a sensor network for chemical plants based on various diagnosability and reliability criteria. 1. Framework. *Ind. Eng. Chem. Res.* **2002**, *41*, 1826.

(11) Musulin, E.; Bagajewicz, M.; Nougués, J.; Puigjaner, L. Instrumentation Design and Upgrade for Principal Components Analysis Monitoring. *Ind. Eng. Chem. Res.* **2004**, *43*, 2150.

(12) Rodriguez, L.; Cedeño, M.; Sánchez, M. Optimal Sensor Network Upgrade for Fault Detection using Principal Component Analysis. *Ind. Eng. Chem. Res.* **2016**, *55*, 2359.

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

(14) Ramaker, H.; Van Sprang, E.; Westerhuis, J.; Smilde, A. K. Fault Detection Properties of Global, Local and Time Evolving Models for Batch Process Monitoring. *J. Process Control* **2005**, *15*, 799.

(15) Westerhuis, J.; Gurden, S.; Smilde, A. Generalized Contribution Plots in Multivariate Statistical Process Monitoring. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 95.

(16) Box, G. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. *Ann. Math. Stat.* **1954**, *25*, 290.

(17) Alcala, C.; Qin, S. Reconstruction-Based Contribution for Process Monitoring. *Automatica.* **2009**, *45*, 1593.

(18) Nguyen, D.; Bagajewicz, M. New efficient breadth first/level traversal tree search method for the design and upgrade of sensor networks. *AIChE J.* **2011**, *57*, 1302.

(19) Russell, E.; Chiang, L.; Braatz, R. *Data-Driven Methods for Fault Detection and Diagnosis in Chemical Processes*; Springer-Verlag: London, 2000.

DOI: 10.1021/acs.iecr.6b02288
Ind. Eng. Chem. Res. XXXX, XXX, XXX−XXX