An improved PCA scheme for sensor FDI: Application to an air quality monitoring network

Article *in* Journal of Process Control · July 2006

DOI: 10.1016/j.jprocont.2005.09.007

3 authors:

Mohamed-Faouzi HARKAT
Badji Mokhtar - Annaba University

Gilles Mourot
University of Lorraine

José Ragot
University of Lorraine

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Process Control 16 (2006) 625–634

JOURNAL OF PROCESS CONTROL

www.elsevier.com/locate/jprocont

An improved PCA scheme for sensor FDI: Application to an air quality monitoring network

Mohamed-Faouzi Harkat [a,*], Gilles Mourot [b], José Ragot [b]

[a] Université Badji Mokhtar, ANNABA, Faculté des Sciences de l'Ingénieur, Département d'Électronique, BP. 12, Sidi Amar, 23000 Annaba, Algérie

[b] Institut National Polytechnique de Lorraine, Centre de Recherche en Automatique de Nancy, UMR-CNRS INPL/UHP 7039, 2, Avenue de la Forêt de Haye 54516 Vandoeuvre-lès-Nancy, France

Received 9 January 2005; received in revised form 13 July 2005; accepted 26 September 2005

Abstract

In this paper a sensor fault detection and isolation procedure based on principal component analysis (PCA) is proposed to monitor an air quality monitoring network. The PCA model of the network is optimal with respect to a reconstruction error criterion. The sensor fault detection is carried out in various residual subspaces using a new detection index. For our application, this index improves the performance compared to classical detection index SPE. The reconstruction approach allows, on one hand, to isolate the faulty sensors and, on the other hand, to estimate the fault amplitudes.

Keywords: Fault diagnosis; Principal component analysis; Sensor failure detection and isolation; Variable reconstruction; Air quality monitoring network

1. Introduction

Many human activities produce primary pollutants like nitrogen oxides (NO_2 and NO) and volatile organic compounds (VOC) which formed in the lower atmosphere, by chemical or photochemical reactions, secondary pollutants like ozone. The acceptable concentrations of these pollutants, harmful for human health and the environment, are defined by European standards. Air quality monitoring networks have the following main missions: the measurement network management (recording of pollutant concentrations and a range of meteorological parameters related to pollution events) and the diffusion of data for permanent information of population and public authorities in reference to norms. To ensure these missions, the validity of the delivered information is essential before any use of measurements, mainly those of pollutants. Moreover, the geographical area monitored by a network being large, fault diagnosis procedure also enables to optimize the maintenance actions. Therefore sensor validation is an issue of great importance for the development of reliable environmental monitoring and management systems. In collaboration with AIRLOR, an air quality monitoring network (France), the aim of this work is to develop a method to perform sensor validation, mainly sensors measuring the concentrations of ozone and nitrogen oxides.

The task for sensor validation is to detect, isolate, and identify faulty sensors by examining the sensor measurements. The sensor validation is usually performed either using "outlier" detection methods which only enable to identify extreme values out of measurement range or manually by an operator. Unfortunately, this latter approach is too subjective and often impractical in real-time due to high process dimensionality which produces a large amount of collected data.

* Corresponding author.
E-mail address: Mohamed-faouzi.Harkat@ensem.inpl-nancy.fr (M.-F. Harkat).

0959-1524/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jprocont.2005.09.007

However, the availability of many sensors in the network provides valuable redundancy for fault detection and isolation because some sensor measurements are highly correlated under normal conditions. The analytical redundancy approach consists of checking the consistency between the measurements and the estimates provided by the relationships existing between the various variables of the process [3]. This analysis may lead to detect and isolate the faulty sensors. In most practical situations, fault diagnosis needs to be performed in the presence of disturbances, noise and modelling errors. These mathematical relations between the plant variables generally take two forms [4]. Analytical redundancy methods used explicit input–output model usually derived from system identification. However, in the considered sensor network, this explicit formulation of the redundancy relationships may be difficult to obtain (owing to complexity of the process and high process dimensionality) because it must be guided mainly by performance criteria of the diagnosis system. As an alternative, implicit modelling approaches, which are data-driven techniques (like principal component analysis), are particularly well adapted to reveal linear relationships among the plant variables without formulating them explicitly. Moreover the number of these relations could be determined by minimizing a criterion based on the best reconstruction of the variables with respect to the number of principal components in the PCA model [12]. PCA has some other nice features. It can handle high dimensional and correlated process variables, provides a natural solution to the errors-in-variables problem and includes disturbance decoupling [4].

The widely used detection index SPE (squared prediction error) indicates how much each sample deviates from the PCA model. Indeed SPE performs fault detection in the the residual space. However, Harkat et al. [6] have shown experimentally on real data that the SPE is sensitive to the modelling errors. Indeed modelling errors could be projected into the residual space which results in residuals with higher variance than the others. Then the SPE will be heavily biased in favour of those equations with the largest residual variance whereas the residuals with the smallest residual variances are most useful for sensor fault diagnosis because they are associated with linear relationships.

After a fault has been detected, it is important to isolate faulty sensor. In the PCA framework, the well known isolation approaches are residual enhancement, contribution plots and variable reconstruction methods [13]. The residual enhancement technique generates structured residuals that selectively respond to subsets of faults [5]. Such residuals may be obtained by algebraic transformation, or by a direct technique that involves a bank of partial PCA models [4]. However, for high dimensionality process, it is not always possible to find the algebraic transformation that enables to obtain the desired isolation properties because these properties are only defined according to the occurrence of the faults in the residuals without taking into account the sensitivities of the residuals to the faults. These comments are also true for the direct method. For fault isolation contribution plots show the contribution of each process variable to the detection statistic [8]. It is assumed that the process variable with the highest contribution is likely the faulty one. However, the contribution plots may not explicitly identify the cause of an abnormal condition [9], and sometimes lead to incorrect conclusions [13]. An alternative approach for fault isolation is the variable reconstruction method proposed by Dunia et al. [2]. It is based on the idea that the influence of fault on the detection index is eliminated when the faulty variable is reconstructed using the PCA model from the variables without defect.

In this paper, a sensor fault detection and isolation procedure based on principal component analysis is proposed to monitor an air quality monitoring network. Taking into account the high redundancy among the process variables, this procedure is based on the variable reconstruction approach in order to design the PCA model of the network, to isolate the faulty sensors and to estimate the fault amplitudes. To improve fault detection, a new fault detection index is proposed which monitors the last principal components (which have smallest variances) in different residual subspaces, starting with a single direction (the last principal direction) and gradually increasing the dimension of the residual to the full residual space. Mathematically, it is difficult to prove that the new index has better fault sensitivity than the SPE because they perform fault detection in different residual spaces where the sensitivities of the residuals to the faults may be different. Therefore experimental data set was used to evaluate the performance of the proposed index. For our application, this index improves the fault detection compared to classical detection index SPE.

First, we present the principle of PCA modelling based on the variable reconstruction approach. Then, after having summarized the principle of sensor fault detection in the PCA framework, we propose our new detection index. For isolation of the faulty sensors, we combine the proposed index and the variable reconstruction principle. In Section 4, we present the application of the proposed method to sensor fault detection and isolation of an air quality monitoring network in Lorraine. Conclusions and future works are finally presented in the last section.

2. PCA fault detection and isolation

2.1. PCA modelling

Principal component analysis is one of the most popular statistical methods, for extracting information from measured data, which finds the directions of significant variability in the data by forming linear combinations of variables.

Let us consider $\mathbf{x}(k) = [x_1(k) \quad x_2(k) \quad \cdots \quad x_m(k)^{\mathrm{T}}]$ the vector formed with m observed plant variables at time instant k. Define the data matrix $X = [\mathbf{x}(1) \quad \mathbf{x}(2) \quad \cdots \quad \mathbf{x}(N)]^{\mathrm{T}} \in \Re^{N\times m}$ with N samples $\mathbf{x}(k)$

$(k = 1, \ldots, N)$ which is representative of normal process operation.

PCA determines an optimal linear transformation of the data matrix X in terms of capturing the variation in the data:

$$T = XP \quad \text{and} \quad X = TP^{\mathrm{T}} \tag{1}$$

with $T = [\,t_1 \quad t_2 \quad \cdots \quad t_m\,] \in \Re^{N\times m}$, where the vectors t_i are called scores or principal components and the matrix $P = [\,p_1 \quad p_2 \quad \cdots \quad p_m\,] \in \Re^{m\times m}$, where the orthogonal vectors p_i, called loading or principal vectors, are the eigenvectors associated to the eigenvalues λ_i of the covariance matrix (or correlation matrix) Σ of X:

$$\Sigma = P\Lambda P^{\mathrm{T}} \quad \text{with } PP^{\mathrm{T}} = P^{\mathrm{T}}P = I_m \tag{2}$$

where $\Lambda = \mathrm{diag}(\lambda_1 \cdots \lambda_m)$ is a diagonal matrix with diagonal elements in decreasing magnitude order.

The partition of eigenvectors and principal components matrices gives:

$$P = \left[\hat{P}_\ell | \tilde{P}_{m-\ell}\right], \quad T = \left[\hat{T}_\ell | \tilde{T}_{m-\ell}\right] \tag{3}$$

where ℓ will be defined below.

Eq. (1) can be rewritten as

$$X = \hat{T}_\ell \hat{P}_\ell^{\mathrm{T}} + \tilde{T}_{m-\ell}\tilde{P}_{m-\ell}^{\mathrm{T}} = \hat{X} + E \tag{4}$$

with

$$\hat{X} = X\hat{C}_\ell \quad \text{and} \quad E = X\tilde{C}_{m-\ell} \tag{5}$$

where the matrices $\hat{C}_\ell = \hat{P}_\ell \hat{P}_\ell^{\mathrm{T}}$ and $\tilde{C}_{m-\ell} = I_m - \hat{C}_\ell$ constitute the PCA model.

The matrices $\hat{X}$ and E represent, respectively, the modelled variations and nonmodelled variations of X based on ℓ components ($\ell < m$). The first ℓ eigenvectors $\hat{P}_\ell \in \Re^{m\times\ell}$ constitute the representation space whereas the last $(m-\ell)$ eigenvectors $\tilde{P}_{m-\ell} \in \Re^{m\times m-\ell}$ constitute the residual space.

The identification of the PCA model thus consists in estimating its parameters by an eigenvalue/eigenvector decomposition of the matrix Σ and determining the number of principal components ℓ to retain. Here PCA is used for extracting redundancy relationships between the variables. In most practical cases (noisy measurements), the small eigenvalues indicate the existence of linear or quasilinear relations among the process variables. However the distinction between significant or not significant eigenvalues may not be obvious due to modelling errors (disturbances and nonlinearities) and noise. A key issue to develop a PCA model is to choose the adequate number of principal components.

2.1.1. Determining the number of principal components

Most methods to determine the number of principal components are rather subjective in the general practice of PCA [14]. However, the number of principal components has a significant impact on each step of the sensor fault detection, isolation and reconstruction scheme, such as the ability to detect small faults, the degree of freedom for fault identification and the accuracy of reconstruction.

Qin and Dunia [12] proposed to determine ℓ by minimization of the variance of reconstruction error. The variable reconstruction consists in estimating a variable from others plant variables using the PCA model, i.e. using the redundancy relations between this variable and the others. The reconstruction accuracy is thus related to the capacity of the PCA model to reveal the redundancy relations among the variables, i.e. to the number of principal components. Indeed, if too many principal components are chosen, each variable tends to rely too much on itself, which means its relationship to other variables is weakened. If too few principal components are used, the model is inaccurate to represent the normal variation of the data and thus results in poor reconstruction.

First, the variable reconstruction approach is presented. Let us denote $\mathbf{x}_j(k)$ the measurement vector where the jth component is reconstructed. The reconstruction of the jth variable, while taking into account the reconstruction condition $c_{jj} \neq 1$, is

$$\mathbf{x}_j(k) = G_j\mathbf{x}(k) \tag{6}$$

with

$$G_j^{\mathrm{T}} = [\,\xi_1 \quad \cdots \quad g_j \quad \cdots \quad \xi_m\,] \quad g_j^{\mathrm{T}} = \frac{[\,c_{-j}^{\mathrm{T}} \quad 0 \quad c_{+j}^{\mathrm{T}}\,]}{1 - c_{jj}} \tag{7}$$

where $\xi_j = [\,0 \quad \cdots \quad 1 \quad \cdots \quad 0\,]^{\mathrm{T}}$ is the the jth column of an identity matrix, $c_j^{\mathrm{T}} = [\,c_{1j} \quad c_{2j} \quad \cdots \quad c_{mj}\,] = [\,c_{-j}^{\mathrm{T}} \quad c_{jj} \quad c_{+j}^{\mathrm{T}}\,]$ is the jth column of $\hat{C}_\ell$ and the subscripts $-j$, $+j$, respectively, allow to denote a vector formed by the first $(j-1)$ and the last $(m-j)$ elements of the considered vector.

The variance of the reconstruction error, in the direction ξ_j, is defined as

$$\rho_j(\ell) = \frac{\tilde{\xi}_j^{\mathrm{T}}\Sigma\tilde{\xi}_j}{\left(\tilde{\xi}_j^{\mathrm{T}}\tilde{\xi}_j\right)^2} \tag{8}$$

where $\tilde{\xi}_j = (I - \hat{C}_\ell)\xi_j$.

To find the number of principal components, the normalized VRE (variance of reconstruction error) is minimized with respect to the number ℓ:

$$\mathrm{VRE}(\ell) = \sum_{j=1}^{m} \frac{\rho_j(\ell)}{\xi_j^{\mathrm{T}}\Sigma\xi_j} \tag{9}$$

An important feature of this approach is that the proposed criterion has a minimum corresponding to the best reconstruction [12]. Moreover these authors proposed to discard from process monitoring the variables for which the individual variances of reconstruction error are greater than their variances. Indeed variables which are little correlated with others cannot be reliably reconstructed from the PCA model.

2.2. Fault detection

After the PCA model has been built, we now examine its use for sensor fault detection.

2.2.1. Residual generation

Let us consider a new sample vector, $\mathbf{x}(k)$. Following decomposition (4), this vector can also be represented as

$$\mathbf{x}(k) = \hat{\mathbf{x}}(k) + \mathbf{e}(k) \tag{10}$$

where $\hat{\mathbf{x}}(k) = \hat{C}_\ell \mathbf{x}(k)$ is the estimation vector, $\mathbf{e}(k) = (I - \hat{C}_\ell)\mathbf{x}(k)$ is the vector of estimation errors.

The principal component vector is given by

$$\mathbf{t}(k) = P^{\mathrm{T}}\mathbf{x}(k) = [\,\hat{\mathbf{t}}_\ell(k) \quad \tilde{\mathbf{t}}_{m-\ell}(k)\,] \tag{11}$$

where

$$\hat{\mathbf{t}}_\ell(k) = \hat{P}_\ell^{\mathrm{T}}\mathbf{x}(k), \quad \tilde{\mathbf{t}}_{m-\ell}(k) = \tilde{P}_{m-\ell}^{\mathrm{T}}\mathbf{x}(k) \tag{12}$$

There is an equivalence between the residual vector and the last principal component vector $\tilde{\mathbf{t}}_{m-\ell}$:

$$\mathbf{e}(k) = \tilde{P}_{m-\ell}\tilde{\mathbf{t}}_{m-\ell}(k) \tag{13}$$

So, it becomes simpler to work with the residual vector $\tilde{\mathbf{t}}_{m-\ell}$ with dimension $(m-\ell)$.

Let us consider now the fault propagation on the residual vector $\tilde{\mathbf{t}}_{m-\ell}$. In the presence of fault with a magnitude $\mathbf{f}(k)$, the measurement vector $\mathbf{x}(k)$ can be expressed as

$$\mathbf{x}(k) = \mathbf{x}^\circ(k) + \epsilon(k) + \xi_j\mathbf{f}(k) \tag{14}$$

where $\mathbf{x}^\circ(k)$ is the fault-free measurement vector, $\epsilon(k)$ is the white measurement noise vector and ξ_j is the fault direction (fault on the jth component of $\mathbf{x}(k)$).

From (12) and (14), the residual vector is given by

$$\tilde{\mathbf{t}}_{m-\ell}(k) = \underbrace{\tilde{P}_{m-\ell}^{\mathrm{T}}\mathbf{x}^\circ(k)}_{=0} + \tilde{P}_{m-\ell}^{\mathrm{T}}\epsilon(k) + \tilde{P}_{m-\ell}^{\mathrm{T}}\xi_j\mathbf{f}(k) \tag{15}$$

In the fault-free case, the expectation of the residual vector is zero. In the presence of faults, the expectation of the residual vector is no more zero and the fault affects all the components of residual vector.

2.2.2. Definition of the detection index D_i

Sensor fault detection using PCA is performed by monitoring the residuals. The SPE (squared prediction error) is a statistic that measures the lack of fit of the PCA model to the data. At time k, the detection index SPE is given by

$$\mathrm{SPE}(k) = \tilde{\mathbf{t}}_{m-\ell}(k)^{\mathrm{T}}\tilde{\mathbf{t}}_{m-\ell}(k) = \sum_{j=\ell+1}^{m} t_j^2(k) \tag{16}$$

This quantity suggests the existence of an abnormal situation in the data when:

$$\mathrm{SPE}(k) > \delta_\alpha^2 \tag{17}$$

where δ_α^2 is a control limit for SPE given by [1] or estimated using the historical data.

The SPE is formed by summing the squares of residuals obtained from PCA model. However modelling errors could be projected into the residual space which results in residuals with higher variance than the others. Then the SPE will be heavily biased in favour of those equations with the largest residual variance whereas the residuals with the smallest residual variances are most useful for sensor fault diagnosis because they are associated with linear relationships.

To overcome this problem, a new index based on the last principal components is suggested and given, at time k, by the following expression:

$$D_i(k) = \tilde{\mathbf{t}}_i(k)^{\mathrm{T}}\tilde{\mathbf{t}}_i(k) = \sum_{j=m-i+1}^{m} t_j^2(k) \quad \text{with } i = 1, 2, \ldots, (m-\ell) \tag{18}$$

This index is calculated by using successive sum of squared last principal components in different subspaces of the residual space, starting with a single direction (the last principal direction associated to the smallest variance) and gradually increasing the dimension of the residual to the full residual space. Therefore, the index D_i may be less sensitive to the modelling errors than the SPE. However, mathematically, it is difficult to prove that the new index has better fault sensitivity than the SPE because they perform fault detection in different residual spaces where the sensitivities of the residuals to the faults may be different.

Considering the definitions (16) and (18), the expression of the index D_i can be written as follows:

$$D_i(k) = \mathrm{SPE}_i(k) \tag{19}$$

where $\mathrm{SPE}_i(k)$ is a squared prediction error calculated by retaining $(m-i)$ components in the PCA model. The control limits $\delta_{i,\alpha}^2$ for the detection index D_i with a significance level α, can be calculated as in the case of SPE [1].

The process is considered out of control limit if:

$$D_i(k) > \delta_{i,\alpha}^2 \quad i = 1, 2, \ldots, (m-\ell) \tag{20}$$

To improve quality of detection by reducing the rate of false alarms (due to noise), filter EWMA (exponentially weighted moving average) can be applied to the residuals. The filtered residuals are thus obtained:

$$\bar{\tilde{\mathbf{t}}}_i(k) = (1-\gamma)\bar{\tilde{\mathbf{t}}}_i(k-1) + \gamma\tilde{\mathbf{t}}_i(k) \tag{21}$$

$\bar{\tilde{\mathbf{t}}}_i(k)$ being the filtered vector at time k and γ is a forgetting factor chosen in the interval $[0, 1]$.

The control limits $\bar{\delta}_{i,\alpha}^2$ for $\overline{D}_i$ (filtered D_i) are computed as proposed by Qin et al. [11] for the index $\overline{\mathrm{SPE}}$ (filtered SPE).

2.3. Fault isolation

After the presence of a fault has been detected, it is important to identify the fault and apply the necessary corrective actions to eliminate the abnormal data.

The variable reconstruction approach for sensor fault isolation was proposed by Dunia et al. [2]. This approach assumes that each sensor may be faulty (in the case of a single fault) and suggests to reconstruct the assumed faulty sensor using the PCA model from the remaining measurements. By examining the residuals given by PCA model before and after reconstruction, we can determine the faulty sensor.

$\overline{D}_i^{(j)}(k)$ denotes the index $\overline{D}_i(k)$ calculated after reconstruction of the jth variable. To show the propagation of a fault $\mathbf{f}(k)$ in the direction ξ_f on the index $\overline{D}_i^{(j)}$, let us consider the expression of the measurement vector $\mathbf{x}_j(k)$ (6):

$$\mathbf{x}_j(k) = G_j(\mathbf{x}^\circ(k) + \epsilon(k)) + G_j\xi_f\mathbf{f}(k) \tag{22}$$

If we consider the case when the faulty variable is the reconstructed one ($j = f$), then:

$$G_j\xi_f = 0 \tag{23}$$

Therefore, if the faulty variable is reconstructed ($j = f$), the index $\overline{D}_i^{(j)}$ is in the control limit because the fault $\mathbf{f}$ is eliminated (23). If the reconstructed variable is not faulty ($j \neq f$), the index $\overline{D}_i^{(j)}$ being always affected by the fault $\mathbf{f}$, $\overline{D}_i^{(j)}$ is outside its control limit. In summary, when a fault has been detected, all the indices $\overline{D}_i^{(j)}$ are computed, and if $\overline{D}_i^{(j)} \leqslant \bar{\delta}_{i,\alpha}^2$, the jth sensor is considered as a faulty one.

From the above analysis, we can define an identification index based on $\overline{D}_i^{(j)}$, ($j = 1,\ldots,m$) as follows:

$$A_{\overline{D}_i^{(j)}}(k) = \frac{\overline{D}_i^{(j)}(k)}{\bar{\delta}_{i,\alpha}^2} \tag{24}$$

Thus at time k, if the faulty variable is the reconstructed variable, then $A_{\overline{D}_i^{(j)}}(k)$ is lower than one. If the reconstructed variable is not faulty $A_{\overline{D}_i^{(j)}}(k)$ will be larger than one. So, the jth variable for which $A_{\overline{D}_i^{(j)}}(k)$ is lower than one, is the faulty variable.

3. Proposed sensor FDI scheme

The aim of the proposed FDI (Fault Detection and Isolation) scheme is to detect the faulty sensors using the index $\overline{D}_i$ and to identify the faulty sensor using the variable reconstruction approach. The measurement delivered by the faulty sensor is then reconstructed using Eq. (6).

The algorithm for implementing the proposed fault detection and isolation scheme is as follows:

(i) Perform a standard PCA on the data matrix X; determine the model by a proper selection of the number of PC.

(ii) Calculate $\overline{D}_i (i = 1,\ldots,m-\ell)$ and determine the control limit $\bar{\delta}_{i,\alpha}^2$ for each index under normal conditions.

(iii) For each observation $\mathbf{x}(k)$:
  (a) Select the residual subspace, $i = 1$.
  (b) Calculate $\overline{D}_i(k)$.
  if $\overline{D}_i(k) > \bar{\delta}_{i,\alpha}^2$, a fault is detected, go to (c),
  else $i = i + 1$, go to (b),
  repeat until $i = m - \ell$.
  (c) Identify the faulty variable:
  - The index $A_{\overline{D}_i^{(j)}}(k)$, calculated after the reconstruction of the jth variable, which is lower than 1 indicates that this variable is the faulty one.
  - Suggest a replacement value for the faulty measurement (estimation of the fault amplitude) by reconstructing the faulty variable.

4. Application to an air quality monitoring network

In this section the proposed FDI scheme is applied to sensor fault detection, isolation and reconstruction of air quality monitoring network.

4.1. Problem settings

The air quality monitoring network AIRLOR working in Lorraine, France, consists of more than twenty stations placed in rural, peri-urban and urban sites. Each station consists of a set of sensors, dedicated to the acquisition of the following pollutants: carbon monoxide CO, nitrogen oxides NO and NO_2, sulfur dioxide SO_2 and ozone O_3. Moreover, seven stations are dedicated to the acquisition of additional meteorological parameters. The measures are averages calculated over fifteen minutes in order to limit spatial and temporal sampling problems.

The purpose is to detect functioning abnormalities of the sensors principally those of the ozone concentration (O_3) and nitrogen oxides (NO and NO_2). First, let us describe in some words the phenomenon. Ozone is a secondary pollutant produced by complex photochemical reactions between primary pollutants (mainly NO, NO_2 and VOC) emitted into the atmosphere. Their concentrations also depend highly on the vertical and horizontal movements of the atmosphere that are linked to the meteorological conditions. On one hand, ozone is a secondary pollutant whose spatial distribution of the maximum values is rather homogeneous at our local scale. On the other hand, the nitrogen oxides are primary pollutants which are more localized because their concentrations depend directly on the sources of emissions. PCA can handle the high dimension of the measurement network and the high degree of correlation among some variables. However, we can wonder about the performances of PCA since the phenomenon of photochemical pollution exhibits a nonlinear and time-varying behaviour.

In this paper, only six neighbour measurement stations are considered. The matrix X contains 18 variables, v_1 to v_{18}, corresponding, respectively, to ozone O_3 and nitrogen oxides (NO_2 and NO) collected on these six stations. It should be noted here that the paper presents a feasibility study. Thus short period of measurements are only considered here and we simulated sensor faults whose magnitude is chosen according to measurement uncertainties provided by an expert of AIRLOR (approximately 15% of measurement for ozone). However, in the considered period, we have weak, average and high levels of pollutants (Figs. 2–4). Here, only 1080 samples are used: the first 800 samples to elaborate the PCA model and the last samples 801–1080 to test the proposed fault detection and isolation scheme.

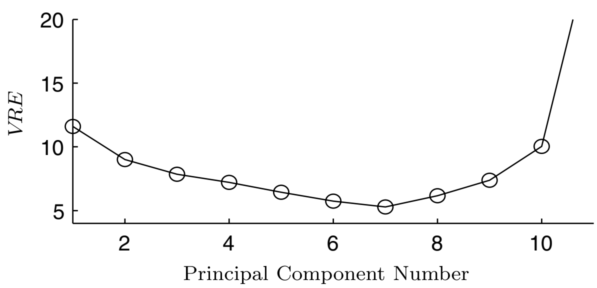

Fig. 1. Variances of the reconstruction error for the number of components in the PCA model.

4.2. PCA modelling

First, the correlation matrix of data matrix X was calculated and an analysis of theses correlations was carried out. Only measurements of the ozone concentrations from the various sites have significant coefficients of correlation (higher than 85%) whereas measurements of NO_x levels from the various sites are poorly correlated. These first results are in agreement with our remarks on the phenomenology of these pollutants.

A preliminary analysis of the data enabled us to note that the relations between the variables could be approximated by static linear relations. The analysis of eigenvalues of the data correlation matrix and the cumulative percentages of correlation explained by each eigenvalue did not make it possible to determine the number of component to be retained because the eigenvalues have an exponential decrease. Fig. 1 shows the evolution of the variance of the reconstruction error according to the number of principal components. This curve having a minimum for $\ell = 7$, a PCA model with seven components was retained with a percentage of cumulative correlation equal to 91%.

In Table 1, the variances of the reconstruction error using seven principal components for each variable are indicated. It should be noted that all variables can be reconstructed because their respective variance is lower than 1. However, variables 9, 11, 12 and 15 having coefficients higher than the other variables (in bold in the table), some detectability difficulties may appear for these variables. These variables are the concentrations of NO and NO_2 on urban sites or peri-urban. Broadly, the ozone con-

Table 1
Variances of the reconstruction error for the variables

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\ell = 7$	0.1	0.1	0.2	0.1	0.4	0.4	0.0	0.3	**0.6**	0.0	**0.5**	**0.6**	0.0	0.4	**0.6**	0.1	0.4	0.4

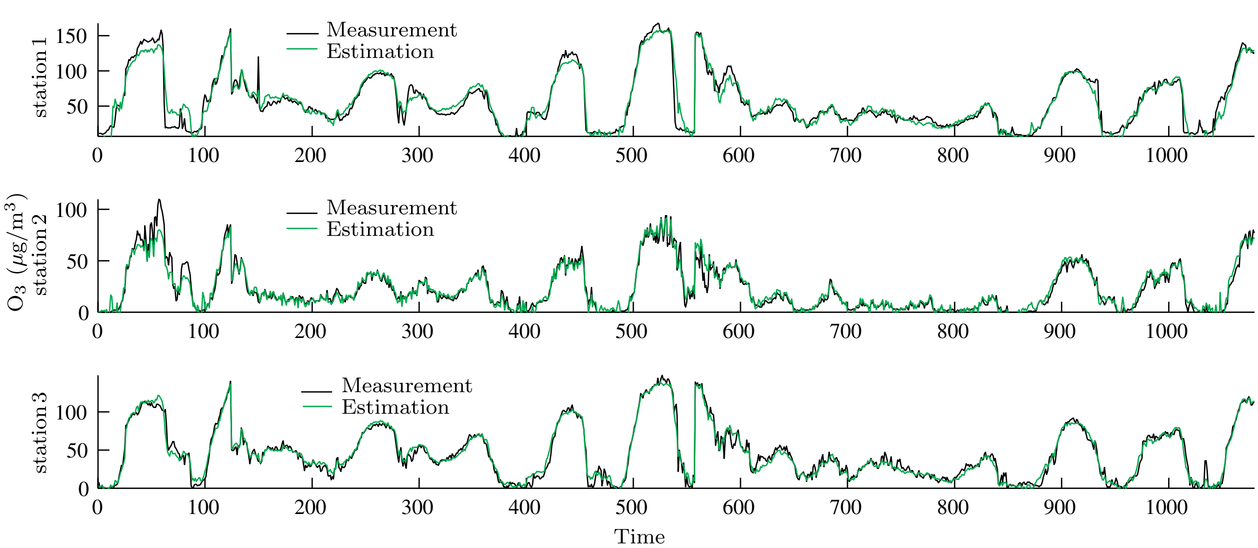

Fig. 2. Measurements and estimations of O_3 level for the first three stations.

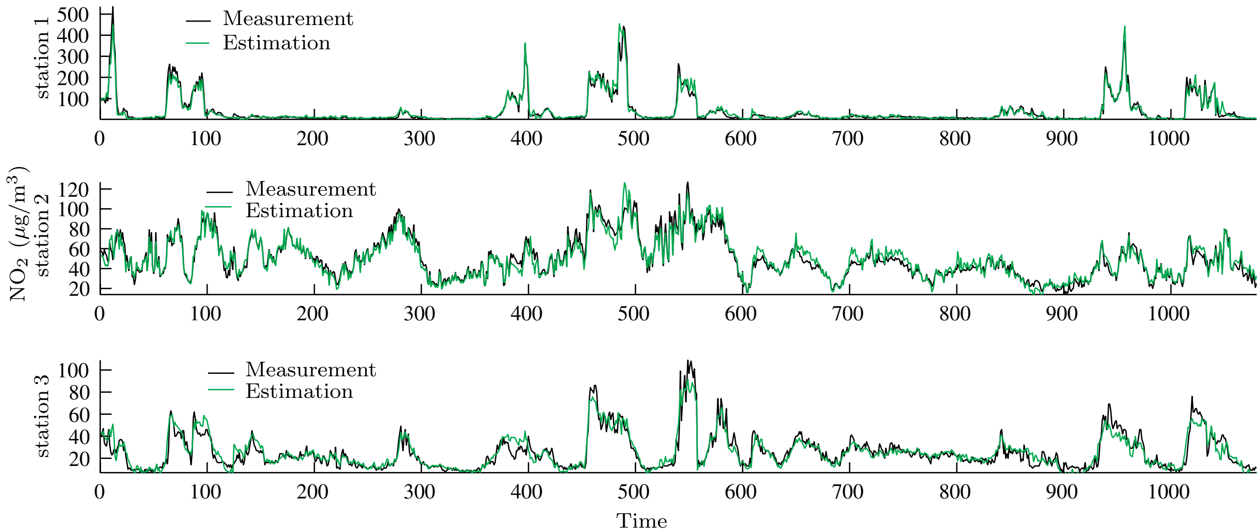

Fig. 3. Measurements and estimations of NO_2 level for the first three stations.

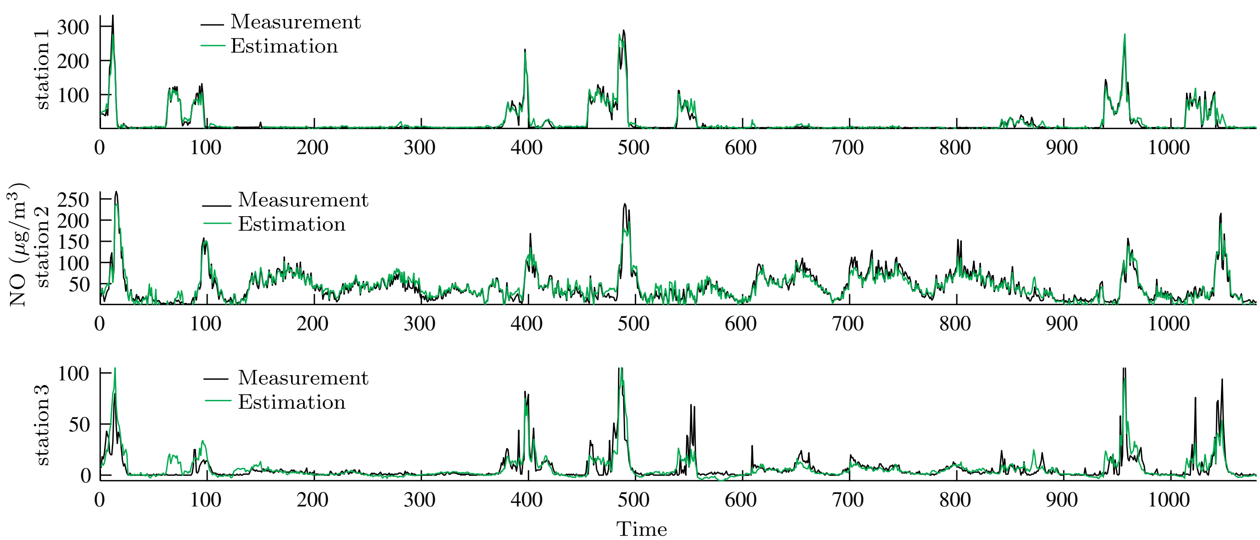

Fig. 4. Measurements and estimations of NO level for the first three stations.

centrations have variances of reconstruction error much weaker than nitrogen oxide levels.

Figs. 2–4, respectively, present measurements and estimations of O_3, NO_2 and NO levels for the first three stations, the estimations being given by the PCA model. These three stations are representative of the whole measurement network. Station 1 is a peri-urban station which has the highest ozone levels. Station 2 is an urban station which has the lowest ozone levels and the highest nitrogen oxide levels. Station 3 behaves like the others peri-urban stations.

By taking into account the nature of the considered process, the results are very satisfactory. With this PCA model, the concentrations of NO, O_3 and NO_2 are generally correctly estimated for weak, average and high levels. However, for some variables we can have modelling errors as shown in Figs. 2 (station 1) and 4 (station 3).

In conclusion, the linear PCA was able to model the relations between the various variables. However, as we could note it, certain variables being less better estimated than others, we now will examine the effect of these modelling errors on the detection and isolation of faulty sensors.

4.3. Sensor fault detection and isolation

Firstly, we will present some simple examples of sensor fault detection in order to show the results that can be obtained using the proposed fault detection, isolation and reconstruction scheme. The simulated fault is an additive step whose amplitude is rather higher than the measurement uncertainties. Then we will present a systematic study to compare the performances of the proposed index $\overline{D}_i$ with those of the $\overline{\mathrm{SPE}}$.

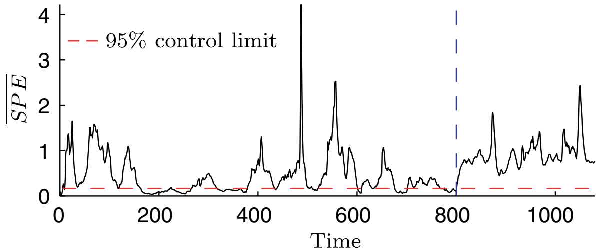

Fig. 5. $\overline{\mathrm{SPE}}$ with a fault on the variable v_7.

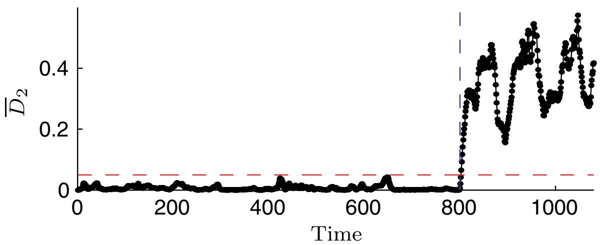

Fig. 6. $\overline{D}_2$ with a fault on the variable v_7.

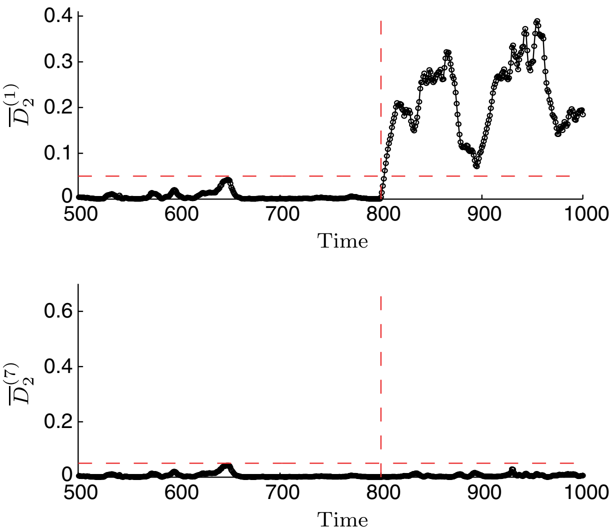

Fig. 7. Localization using the reconstruction approach: two indices $\overline{D}_2^{(j)}$, $(j = 1, 7)$.

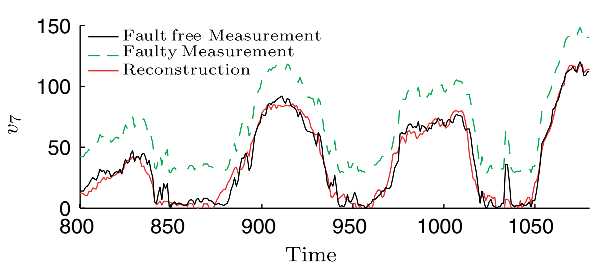

Fig. 8. Reconstruction of the faulty variable v_7 (O_3).

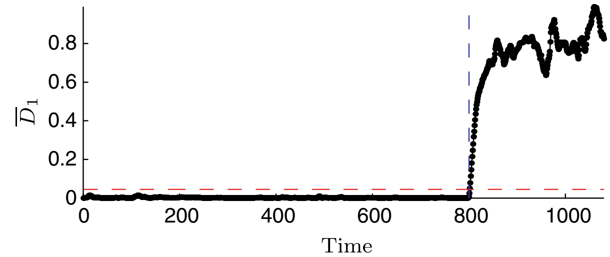

Fig. 9. $\overline{D}_1$ with a fault on the variable v_2.

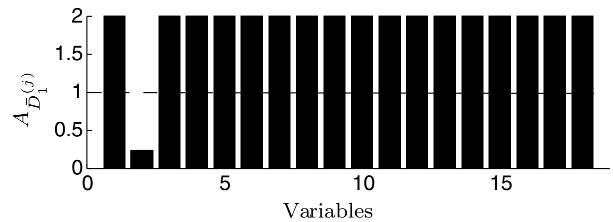

Fig. 10. Localization of the faulty variable v_2 using the index $A_{\overline{D}_1^{(j)}}$.

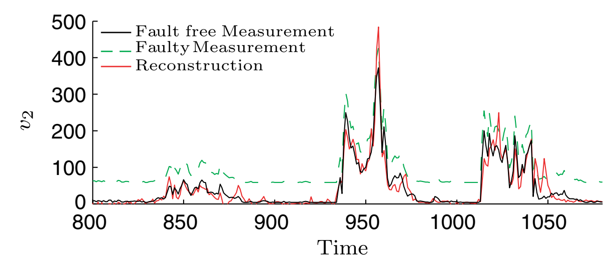

Fig. 11. Reconstruction of the faulty variable v_2 (NO_2).

4.3.1. Some examples

For the first example, a fault was simulated on the variable v_7 from sample 801 to the end of the data file (Fig. 8). The magnitude of the fault amounts to 20% of the range of variation of variable v_7. Fig. 5 shows the detection index $\overline{\mathrm{SPE}}$ for this fault. False alarms occur due to the modelling errors. If the threshold of the $\overline{\mathrm{SPE}}$ is increased in order to be insensitive to these errors, it will impossible to detect this fault. Using the proposed fault detection index, the simulated fault is detected without ambiguity on the $\overline{D}_2$ index using the two last principal components (Fig. 6).

To identify the fault, the reconstruction approach combined with the $\overline{D}_2$ index, clearly indicates that the variable responsible for this out-of-control situation is v_7. Fig. 7 only indicates the time evolution of two indices $\overline{D}_2^{(j)}$, $(j = 1, 7)$, the others being similar to $\overline{D}_2^{(1)}$. The index $\overline{D}_2^{(7)}$ clearly shows that variable v_7 was contaminated by a fault because this index is the only one in its control limit.

Variable v_7 being identified as the faulty variable, then we can reconstruct this variable in order to give a replacement value for the faulty measurement. Fig. 8 shows, for variable v_7, the fault-free measurements, the faulty measurements and the replacement values obtained by reconstruction. It is clear that the reconstructed measurements are good estimations of the fault-free measurements.

For the second example, we will consider nitrogen dioxide variables (NO_2). Now, a fault was simulated on the variable v_2 (NO_2). The simulated fault is detected on the index $\overline{D}_1$ using the last principal component (Fig. 9). The isolation of this faulty variable is depicted in Fig. 10, where the index $A_{\overline{D}_1^{(2)}}$ is lower than one which indicates that the variable v_2 is the faulty variable. Taking into account the nature of this variable, its reconstruction gives an acceptable estimation of the fault-free measurements as it is depicted in Fig. 11.

We carried out the same simulations for the other sensors of O_3, NO_2 and NO. The detection/isolation performances obtained are completely identical for these sensors.

4.4. Performances of the sensor FDI procedure

The performances of the proposed sensor fault detection and isolation procedure are mainly based on the effectiveness of the detection index $\overline{D}_i$ and on the capacity of variable reconstruction approach. To evaluate its performances, we have presented here the results of three tests:

- comparison of the detectability conditions of the indices $\overline{D}_i$ and $\overline{\mathrm{SPE}}$,
- comparison of the good detection ratio of both indices $\overline{D}_i$ and $\overline{\mathrm{SPE}}$,
- test of the quality of the variable reconstruction approach.

4.4.1. Fault detectability conditions

We extended the fault detectability conditions, introduced by Qin et al. [11] for $\overline{\mathrm{SPE}}$, to the index $\overline{D}_i$. The sensitivity to the fault $\vartheta_{i,j}$ of $\overline{D}_i$ compared to $\overline{\mathrm{SPE}}$ is defined as the ratio of the fault amplitudes checking these detectability conditions of the two indices for the same variable j:

$$\vartheta_{i,j} = \frac{\bar{\delta}_\alpha}{\bar{\delta}_{i,\alpha}} \frac{\left\| \tilde{\xi}_j^{(i)} \right\|}{\left\| \tilde{\xi}_j \right\|} \tag{25}$$

where $\tilde{\xi}_j^{(i)} = (I - \hat{C}_{m-i})\xi_j$.

Table 2 presents the results of sensitivity calculation (25) for different variables. Thus, for example, with the same level of confidence ($\alpha = 95\%$), the $\overline{D}_1$ index can detect faults on the first variable 3.2 times smaller than those detectable by the $\overline{\mathrm{SPE}}$. From this table, it is clear that there is, for

Table 3
Good detection ratio (GDR) for $\overline{D}_i$ and $\overline{\mathrm{SPE}}$, good reconstruction ratio (GRR) for different variables

Variables	GDR		GRR (%)
	$\overline{D}_i$ (%)	$\overline{\mathrm{SPE}}$ (%)	
v_1 (O_3)	75	4	94
v_2 (NO_2)	92	4	78
v_3 (NO)	99	7	92
v_4(O_3)	84	21	85
v_7 (O_3)	89	11	90
v_8 (NO_2)	88	45	86
v_{10} (O_3)	89	8	88
v_{11} (NO_2)	80	59	85
v_{13} (O_3)	97	6	88
v_{14} (NO_2)	97	89	85
v_{15} (NO)	100	97	84
v_{16} (O_3)	82	15	90
v_{17} (NO_2)	70	30	86

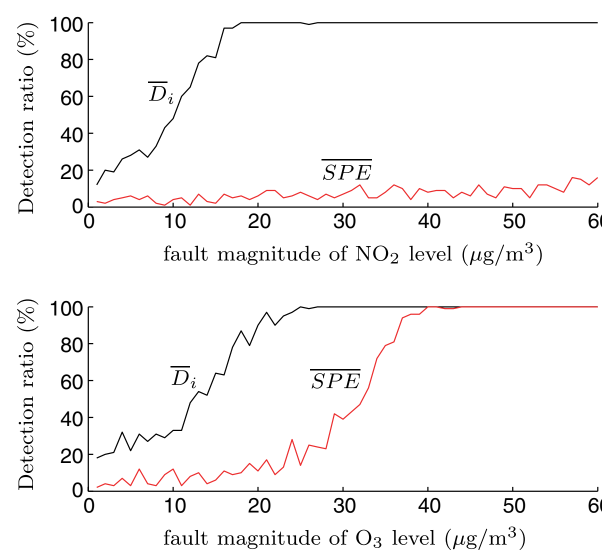

Fig. 12. $\overline{D}_i$ and $\overline{\mathrm{SPE}}$ good detection ratios for two variables.

Table 2
Sensitivity of the index $\overline{D}_i$ compared to $\overline{\mathrm{SPE}}$

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}
ϑ_1	3.2	11.7	10.6	1.1	0.9	0.3	1.2	0.6	0.1	0.9	0.1	0.4	1.9	0.7	0.2	0.2	0.1	0.2
ϑ_2	1.7	6.1	5.5	1.2	1.0	0.2	4.4	2.0	1.7	1.3	0.4	0.4	5.4	1.1	0.1	0.7	0.7	0.2
ϑ_3	1.5	4.3	3.9	1.1	0.8	0.1	3.2	1.4	0.1	4.4	1.1	0.3	3.8	0.9	0.5	2.3	1.0	0.1
ϑ_4	1.3	3.6	3.3	1.9	1.3	0.2	3.8	1.3	0.2	3.8	0.9	0.3	3.8	0.9	0.4	2.6	1.2	0.1
ϑ_5	2.4	2.7	2.6	2.3	1.3	0.5	2.8	1.2	0.4	2.8	0.7	0.2	2.9	0.6	0.5	1.9	0.9	0.2
ϑ_6	2.4	2.2	2.2	2.2	1.4	0.4	2.4	1.0	0.4	2.4	0.6	0.4	2.4	0.7	0.4	2.3	1.0	0.1
ϑ_7	1.8	1.7	1.7	1.8	1.5	1.1	1.8	1.4	0.5	1.8	1.1	0.8	1.8	0.6	0.9	1.8	0.7	0.3
ϑ_8	1.6	1.5	1.5	1.6	1.3	1.2	1.6	1.3	0.5	1.6	0.9	0.8	1.6	0.6	0.8	1.5	1.5	1.47
ϑ_9	1.4	1.3	1.3	1.4	1.2	1.3	1.4	1.2	0.5	1.4	1.0	1.0	1.4	1.1	1.0	1.4	1.3	1.3
ϑ_{10}	1.2	1.1	1.1	1.2	1.2	1.2	1.2	1.1	0.5	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2

every variable, at least one index $\overline{D}_i$ that is more sensitive to the fault than the $\overline{\mathrm{SPE}}$.

4.4.2. Detection ratio of both indices

The second test compares the good detection ratio for the two indices $\overline{\mathrm{SPE}}$ and $\overline{D}_i$. The good detection ratio is defined as the percentage of good detection on a certain number of realizations. Therefore, the best detection index is the one with the highest fault detection ratio.

In order to discriminate between the two indices, the test conditions are very difficult because the fault magnitudes are nearly equal to the estimation errors obtained with the PCA model. The additive fault value is randomly generated within the interval $[10, 30]\ \mu\mathrm{g/m}^3$. Table 3 gives the detection ratios for different variables using the two indices $\overline{D}_i$ and $\overline{\mathrm{SPE}}$. It is very clear that the detection ratios for the index $\overline{D}_i$ are greater than those of the index $\overline{\mathrm{SPE}}$ which confirms the better fault sensitivity of the index $\overline{D}_i$ by comparison with the $\overline{\mathrm{SPE}}$.

In order to plot the curves of good detection ratio for each variable, fault magnitudes are randomly generated within the interval $[1, 60]\ \mu\mathrm{g/m}^3$ for each realization. Two characteristic curves of the results obtained are depicted in Fig. 12. All the curves obtained indicate that the good detection ratios of the index $\overline{D}_i$ are always greater than those of the index $\overline{\mathrm{SPE}}$ whatever the fault magnitudes may be.

4.4.3. Quality of variable reconstruction

Table 3 presents the rate of good reconstruction for each sensor (the value is well reconstructed when the reconstruction error is lower than the standard deviation of the estimation error of the considered variable). Those results show that the variable reconstruction approach gives generally good replacement values for the faulty measurements.

5. Conclusion

In this paper, a sensor fault detection and isolation procedure based on principal component analysis is proposed to monitor an air quality monitoring network. Taking into account the high redundancy among the process variables, this procedure is based on the variable reconstruction approach in order to design the PCA model of the network, to isolate the faulty sensors and to estimate the fault amplitudes. To improve fault detection, a new detection index is proposed which monitors the last principal components (which have smallest variances) in different residual subspaces, starting with a single direction (the last principal direction) and gradually increasing the dimension of the residual to the full residual space. Therefore, the index D_i may be less sensitive to the modelling errors than the SPE. However, mathematically, it is difficult to prove that the new index has better fault sensitivity than the SPE because they perform fault detection in different residual spaces where the sensitivities of the residuals to the faults may be different.

The proposed sensor fault detection, isolation and reconstruction scheme is successfully applied to data collected on an air quality monitoring network. For this application, the index D_i improves the fault detection compared to classical detection index SPE.

The results of this feasibility study, carried out over short periods of measurements, must be confirmed, on one hand, over longer periods of measurements and, on the other hand, by quantifying the performances of the method by a Monte-Carlo simulation. For longer periods of measurements it should be important to take into account the nature of phenomena involved (nonlinear and time-varying behaviour). We plan to use extensions of PCA such as nonlinear PCA [7] or recursive PCA [10].

Acknowledgement

The authors would like to thank AIRLOR which made available the atmospheric pollution data used in this article.

References

[1] G.E.P. Box, Some theorems on quadratic forms applied in the study of analysis of variance problems: Effect of inequality of variance in one-way classification, The Annals of Mathematical Statistics 25 (1954) 290–302.
[2] R. Dunia, S.J. Qin, T.F. Edgar, Identification of faulty sensors using principal component analysis, AIChE Journal 42 (10) (1996) 2797–2812.
[3] P.M. Frank, Analytical and qualitative model-based fault diagnosis: a survey and some new results, European Journal of Control 2 (1993) 6–28.
[4] J. Gertler, T. McAvoy, Principal component analysis and parity relations—a strong duality, in: IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, SAFEPROCESS, Hull, UK, 1997, pp. 837–842.
[5] J. Gertler, W. Li, Y. Huang, T. MacAvoy, Isolation enhanced principal component analysis, AIChE Journal 45 (1999) 323–334.
[6] M.F. Harkat, G. Mourot, J. Ragot, Sensor failure detection of air quality monitoring network, in: IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, SAFEPROCESS, Budapest, Hungary, June 14–16, 2000.
[7] M.F. Harkat, G. Mourot, J. Ragot, Sensor fault detection and isolation of an air quality monitoring network using non linear principal component analysis, in: 16th IFAC World Congress, Prague, Czech Republic, July 4–8, 2005.
[8] T. Kourti, J.F. MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods, Chemometrics and Intelligent Laboratory Systems 28 (1) (1995) 3–21.
[9] T. Kourti, J.F. MacGregor, Recent developments in multivariate SPC methods for monitoring and diagnosing process and product performance, Journal of Quality Technology 28 (4) (1996) 409–428.
[10] W. Li, H. Yue, S. Valle, S.J. Qin, Recursive PCA for adaptive process monitoring, Journal of Process Control 10 (2000) 471–486.
[11] S.J. Qin, Y. Hongyu, R. Dunia, Self validating inferential sensors with application to air emission monitoring, Industrial and Engineering Chemistry Research 36 (1997) 1675–1685.
[12] S.J. Qin, R. Dunia, Determining the number of principal components for best reconstruction, Journal of Process Control 10 (2000) 245–250.