On the structure determination of a dynamic PCA model using sensitivity of fault detection

Mohamed Guerfel * **Kamel Ben Othman** *
Mohamed Benrejeb *

* *LARA Automatique, ENIT, BP 37, 1002 Tunis Belvédère, Tunisia*
(e-mail: guerfel_mohamed@yahoo.fr, kamel.benothman@enim.rnu.tn, mohamed.benrejeb@enit.rnu.tn)

Abstract: This work proposes a dynamic PCA modeling method for dynamical non-linear processes. This method uses fault free data to construct data matrix used to compute the correlation matrix and faulty system data in order to fix the dynamic PCA model parameters (the time-lag and the number of principal components). It is shown that the sensitivity of dynamic PCA-based fault detection depends on the parameters used in the model. This method is tested on a three serial interconnected tanks and subject to fluid circulation faults in its pipes.

Keywords: Process and control monitoring; Modeling; Static PCA; Dynamic PCA; Time-lag; Number of components; System fault detection.

1. INTRODUCTION

The use of multi statistical process control tools also known as MSPC became frequent for the modeling, control and diagnosis of complex and over-instrumented processes (chemicals, microelectronics, pharmaceutical..., see Venk (2003)). Static principal component analysis (PCA) is one among the most popular statistical methods, it was used successfully as a modeling tool for statical and slow dynamics processes in linear or non-linear cases, (see Qin (2003)). The extension of PCA for the dynamical modeling, called DPCA, was proposed in Ku (1995). Other work tackled this subject, like in Lee (2004), Li (2003), Mina (2007), Treasure (2004), Xie (2006). In all methods presented in scientific litterature, the model used as reference in the diagnosis procedure is obtained via the minimization of a criterion depending on the nominal data of the process. However, the obtained model can be inadequate for changes detection purposes since the minimized criterion does not necessarily maximize the changement impact of the process on the computed model (see Tamura (2007), Kano (2002)). Many changement types can affect the process, among them one distinguishes : sensors/actuators failures (see Huang (2000)), performance degradation (see Kano (2002)), operating point changes (see Zhao (2004)) and process structure modification or "system fault" (see Huang (2007)). These changes can be highlighted by various statistical tests chosen according to the changement type to be detected. For further details on these tests (also called residuals), the reader can consult Harkat (2006), Kano (2001), Singhal (2005), Guerfel (2008). This work proposes a modeling method of dynamic, linear or non-linear processes via DPCA. This method jointly uses nominal process data to build the correlation matrix to diagonalize and system fault type data to fix the time-lag and the principal component number to retain for the DPCA model. The paper is divided into the following sections. Section 2 recalls shortly the static PCA modeling and its structural parameter. Section 3 defines the dynamic PCA modeling and its structural parameters. The different changes which can affect a process and the statistical test used in this work to detect the system fault type are defined in section 4. The proposed modeling method permitting the choice of the time-lag and the number of principal component to retain for the DPCA model in the case of dynamical non-linear process is presented in section 5. Section 6 illustrates the application of the method on a three serial tanks subject to fluid circulation faults in their pipes. Finally, the last section provides a concluding summary of this work.

2. STATIC PRINCIPAL COMPONENT ANALYSIS

For the vector $z(k) = [z_1(k)\, z_2(k)\, \ldots\, z_m(k)]^t$, scaled to zero mean and unity variance and containing the m observed inputs/outputs of the process in the instant k, the data matrix Z_N resulting from the juxtaposition of $z(k)$ in different instants is written :

$$Z_N = [\, z(k) \ \cdots \ z(k+N-1)\,]^t \tag{1}$$

The subscripts N designates the number of observations used in the construction of the matrix Z_N.
Modeling a process via static PCA consists in seeking an optimal linear transformation (with respect to a variance criterion) of the original data matrix Z_N into a new one called T and defined as follows :

$$T = Z_N P \quad ; \quad T = [\, t_1 \ \cdots \ t_m \,] \in \mathbb{R}^{N \times m} \tag{2}$$

The vectors $t_q \in \mathbb{R}^N$, $q \in \{1, \ldots, m\}$, called principal components are uncorrelated and arranged in the decreasing variance order. The column vectors p_q of the matrix P

represent the eigenvectors corresponding to the eigenvalues λ_q obtained from the diagonalization of the correlation matrix Σ of Z_N :

$$\Sigma = P\Lambda P^t \quad ; \quad PP^t = P^tP = I_m \tag{3}$$

the notation $\Lambda = diag(\lambda_1 \dots \lambda_m)$ designates the diagonal matrix of eigenvalues arranged in the decreasing magnitude order $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. With the triple partitioning :

$$\Lambda = \begin{bmatrix} \hat{\Lambda} & 0 \\ 0 & \tilde{\Lambda} \end{bmatrix}, \quad P = \left[\, \hat{P} \middle| \tilde{P} \,\right], \quad T = \left[\, \hat{T} \middle| \tilde{T} \,\right] \tag{4}$$

The data matrix can be decomposed in the following form :

$$Z_N = \hat{Z}_N + E_N \ \text{ with } \ \hat{Z}_N = Z_N\hat{C} \ ; \ E_N = Z_N\tilde{C} \tag{5}$$

The matrices $\hat{C} = \hat{P}\,\hat{P}^t$ and $\tilde{C} = I_m - \hat{C}$ form the static PCA model of the process (for further details see Jollife (2003)). The matrices $\hat{Z}_N$ and E_N represent, respectively, the modeled and the non modeled variations of Z_N from ℓ components ($\ell < m$). The first ℓ eigenvectors forming the matrix $\hat{P} \in \mathbb{R}^{m\times\ell}$ constitute the representation space whereas the last $(m-\ell)$ eigenvectors forming the matrix $\hat{P} \in \mathbb{R}^{m\times\ell}$ constitute the residual space.
The identification of the static PCA model thus consists in estimating its parameters by an eigenvalue/eigenvector decomposition of the matrix Σ and determining its structural parameter which is the number of principal components ℓ to retain. An incorrect choice (too large or too small) of ℓ could mask the changes occurring in the modeled process or gives false alarms which affect the change detection procedure (see Qin (2003)). Many methods were proposed to fix ℓ in the static PCA model. The reader can find more details in Valle (1999). Most of these methods are heuristic and give a subjective number ℓ (see Harkat (2005)). In order to mitigate the disvantages of the heuristic methods, Qin and Dunia have proposed to fix ℓ via the minimization of a criterion called VNR which represents the variance of the reconstruction error of the process variables (see Qin (2000)). However, it is noted that the VNR criterion underestimates the number ℓ exact to retain in real applications cases (see Valle (1999)). All the methods aiming at the determination of ℓ seek to find its theoretical or exact value called (ℓ_{th}) which represents the theoretical number of linear or quasi-linear relations existing between the different components of $z(k)$. These methods are sensitive to the signal noise ratio and depend on the nature of the process non linearity. It is also noted that ℓ can be different from ℓ_{th} in the case of models built for diagnosis purposes provided that the static PCA model (constructed with ℓ components) can detect changes (see Frank (2000)). From this idea was born a new process modeling method via static PCA. Proposed in Tamura (2007), this method uses nominal process data to build the correlation matrix which will be diagonalized and faulty data in order to fix ℓ.

3. DYNAMIC PRINCIPAL COMPONENT ANALYSIS

Dynamic principal component analysis proposed in Ku (1995) and known as DPCA aims at finding dynamical linear relations between the process variables. The principle of this method is identical to the static PCA. Starting from a scaled to zero mean and unity variance data vector $z^d(k) = [z^t(k)\ z^t(k-1)\ \dots\ z^t(k-s)]$, where s designates the used time-lag, the data matrix $Z_N^d\,(k,s) \in \mathbb{R}^{N\times m(s+1)}$ is built as follows :

$$Z_N^d\,(k,s) = \begin{bmatrix} z^t\,(k) & \dots & z^t\,(k-s) \\ z^t\,(k+1) & \dots & z^t\,(k-s+1) \\ \vdots & \ddots & \vdots \\ z^t\,(k+N-1) & \dots & z^t\,(k+N-1-s) \end{bmatrix} \tag{6}$$

with $N > m(s+1)$ and $k > s$.
The correlation matrix obtained from $Z_N^d\,(k,s)$, noted Σ_d, is computed and diagonalized in order to obtain the eigenvectors and the eigenvalues matrices noted respectively P_d and Λ_d. Each one of these two matrices is divided into two parts the first corresponding to the representation space ($\hat{\Lambda}_d,\ \hat{P}_d$) and the second corresponding to the residual space ($\tilde{\Lambda}_d,\ \tilde{P}_d$). The principal components vector noted $t^d \in \mathbb{R}^{1\times m(s+1)}$, can be computed in an instant k as follows :

$$t^d(k) = \left[\, \hat{t}^d(k) \mid \tilde{t}^d(k) \,\right] = z^d(k)P_d = z^d(k)\left[\, \hat{P}_d \mid \tilde{P}_d \,\right] \tag{7}$$

The structural parameters in DPCA modeling are the number of principal components ℓ and the time-lag s. The number ℓ can be fixed via the methods used in static PCA after the choice of s which is a very delicate problem. The modeling of data obtained from dynamic process via static PCA constructs an approximate static model of the real process and does not reveal its exact structure (see Ku (1995)). It is possible to detect modifications in dynamical processes via static PCA as in Harkat (2006) and Sharmin (2008), but the theoretical bases of the method are lost since the principal components are no longer uncorrelated and do not follow a normal multivariate statistical distribution. In this case, it will be very difficult to detect small changes in the process parameters as long as the variation domain of the process variales remain the same before and after the change. For a well chosen time-lag $s = s_{min}$, all the static and dynamic relations ruling the process will be represented by the last eigenvectors corresponding to the smallest eigenvalues of Σ_d computed from $Z_N^d\,(k, s_{min})$. Taking a time-lag s higher than s_{min} in the construction of $Z_N^d\,(k,s)$ used for the computation of the DPCA model will not bring any supplementary information but will add redundant relations which were obtained from the construction of the DPCA model using the matrix $Z_N^d\,(k, s_{min})$ (see Ku (1995)). Many methods were proposed for the choice of s. They seek to find the theoritical $s = s_{min}$, most of them are heuristic as in Ku (1995) or resulting from the identification techniques as AIC, see Akaike (1974), Larimore (1990), Li (2003) and MDL, see Simoglou (2002), Rissanen (1978) which privilege the approximation of the data matrix. None of those methods was built in the purpose of minimization of s compared to fault detection.

4. STATISTIC USED FOR SYSTEM FAULT DETECTION

The physical processes are subject to changes in their operating conditions. In the case of non stationary processes, these changes can be sensors/actuators failures (see Huang (2000)), operating point changes, performance degradation or process structure modification. The operating point changes are characterized by an augmentation in the mean of one or many inputs (see Zhao (2004)). The process performance degradation can be expressed as an augmentation in the variance of one or many process variables under the hypothesis of independent and identically distributed noise (see Kano (2002)). The process structure modifications known as "system fault" appear as changes in the structure or a modification in its model parameters (see Huang (2007)). All these changes can be highlighted by various statistical tests chosen according to the changement type to be detected. Only the system fault type is included in this work. The best indices for the detection of such modification are the ones based on the residual space (see Guerfel (2008)). For this reason, one proposes the use of the D_i statistic, $i = 1, 2, \ldots, (m(s+1) - \ell)$, which is defined in Harkat (2006) as the sum of squared last principal components. This statistic is computed every instant k :

$$D_i(k) = \sum_{j=m(s+1)-i+1}^{m(s+1)} \left(\tilde{t}_j^d(k)\right)^2 \tag{8}$$

The variable $\tilde{t}_j^d(k)$ designates the j^{th} principal component obtained at the instant k. The index D_i represents an SPE (see box (1954)) computed from a DPCA model with $(m(s+1)-i)$ components, its detection threshold $\tau^2_{i,\alpha}$, can be computed in the following way :

$$\tau^2_{i,\alpha} = g^{(i)} \chi^2_{h^{(i)},\alpha} \tag{9}$$

The notation χ^2 designates the *Chi*-square distribution, α designates the used confidence limit, $g^{(i)}$ and $h^{(i)}$ are defined as follows :

$$g^{(i)} = \frac{\sum\limits_{j=m(s+1)-i+1}^{m(s+1)} \lambda_j^2}{\sum\limits_{j=m(s+1)-i+1}^{m(s+1)} \lambda_j} \qquad h^{(i)} = \frac{\left(\sum\limits_{j=m(s+1)-i+1}^{m(s+1)} \lambda_j\right)^2}{\sum\limits_{j=m(s+1)-i+1}^{m(s+1)} \lambda_j^2} \tag{10}$$

The quantities $\lambda_j,\ j \in \{m(s+1) - i + 1, \ldots, m(s+1)\}$, designate the j^{th} eigenvalues of Σ_d. A system fault is detected if the D_i index is higher than its threshold $\tau^2_{i,\alpha}$.

5. PROPOSITION FOR DYNAMICAL PROCESSES MODELING

The proposed method is similar to the one defined in Tamura (2007) and called MDM abreviation of *Multi Dimensionnal Monitoring*. Contrary to the MDM, the proposed method allows not only the choice of ℓ but also the choice of the minimum time-lag s to retain for the data matrix $Z_N^d(k,s)$ used in the construction of the DPCA process model. The principle of the method is the following :

(1) Begin method
(2) Initialization $S_{init} = 0$ and $\ell = 0$
(3) Build $Z_N^d(k, S_{init})$, Compute Σ_d, P_d and Λ_d
(4) Compute D_i from system fault data for i varying from 1 to the number of column in $Z_N^d(k,s)$ minus 1
(5) If the fault is detected with any of D_i then go to step 7 else go to step 6
(6) $S_{init} = S_{init} + 1$ go to step 3
(7) $s = S_{init}$ and ℓ is equal to the difference between the number of column in $Z_N^d(k,s)$ and the largest value of i which permits the fault detection.
(8) End method

The disadvantage of the proposed method lies in the fact that a knowledge of information on the system fault is necessary to ensure the choice of structural parameters (s and ℓ) to be retained for the DPCA model. From another point of view, if information on the system fault is available, this method becomes very attractive because it determines the simplest model allowing the system fault detection. Figure (1) summarizes the algorithm of the method in the case of a single system fault "j" affecting the modeled process.

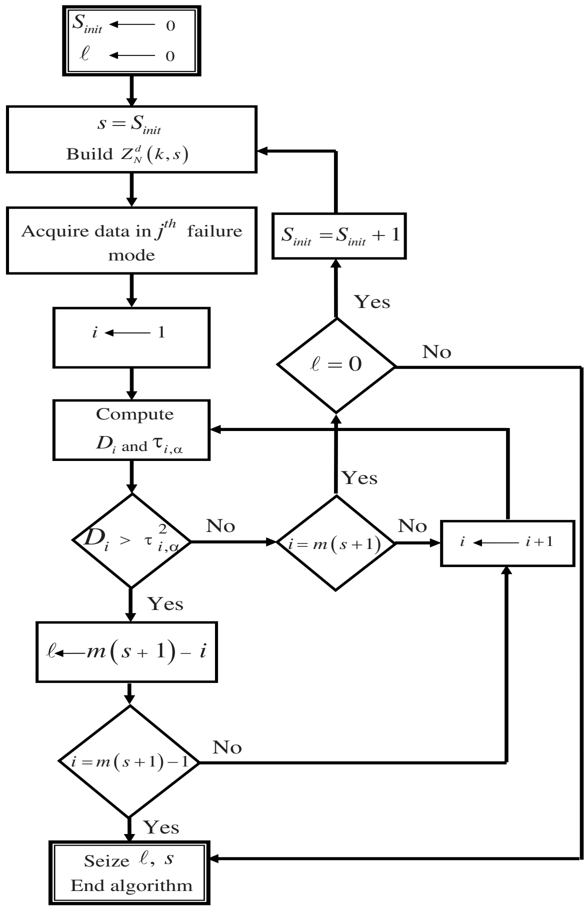

Fig. 1. Algorithm of the proposed method for the determination of s and ℓ in the DPCA model

In order to supress false alarms, the process is considered in failure mode ($D_i > \tau^2_{i,\alpha}$), if D_i has shown eight succeeding values larger than $\tau^2_{i,\alpha}$. The value "eight" is determined in an empirical way and must be adjusted according to the treated application.

6. APPLICATION IN THE MODELING OF THREE TANK SYSTEM

The modeled process illustrated in figure (2), is formed by three identical serial tanks. It contains two inputs : flows q_1, q_2 and three outputs H_1, H_2 and H_3 representing respectively the heights in the first, second and third tanks.

These tanks are interconnected at the bottom by pipes. Two valves V_3 and V_2, separating respectively tank 2 from tank 3 and tank 2 from the outside are introduced in order to model the flows perturbations in the pipes. For a sampling period equal to one second, the discrete process equations are :

$$\begin{cases} H_1(k) = A^{-1}\left(q_1(k) + q_{31}(k) - q_{10}(k)\right) + H_1(k-1) \\ H_2(k) = A^{-1}\left(q_2(k) - q_{23}(k) - q_{20}(k)\right) + H_2(k-1) \\ H_3(k) = A^{-1}\left(q_{23}(k) - q_{31}(k)\right) + H_3(k-1) \\ q_{10}(k) = K_1\sqrt{H_1(k)} \\ q_{20}(k) = K_2\sqrt{H_2(k)} \\ q_{31}(k) = K_{31} f\left(H_3(k) - H_1(k)\right) \\ q_{23}(k) = K_{23} f\left(H_2(k) - H_3(k)\right) \end{cases} \tag{11}$$

where A equal to $0.01539\,m^2$, designates the tank section. The constants K_1, K_2, K_{31} and K_{23} respectively equal to $1.816\,e^{-4}$, $9.804\,e^{-5}$, $1.005\,e^{-4}$ and $7.804\,e^{-5}$ are the process characteristics. The term $f(\centerdot)$ designates a non linear function defined as follows :

$$f(x) \triangleq sign(x)\sqrt{|x|} \tag{12}$$

The measured process variables are the inputs z_1^b, z_2^b and

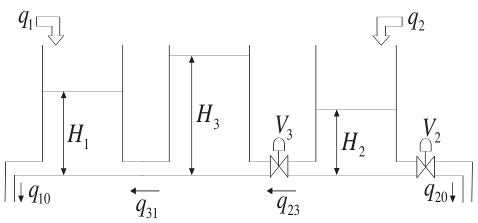

Fig. 2. Three tanks system

the outputs z_3^b, z_4^b and z_5^b. Theses measures are related in the instant k to the physical values via the following equations :

$$\begin{cases} z_1^b(k) = H_1(k) + \varepsilon_1(k) \\ z_2^b(k) = H_2(k) + \varepsilon_2(k) \\ z_3^b(k) = H_3(k) + \varepsilon_3(k) \\ z_4^b(k) = q_1(k) + \varepsilon_4(k) \\ z_5^b(k) = q_2(k) + \varepsilon_5(k) \end{cases} \tag{13}$$

The quantities $\varepsilon_r(k)$, $r \in \{1, \ldots, 5\}$ designate gaussian centered measurement noise. Its standard deviation is equal to 3% of that of the entries. The flows q_1 and q_2 are expressed in m^3/s. They are chosen to be random durations crenels with variable amplitudes respectively in $[3.20, 6.71] \times 10^{-5}$ for q_1 and in $[5.73, 9.57] \times 10^{-5}$ for q_2. The tanks initial heights are expressed in meter. Their values are 0.147, 0.276 and 0.195 respectively for the first, second and third tank. The system is firstly simulated under nominal operation during 4000 samples. After centering and reducing the inputs/outputs measures, the vector z is built at each instant k as follows :

$$z(k) = \left[\, z_1(k)\ z_2(k)\ z_3(k)\ z_4(k)\ z_5(k) \,\right]^t \tag{14}$$

where $z_r(k)$ designates the centered and reduced value of $z_r^b(k)$. The data matrix is constructed via (1). It will be used in the computation of the matrices Λ and P.

In the dynamical case, the vector z^d is constructed in an instant k for a time-lag s using time-lagged vectors z obtained in the static case as following :

$$z^d(k) = \left[\, z^t(k)\ z^t(k-1)\ \ldots\ z^t(k-s) \,\right] \tag{15}$$

The data matrix $Z_N^d(k,s)$ is built via (6). It will be used for the computation of the matrices Λ_d and P_d.

Figure 3 shows the process scree plot for a time-lag s respectively equal to zero, one and two. The eigenvalues of Σ (built for $s=0$) are not null and do not indicate the presence of any linear or quasi linear relation between the measured process variables at the same instant. The last three eigenvalues of Σ_d built for $s=1$ are quasi null and show the existance of three quasi linear relations verified by the measured process variables between two sampling instants. The correlation matrix Σ_d built for $s=2$ shows six quasi null eigenvalues. They indicates the presence of six quasi linear relations verified by the measured process variables between three sampling instants. The approximation of the non linear relations verified by the measured process variables may be better if one uses time-lags higher than two but the computation complexity will increase as well.

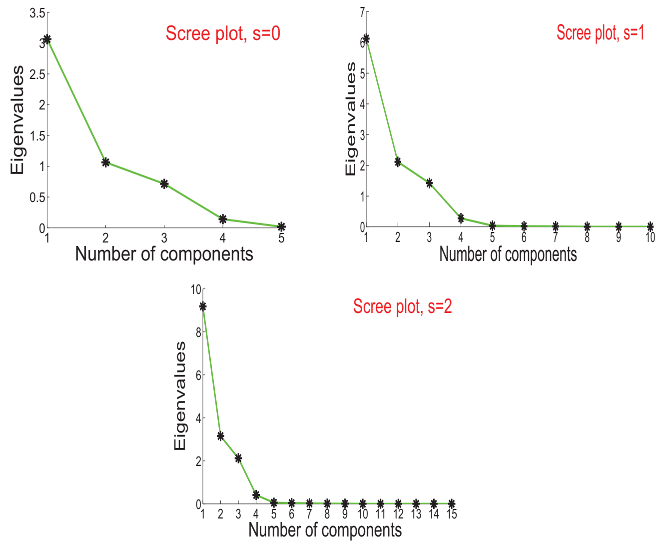

Fig. 3. Scree plots for $s=0$, $s=1$ and $s=2$

Twenty measures representing a system fault are generated in order to fix the time-lag s used for the construction of $Z_N^d(k,s)$. This system fault represents a variation of 1.1 of K_2 from its nominal value. The application of the algorithm in section 5 gives that $s=1$ and $\ell=7$ are sufficient to detect such a fault.

A second simulation shown in figure 4 is realized during 4000 instants and perturbations in flows circulation are introduced by varying K_2 in the following way. For the first 1000 instants, K_2 is equal to its nominal value. In the second 1000 instants, K_2 is equal to 1.1× its nominal value. In the third 1000 instants, K_2 is equal to 1.2× its nominal value. For the latest 1000 instants, K_2 is equal to 1.3× its nominal value.

The evolution of the statistics D_i, $i \in \{i, \ldots, 4\}$, respectively for $s=0$ and $s=1$ in the case of the second simulation are shown in figure 5 and 6. On one hand, the flows perturbations are not detectable with the statical PCA model. The statistics D_i obtained from its application on

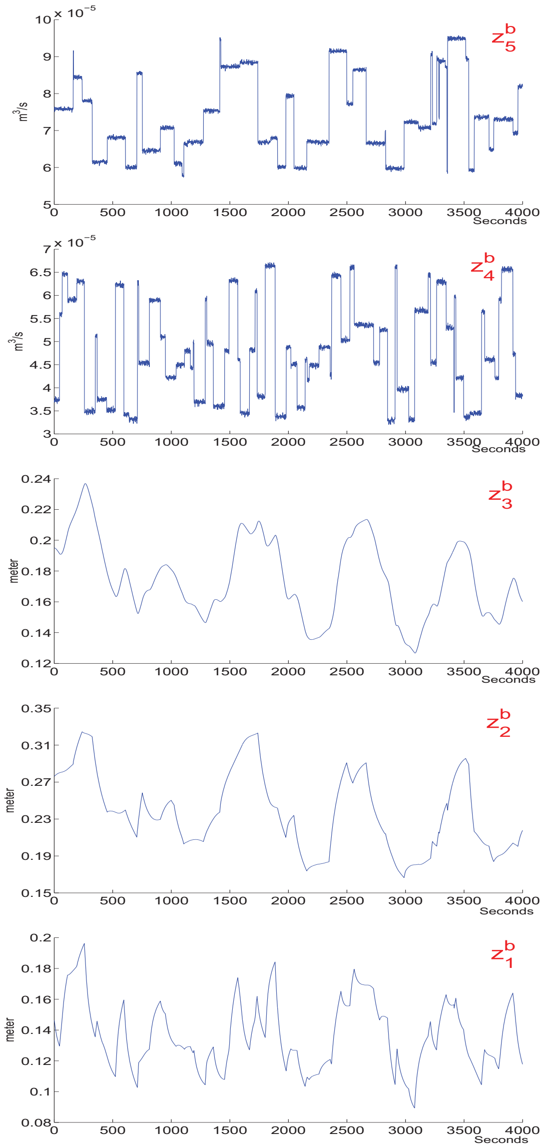

Fig. 4. Evolution of flows z_5^b, z_4^b and heights z_3^b, z_2^b and z_1^b in the second simulation

the current process data are always under their thresholds (figure 5) excepting some aberrant values. On the other hand, the flows perturbations are well detected with a DPCA built with $s = 1$. The statistics D_i, $i \in \{1, \ldots 3\}$, obtained from its application on the current process data allow the detection of all the system faults that were simulated. The maximal dimension of the residual space that can detect the perturbations is equal to three. The time-lag s used to construct the data matrix from which the DPCA model is built depending on the magnitude of the system fault to be detected. Increasing the value of s may lead to a better linear approximation of the real non linear relations existing between the time-lagged process measurements which can reduce the magnitude of the system fault to be detected.

7. CONCLUSION

The proposed method permits the estimation of the time-lag s and the choice of the principal component number ℓ used in the process modeling via DPCA. Contrary to the majority of the existing methods which use data gathered during nominal operation conditions to estimate s and ℓ, the proposed method uses data representing nominal process operating condition to build the data matrix which is used in the computation of DPCA model and other process data belonging to the system fault type to compute the structural parameters s and ℓ. The suggested method proves to be interesting if information relating to the system fault mode is a priori available. In this case, the method determines the least complex model allowing the detection of the considered system fault. Built around a

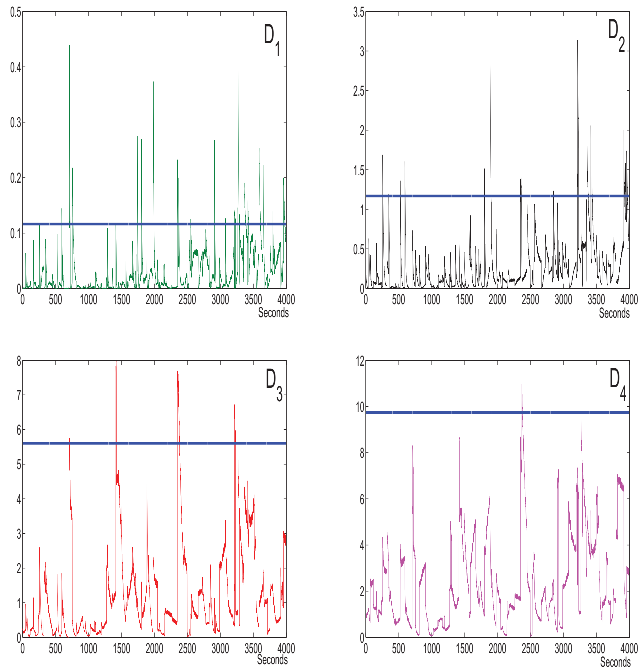

Fig. 5. Evolution of $D_i, i \in \{1, \ldots, 4\}$ in the case of $s = 0$

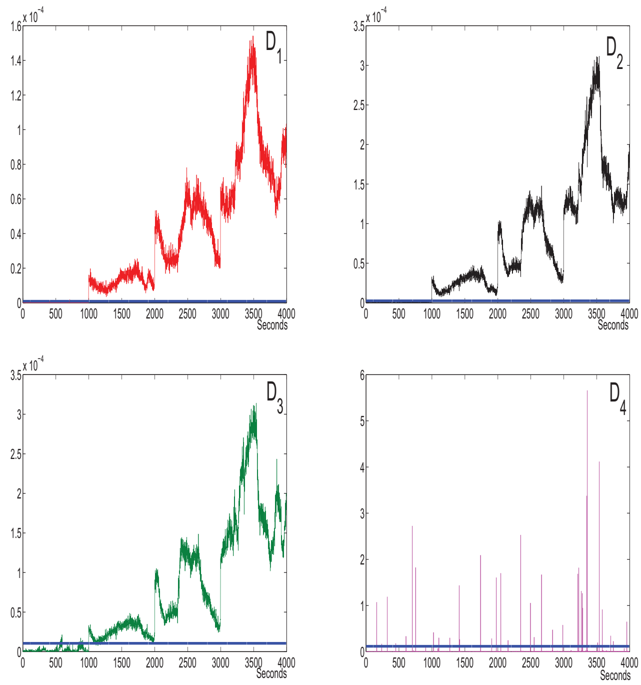

Fig. 6. Evolution of $D_i, i \in \{1, \ldots, 4\}$ in the case of $s = 1$

particular operating point, this method can be sensitive to operating point changes. In the abscence of a system fault, the obtained model presents a risk of generating false alarms due to a shift of the current process variables operating point. The extension of the method to multiple system fault cases and the minimization of the false alarms due to the operating point shift will be considered in forthcoming works.

REFERENCES

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, volume 19, pages 716–723, 1974.

G. E. P. Box. Some theorems on quadratique forms applied in the study of analysis of variance problems : Effect of inequality of variance in one-way classification. *The Annals of Mathematical Statistics*, volume 25, pages 290–302, 1954.

P.M. Frank, E. A. Garcia, B. Köppen-Seliger. Modelling for fault detection and isolation versus modelling for control. *Mathematics and Computers in Simulation*, volume 53, pages 259–271. 2000.

M. Guerfel, G. Mourot, J. Ragot, M. Benrejeb, K. Benothman. Comparaison des indices de détection de changements des modes de fonctionnement par ACP. Cas des indices basés sur l'estimation d'état. *1st International Workshop on Systems Engineering Design & Applications, SENDA'08*, Monastir, Tunisia 24-26 October, 2008.

M. F. Harkat, G. Mourot, J. Ragot. Diagnostic de fonctionnement de capteurs d'un réseau de surveillance de la qualité de l'air par analyse en composantes principales. *Journal européen des systèmes automatisés*, volume 39. pages 417–436, 2005.

M. F. Harkat, G. Mourot, J. Ragot. An improved PCA scheme for sensor FDI: Application to an air quality monitoring network. *Journal of Process Control*, volume 16, pages 625–634, 2006.

Y. Huang, J. Gertler, T. McAvoy. Sensor and actuator fault isolation by structured partial PCA with nonlinear extensions. *Journal of Process Control*, volume 10, pages 444-459, 2000.

H.P. Huang, C.C. Li, J.C. Jeng. Multiple multiplicative fault diagnosis for dynamic processes via parameter similarity measures *Industrial and Engineering Chemistry Research*, volume 46, pages 4517-4530, 2007.

I. T. Jolliffe. *Principal component analysis.* Springer-Verlag, New York, 2003.

M. Kano, H. Ohno, S. Hasebe, I. Hashimoto. New multivariate statistical process monitoring method using principal component analysis. *Computers and Chemical Engineering*, volume 25, pages 1103–1113, 2001.

M. Kano, K. Nagao, S. Hasebe, I. Hashimoto, H. Ohno, Strauss R. Comparison of multivariate statistical process control monitoring methods with applications to the Eastman challenge problem. *Computers and Chemical Engineering*, volume 26, pages 161–174, 2002.

W. Ku, R. H. Storer, C. Georgakis. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, volume 30, pages 179–196, 1995.

W. E. Larimore. Canonical variate analysis in identification, filtering and adaptative control. *29th Conference on Decision and Control*, Honolulu, Hawai Islands December, pp. 635–639, 1990.

C. Lee, S. W. Choi, I. B. Lee. Sensor fault identification based on time-lagged PCA in dynamic processes. *Chemometrics and Intelligent Laboratory Systems*, volume 70, pages 165–178, 2004.

W. Li, S. J. Qin. Consistent dynamic PCA based on errors-in-variables subspace identification. *Journal of Process Control*, volume 11, pages 661–678, 2003.

J. Mina, C. Verde. Fault detection for large scale systems using dynamic principal components analysis with adaptation. *International Journal of Computers, Communications & Control*, volume 2, pages 185–194, 2007.

S. J. Qin, R. Dunia. Determining the number of principal components for best reconstruction. *Journal of Process Control*, volume 10, pages 245–250, 2000.

S. J. Qin. Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, volume 17, pages 480–501, 2003.

R. Sharmin, S. L. Shah, U. Sundararaj. A PCA Based Fault Detection Scheme for an Industrial High Pressure Polyethylene Reactor. *Macromolecular Reaction. Engineering*, volume 2, pages 12–30, 2008.

A. Singhal, D. E. Seborg. Clustering multivariate time-series data. *Journal of chemometrics*, volume 19, pages 427–438, 2005.

A. Simoglou, E. B. Martin, A.J. Morris. Statistical performance monitoring of dynamic multivariate process using state space modelling. *Computers and Chemical Engineering*, volume 26, pages 909–920, 2002.

J. Rissanen. Modelling by shortest data description. *Automatica*, volume 14, pages 465–471, 1978.

M. Tamura, S. Tsujita. A study on the number of principal components and sensitivity of fault detection using PCA. *Computers and Chemical Engineering*, volume 31, pages 1035–1046, 2007.

R. J. Treasure, U. Kruger, J. E. Cooper. Dynamic multivariate statistical process control using subspace identification. *Journal of Process Control*, volume 11, 661–678, 2004.

S. Valle, L. Weihua, S. J. Qin. Selection of the number of principal components : The variance of the reconstruction error criterion with a comparison to other methods. *Industrial and Engineering Chemistry Research*, volume 38, pages 4389–4401, 1999.

V. Venkatsubramanian, R. Rengaswamy, K. Yin, S. N. Kavuri. A review of fault detection and diagnosis; part I: quantitative model-based methods. *Computers and Chemical Engineering*, volume 27, pages 293–311, 2003.

M. Wax, I. Ziskind. Detection of the number of coherent signals by the MDL principle. *IEEE Transactions on Acoustic, Speech and signal processing*, volume 37, pages 1190–1195, 1989.

L. Xie, J. Zhang, S. Wang. Investigation of Dynamic Multivariate Chemical Process Monitoring. *Chinese Journal of Chemical Engineering*, volume 14, pages 559–568, 2006.

S. J. Zhao, J. Zhang, Y. M. Xu. Monitoring of processes with multiple operating modes through multiple principle component analysis models. *Industrial Engineering Chemical Research*, volume 43, pages 7025–7035, 2004.