

Nonlinear Statistical Process Monitoring based on Competitive Principal Component Analysis

Messaoud RAMDANI¹ Khaled MENDACI²

¹ Electronics Department, University of Badji-Mokhtar Annaba, BP. 12, 23000, Annaba, Algeria

² Science and Technology Department, University of Larbi Ben Mhidi Oum El Bouaghi, 04000, Algeria

Abstract

Traditional process monitoring techniques assume the normal operating conditions (NOC) to be distributed normally. However, for processes with more than one operating regime, building a single subspace model to monitor the whole process operation performance may not be efficient and will lead to high rate of missing alarm. To handle this situation, a monitoring strategy using multiple subspace models is presented in this paper. From the experimental results using a simulation model of a continuous flow aerated bioreactor for wastewater treatment in pulp and paper industry it has been shown that the proposed approach is very promising.

Keywords: Statistical Process monitoring, fuzzy clustering, local statistics, confidence limits, nonlinear systems.

1. Introduction

Nowadays, the significant advances in process sensing technologies make it possible to collect huge amounts of measured data. However, it is very difficult to extract the valuable information about the state of the process. The inefficient handling of the process measurements is mainly due to the presence of nonlinear relationships between the variables. Thus, there is a need for an effective information processing approach to monitor nonlinear and complex systems.

Modern model-based condition monitoring is based on information redundancy principle. In fact, abnormal states can be detected by a consistency checking between an observed behavior as indicated by sensors and an expected behavior provided by mathematical models [6], [12]. The models may be explicit, obtained from first principles or system identification [8], [20] or implicit, obtained by principle component projection [15]. For fault isolation, some structured residuals, which respond to subsets of faults [7], [9] may be generated by algebraic transformation. In the principal component framework, the direct computation involves structured partial principal component models [3]. Alternative fault isolation techniques, in the PCA framework, involve contribution charts [15], and the sensor validity index [5] based on the reconstruction principle.

Over the last decade, the multivariate statistical process control (MSPC) has been used as a performance monitoring tool in a wide range of industries. MSPC aims to remove the redundancy often observed in the recorded variables by defining a reduced set of latent variables. PCA is one of the most popular MSPC techniques. More precisely, the variations in the original measurement spaces are described by a reduced set of latent variables without the loss of significant contributions. Then fault detection and diagnosis (FDD) is accomplished in the low-dimensional space by monitoring the sum of prediction error (SPE) and principal component scores charts [15]. According to the principle of PCA, the loading plot can provide the relationship between the original variables, which can be utilized to identify faults because most anomalies may destroy the relationships between process variables.

Unfortunately, if the process exhibits multiple operating regimes, the application of linear PCA gives an excessive number of false alarms or alternatively, missed detection of processes faults, which significantly compromises the reliability of the monitoring system. To overcome the above limitations, an extensive research has been carried out and various solutions have been proposed. Nonlinear extensions have been reported by Hastie [10] using principal curves, Kramer [16] using auto associative neural networks, Qin and McAvoy [18] using partial least squares (PLS) based on neural networks. Webb [23] and Wilson [24] proposed a nonlinear extension of PCA using radially symmetric kernel functions, and radial basis function (RBF) networks, respectively.

This paper is organized as follows. Section 2 describes the statistical monitoring principles. In section 3, the competitive principal component analysis is developed. In section 4, a simulation case study using a biological plant is performed and the results are discussed. Finally, some concluding remarks as well as some possible improvements are given in section 5.

2. Multivariate statistical process monitoring principles

In order to introduce the necessary notations used in the paper, the PCA theory is briefly reviewed. Let us consider a process data matrix $X(N \times n)$ composed of N

sample vectors with n variables measurements collected when the process is in control.

Denote the correlation matrix of X as $\Sigma = X^T X / (N - 1)$, and performing singular value decomposition (SVD) to the matrix Σ yields

$$\Sigma = U \Lambda U^T \quad (1)$$

where $U_{n \times n}$ is a unitary matrix and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal one containing the eigenvalues in decreasing order. The column vectors in the matrix $U = [u_1, u_2, \dots, u_n]$ forms a new orthogonal base of space \mathcal{R}^n , and the first q ($< n$) linear independence vectors $\hat{P} \equiv U_q = [u_1, u_2, \dots, u_q]$ of U spans the principal component subspace \hat{S} . The other $n - q$ vectors $\tilde{P} \equiv U_{n-q} = [u_{q+1}, u_{q+2}, \dots, u_n]$ of U spans the residual space \tilde{S} . Selecting the correct number of principal components to retain in a model is a nontrivial task, and there are many subjective rules that can be used to help with this task. The number of principal components to retain can be selected as its cumulative percent variance (CPV) is larger than a prescribed threshold. The data vector $x \in \mathcal{R}^n$ can be decomposed as

$$x = \hat{x} + \tilde{x} = \hat{C}x + \tilde{C}x \quad (2)$$

where $\hat{x} \in \hat{S}$ and $\tilde{x} \in \tilde{S}$ are projection of x on the subspaces \hat{S} and \tilde{S} , respectively. The matrix $\hat{C} = U \hat{U}^T$ and $\tilde{C} = U \tilde{U}^T$. The score vector in the space model space $t = \hat{P}^T x \in \mathcal{R}^q$ is a reduced, q -dimensional representation of the observed vector x . On the other hand, the residual $e = (I - \hat{P} \hat{P}^T)x \in \mathcal{R}^n$, represents the portion not explained by the PCA model. The PCA statistical monitoring model is build on two hypothesis tests in subspaces \hat{S} and \tilde{S} . The statistic used in \hat{S} is the Hotelling T^2 which is defined as

$$T^2 = \left\| \Lambda_q^{-1/2} U_q x \right\|^2 \leq T_{\text{lim}}^2 \quad (3)$$

with $\Lambda_q = \text{diag}(\lambda_1, \dots, \lambda_q)$. The statistic used in \tilde{S} is the Q-statistic or the sum of prediction error (SPE), which is defined as below

$$\text{SPE} = \left\| \tilde{C}x \right\|^2 \leq \delta_\alpha^2 \quad (4)$$

where δ_α^2 is the control limit for SPE index. The $100(1 - \alpha)\%$ control limit for T^2 is calculated by means of an F-distribution as follows:

$$T_{\text{lim}}^2 = \frac{n(N-1)}{N-n} F(n, N-1; \alpha) \quad (5)$$

where $F(n, N-1; \alpha)$ is an F-distribution with degrees of freedom n and $N-1$, and with level of significance α . A confidence limit expression for SPE , when \mathbf{x} follows a normal distribution, has been developed by Jackson and Mudholkar [13]. Since then, it has often used as a limit for fault detection [5]. Box [2] has previously shown that:

$$\delta_\alpha^2 = \mathbf{g} \chi_{n, \alpha}^2 \quad (6a)$$

$$\mathbf{g} = \theta_2 / \theta_1 \quad (6b)$$

$$\mathbf{h} = \theta_1^2 / \theta_2 \quad (6c)$$

where $\theta_i = \sum_{j=q+1}^m \lambda_j^i$ for $i = 1, 2$. To reduce false alarms, exponentially weighted moving average (EWMA) filter can be applied to the residuals \mathbf{e} , but it introduces unfortunately a delay in detecting faults. The general EWMA expression for residual is:

$$\bar{\mathbf{e}}(k) = (I - \beta) \bar{\mathbf{e}}(k-1) + \beta \mathbf{e}(k) \quad (7)$$

$$\overline{\text{SPE}}(k) = \left\| \bar{\mathbf{e}}(k) \right\|^2 \quad (8)$$

where $\bar{\mathbf{e}}(k)$ and $\overline{\text{SPE}}(k)$ are the filtered residuals and square prediction error respectively. $\beta = \gamma I$ denotes a diagonal matrix whose diagonal elements γ are forgetting factors for the residuals. This new test suggests the existence of an abnormal condition when:

$$\overline{\text{SPE}}(k) > \bar{\delta}_\alpha^2 \quad (9)$$

$$\bar{\delta}_\alpha^2 = \frac{\gamma}{2-\gamma} \delta_\alpha^2$$

The SPE or $\overline{\text{SPE}}$ test for sensor fault detection based PCA, were largely used, but it presents a major disadvantage. Indeed, it may be that the supervised process presents a nonlinear behavior, which is the case in the majority of situations, and as the PCA is a linear technique, in addition to the measurement noise, residuals will be also affected by modelling errors.

3. Local principal component analysis

Local Principal Component analysis (Local PCA) is a useful tool for finding local features of large scale data bases. The basic idea behind it is to partition the whole data set into several small subregions and find the corresponding linear structures. To this end, several approaches have been proposed since 70's of the last century. Kambhatla *et al.* [14] and [11] proposed iterative algorithms based on the reconstruction distances. In [14], a two stage algorithm composed of the hard clustering of data sets and the estimation of local principal components in each cluster. Hinton *et al.* [11] extended the idea to a soft version based on the EM algorithm [4] in which the assignments are considered "missing data" and the responsibility of a principal component analyzer for each data point is estimated by using the corresponding reconstruction cost.

Roweis [19] and Tipping *et al.* [21] defined probabilistic models for PCA in which all of the model parameters are estimated through the maximization of a single maximum likelihood function taking into account the probabilistic density functions. For density estimation, the Probabilistic Principal Component Analysis (PPCA) mixture model [21] is usually better than the Gaussian Mixture Models (GMMs) when modeling high-dimensional data with a small number of samples. In order to capture the underlying local structures, cluster analysis has also been developed. The Fuzzy c -Means algorithms and its variants are closely related to GMMs [1]. Ichihashi *et al.* proposed a clustering algorithm, which is similar to the EM algorithm for

GMMs, by using the fuzzification technique with a K-L regularization information.

The local principal components can be obtained by local probability density and cluster analysis techniques. In density estimation, the likelihood function gives a measures of how well the probability density function (pdf) fits the given data set. The best fitting pdf for the data set will be defined by a parametric set that maximizes the likelihood. There is thus a need to find an estimate of these maximum-likelihood (ML) parameters. The EM algorithm is well known for this purpose. In cluster analysis, the set of samples is partitioned into subsets, where samples in the same cluster are somehow more similar than other samples in other cluster.

The fuzzy clustering algorithms are based on the minimization of an objective function defined by the sum of the weighted squares between the data points and the cluster prototypes as follows:

$$J = \sum_{i=1}^c \sum_{k=1}^N (\mu_{i,k})^m d^2(x_k, \eta_i) \quad (10)$$

where $\mu_{i,k}$ represents how the degree of membership of the observation x_k is in the i th cluster ($i = 1, \dots, c$) and $m \in [1, \infty)$ is a weighting exponent that determines the fuzziness of the resulting clusters (usually $m = 2$). Here, the prototype concept plays a central role. In the traditional FCM, a prototype represent a cluster center, and the fuzzy c -varieties (FCV) [1] is a linear fuzzy clustering technique that captures the local linear structures of the data sets and is regarded as a local PCA technique because the prototypes of cluster are estimated by solving the eigenvalue problem of fuzzy scatter matrices.

In the paper, the local PCA techniques are used for local subspace learning under a probabilistic framework, in which the responsibility of each data point for its generation is shared amongst all of the principal component analyzers instead of being assigned to only one analyzer. The task is solved in an iterative way by minimizing an objective function in which the distance is inversely proportional to the probability that the k th data point belongs to the i th cluster, $p(x_k|\eta_i)$. The data are assumed to be normally distributed random variables with expected value v_i and covariance matrix F_i .

The identification of a mixture of Gaussians that represents the $p(x_k|\eta)$ probability density function expanded in a sum of c clusters

$$p(x_k|\eta) = \sum_{i=1}^c p(x_k|\eta_i)p(\eta_i) \quad (11)$$

where $p(x_k|\eta_i)$ distribution generated by the i th cluster is represented by the Gaussian function

$$p(x_k|\eta_i) = \frac{1}{(2\pi)^{n/2} \sqrt{|F_i|}} \exp\left(-\frac{1}{2}(x_k - v_i)^T F_i^{-1} (x_k - v_i)\right) \quad (12)$$

and $p(\eta_i)$ is the unconditional cluster probability (normalized such that $\sum_{i=1}^c p(\eta_i) = 1$ holds, where η_i

represents the parameters of the i th cluster, $\eta_i = \{p(\eta_i), v_i, F_i | i = 1, \dots, c\}$. The clustering is based on the following $d^2(x_k, \eta_i)$ distance measure

$$\frac{1}{d^2(x_k, \eta_i)} = p(\eta_i) \times p(x_k|\eta_i) = \alpha_i \frac{1}{(2\pi)^{r/2} \sqrt{|A_i|}} \exp\left(-\frac{1}{2}(x_k - v_i)^T A_i^{-1} (x_k - v_i)\right) \quad (13)$$

which consists of two terms. The first $\alpha_i = p(\eta_i)$ represents the a priori probability of the clusters, while the second represents the distance between the cluster prototype and the data in the feature space and r is the rank of A_i ; distance norm of of the i th cluster. The case of Mahalanobis distance corresponds to $A_i = F_i$, where

$$F_i = \frac{\sum_{k=1}^N (\mu_{i,k})^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^N (\mu_{i,k})^m} \quad (14)$$

However, when the variables are highly correlated which is a common situation in most industrial plants, the F_i covariance matrices can be ill conditioned and cannot be inverted. To handle this problem, the PCA is used to project the correlated high-dimensional data onto a hyperplane which is useful for the analysis of the data. This mapping uses only the first few q nonzero eigenvalues and the corresponding eigenvectors of the $F_i = U_i \Lambda_i U_i^T$, covariance matrix, decomposed to the Λ_i matrix that includes the eigenvalues of F_i in its diagonal in decreasing order, and to the U_i matrix that includes the eigenvectors corresponding to the eigenvalues in its

column, $z_{i,k} = \Lambda_{i,q}^{-1/2} U_{i,q}^T U_{i,q} x_k$ when the hyperplane of the PCA model has adequate number of dimensions, the distance of the data from the hyperplane becomes small, so the projection of the data into this q -dimensional hyperplane does not cause significant reconstruction error

$$Q_{i,k} = \|(x_k - \hat{x}_k)\|^2 = x_k^T (I - U_{i,q} U_{i,q}^T) x_k \quad (15)$$

It is worth noticing that the competitive of principal component analysis is made up by a *projective step* \mathcal{F} and a *generative step* \mathcal{G} , as shown in Fig 1. The projective step \mathcal{F} is a nonlinear mapping from the measurement space to the latent subspace, giving the scores $z = \mathcal{F}(x)$ and the generative step \mathcal{G} is a de-mapping $\hat{x} = \mathcal{G}(z)$. Once, the mixture of PCA has been designed

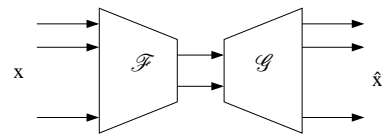


Figure 1: The competitive principal component analysis model which is made up by a *projective step* \mathcal{F} and a *generative step* \mathcal{G} .

with respect of the reconstruction principle, the objective is to perform nonlinear statistical process monitoring. This task can be carried out in two ways or strategies, as shown in Fig. 2. In the first strategy, we define a *global fuzzy measure* as a combination of the local

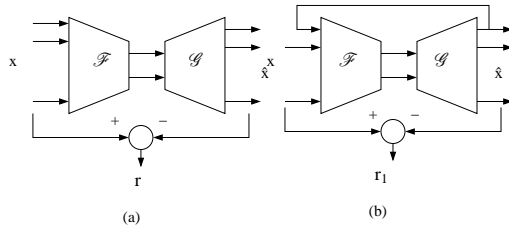


Figure 2: Competitive principal component analysis for fault diagnosis (a) Computation of residual signals (b) Nonlinear variable reconstruction.

statistics, by considering that the j th variable (or sensor) is faulty and define the vector:

$$\hat{x}_{j,(i)} = [x_{1:j-1} \ z_{j,(i)} \ x_{j+1:s}]^T \quad (16)$$

in which $z_{j,(i)}$ is the reconstructed value for the j th variable by the i th local PCA model. The global vector is then given by

$$\hat{x}_j = \sum_{i=1}^c \mu_i (\hat{x}_{j,(i)}) \hat{x}_{j,(i)} \quad (17)$$

In this case, the global SPE is defined by

$$SPE = \sum_{i=1}^c \mu_i (\hat{x}_{j,(i)}) SPE_i \quad (18)$$

For the fault detection task, we introduce a *soft* variant of the control limits δ_{α}^2 , defined by the following equation:

$$\delta_{\alpha}^2 = \sum_{i=1}^c \mu_i (\hat{x}_{j,(i)}) \delta_{\alpha,i}^2 \quad (19)$$

where $\delta_{\alpha,i}^2$ is the control limit in the residual subspace \tilde{S}_i of the i th local model. Thus, we define the quantity E_j as follows:

$$E_j = \sum_{i=1}^c \mu_i (\hat{x}_{j,(i)}) E_{j,(i)} \quad (20)$$

which represents actually a specific global SPE for the j th reconstructed variable. $E_{j,(i)}$ is the local SPE of the i th local PCA for the j th reconstructed variable. In order to isolate sensor faults, we propose the following new validity index:

$$\eta_j^2 = \frac{E_j}{SPE} \quad (21)$$

This new index can be considered as a generalization of the traditional SVI derived from a single PCA model [5] to a mixture of local linear PCA models.

Another strategy to monitor the sensors is to make use of the residuals from the statistics of the j th sensor in the form of a difference between the measurements and the reconstructed variables:

$$r_{j,(i)} = x_j - \hat{x}_{j,(i)} \quad (22)$$

The residuals can be used for the fault isolation.

4. Simulation results

In order to demonstrate the effectiveness of the proposed scheme, a biological system, which consists of a continuous aerated bioreactor for waste water treatment in pulp and paper industry is selected as a typical nonlinear process, Fig. 3. The bioreactor contains a mixed microbial population growing in a blending of two types of substrates, an energetic and an xenobiotic one. The main pollutant is the xenobiotic substrate because the energetic substrate is easily biodegradable [22]. Let $c(t)[g/l]$: the biomass concentration.

$s(t)[g/l]$: the concentration of the xenobiotic pollutant substrate.

$e(t)[g/l]$: the concentration of the energetic substrate.

$u_1(t)[l/s]$: flow rate of clear water feeding the reactor.

$u_2(t)[l/s]$: flow rate of waste water. The waste water contains certain maximal substrate concentrations:

$s_a^{max}[g/l]$ and $e_a^{max}[g/l]$.

μ_{sm} and μ_{em} : the growth rates of the substrates.

$Y_{c/s}$ and $Y_{c/e}$: yield coefficients.

K_s and K_e : Michaelis-Menten constants.

a_s and a_e : Inhibiting coefficients.

The differential equations describing the behaviour of the process are:

$$\begin{cases} \dot{c}(t) = \mu_{sm} \left(\frac{s(t)}{K_s + s(t) + a_e e(t)} + \mu_{em} \frac{e(t)}{K_e + e(t) + a_s s(t)} \right) c(t) \\ \quad - u_1(t) c(t) - u_2(t) c(t) \\ \dot{s}(t) = - \frac{\mu_{sm}}{Y_{c/s}} \frac{s(t)}{K_s + s(t) + a_e e(t)} c(t) - u_1(t) s(t) \\ \quad + (s_a^{max} - s(t)) u_2(t) \\ \dot{e}(t) = - \frac{\mu_{em}}{Y_{c/e}} \frac{e(t)}{K_e + e(t) + a_s s(t)} c(t) - u_1(t) e(t) \\ \quad + (e_a^{max} - e(t)) u_2(t) \end{cases}$$

The above analytic model is used as a virtual system

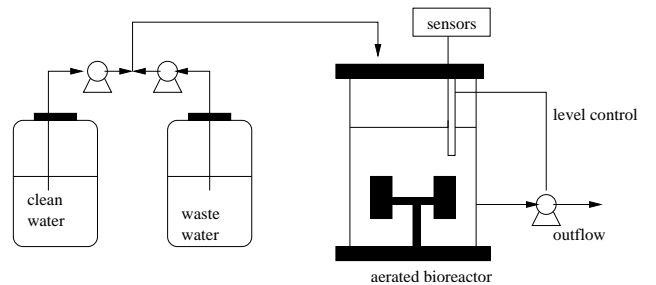


Figure 3: Schematic representation of the aerated bioreactor for wastewater treatment

to generate data in an open loop mode. The collected data for the normal condition is used to build the monitoring model of competitive principal component analyzers. The data vector of the k th sample is given by

$$x(k) = [c(k) \ s(k) \ e(k) \ u_1(k) \ u_2(k)]^T$$

The reconstruction of the five variables in the fault free case, where each variable is reconstructed from the remaining sensors is shown in Fig. 4. It is clear that despite the nonlinearity of the process, the multiple PCA model is able to reconstruct the data. Fig. 5 shows the performance of a mixture of two local PCA, i.e. the

number of principal components $q = 2$. Fig. 6 shows the statistics of the local models and their combination for the sensor 3 for a fault on the first variable from the sample 50. In the case of a fault in sensor 3, Fig. 7 gives the residuals obtained after the reconstruction of the five variables. The results indicate clearly that the task of fault detection and isolation can be carried out in an effective way since the generated residuals are structured.

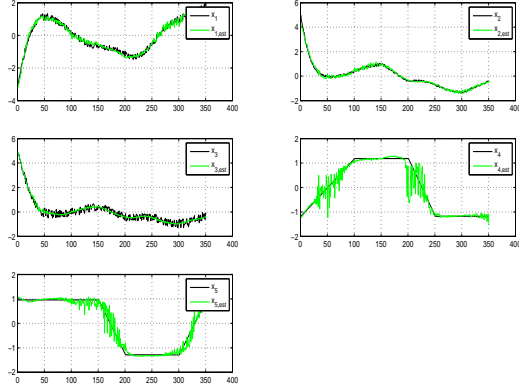


Figure 4: The reconstruction of the variables in the fault free case. The variables x_j and their reconstructed values \hat{x}_j , $j = 1, \dots, 5$.

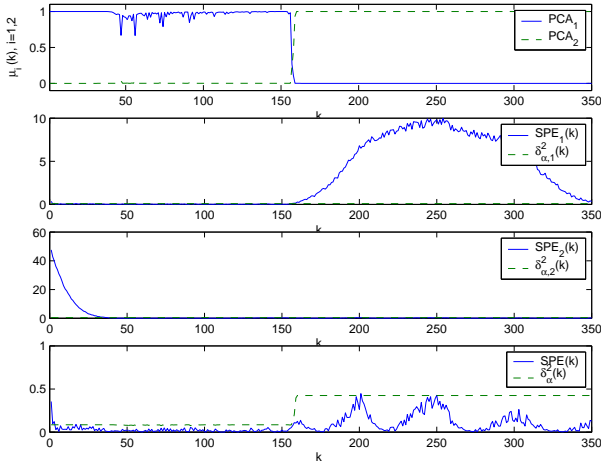


Figure 5: Performance of a mixture of PCA for a training data composed of 300 points (a) the activation degrees of the two local models (b) global SPE (c) local SPE of the PCA_1 model (d) local SPE of the PCA_2 model.

5. Conclusion

This paper has studied the monitoring of nonlinear processes based on competitive principal component analyzers experts. The proposed scheme allows the reconstruction of a faulty sensor by using the information provided by other sensors. The on-line implementation of the indicators requires the use of filters. The simulation

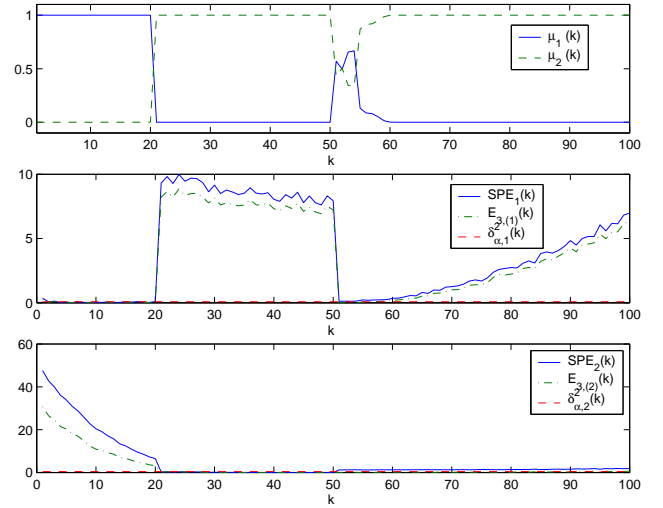


Figure 6: Case of offset defect in sensor 3 .

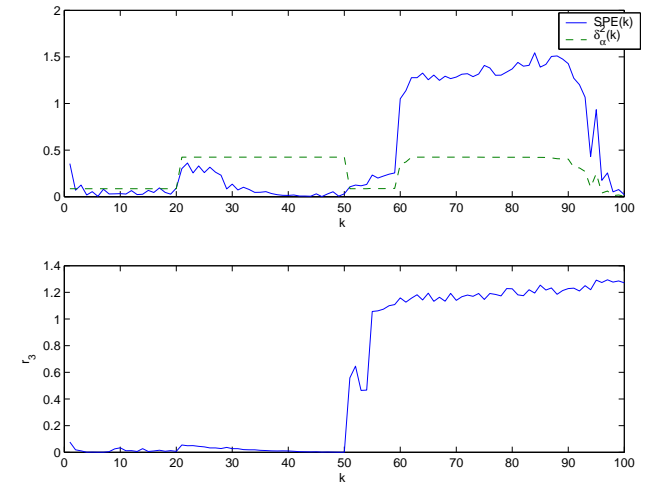


Figure 7: The residuals derived in the case of an offset in the sensor 3 starting at the sample 50.

study on a biological depollution process has demonstrated that it is reliable than linear PCA. Further work of the proposed nonlinear monitoring scheme in dynamic context is needed.

References

- [1] Bezdek, J. C.: Pattern recognition with fuzzy objective function algorithms. Plenum Press, (1981)
- [2] Box G. E. P.: 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
- [3] Cao, J., Gertler, J.: Partial PCA-Based optimal structured residual design for fault isolation. American Control Conf. (2004) 4420–4425
- [4] Dempster, A. P., Laird, N. M., Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm. J. of the Royal Statistical Society, Series B. **39** (1979) 1–38

- [5] Dunia, R., Qin, J., Edgar, T. F., McAvoy, T. J.: Use of principal component analysis for sensor fault identification. *Computers Chem. Eng.* **20** (1996) 713–718
- [6] Frank, P. M.: Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. *Automatica*. **26** (1990) 459–474
- [7] Gertler, J.: A new structural framework for parity equation based failure detection and isolation. *Automatica*. **26** (1990) 381–388
- [8] Gertler, J.: Fault detection and diagnosis in engineering systems, Marcel Dekker. New York, (1998)
- [9] Gertler, J., Li, W., Huang, Y., McAvoy, T.: Isolation-enhanced Principal Component analysis. *AIChE Journal*. **45** (1999) 323–328
- [10] Hastie, T., Stuetzle, W.: Principal curves. *J. American Statistical Association*, **84** (1989) 502–517
- [11] Hinton, G. E., Dayan, P., Revow, M.: Modeling the manifolds of images of handwritten digits. *IEEE Trans. on Neural networks*. **8** (1997) 65–74
- [12] Isermann, R.: Supervision, fault-detection and fault-diagnosis methods. An introduction. *Control Eng. Practice*. vol. 5, pp. 639–652, 1997.
- [13] Jackson, J. E. and Mudholkar, G. S., Control procedures for residuals associated with principal component analysis, *Technometrics*, **21** (1979)
- [14] Kambhatla, N. Leen, T. K.: Dimensionality reduction by local principal component analysis. *Neural computation*. **9** (1997) 1493–1516
- [15] Kourti, T., MacGregor, J.: Statistical Process Control of Multivariate Processes. *Control Eng. Practice*. **3** (1995) 403–414
- [16] Kramer, M. A.: Autoassociative Neural networks. *Computers Chem. Eng.* **16** (1992) 313–328
- [17] McAvoy, T. J., Ye, N.: Base control for the Tennessee Eastman problem. *Comput. and chem. Eng.* **18** (1994) 383–413
- [18] Qin, S. J., McAvoy, T. J.: Nonlinear PLS Modeling using Neural Networks. *Comput. and chem. Eng.* **16** (1992) 379–391
- [19] Roweis, S.: EM algorithms for PCA and SPCA. *Advances in Neural Information Processing Systems*, Eds. M. I. Jordan, M. J. Kearns and S. A. Solla, MIT Press. **10** (1998) 626–632
- [20] Simani, S., Fantuzzi C., Patton R. J.: Model-based fault diagnosis in dynamic systems using identification techniques. Springer, London, (2003)
- [21] Tipping, M. E., Bishop, C. M.: Mixtures of probabilistic principal component analysis. *Neural Computation*. **11** (1999) 443–482
- [22] C.BEN YOUSSEF, B. DAHHOU: Multivariable adaptive control and estimation of a nonlinear wastewater treatment process. 3rd IEEE Mediterranean Symposium on New Directions in Control and Automation, Limassol (Chypre), July 11-13, (1995) 220–224
- [23] Webb, A. R.: An Approach to non-linear principal components analysis using radially symmetric kernel functions Nonlinear PLS Modeling using Neural Networks. *Statistics and Computing*. **6** (1996) 159–168
- [24] Wilson, D. J. H., Irwin, G. W., Lightbody, G.: RBF Principal manifolds for process monitoring. *IEEE trans. Neural Networks*. **10** (1999) 1424–1434