Use of NLPCA for Sensors Fault Detection and Localization Applied at WTP

K. Bouzenad, M. Ramdani, N.Zermi,K.Mendaci
Laboratory of Automatic and Signals Annaba (LASA)
Department of Electronics
Badji-Mokhtar University
BP. 12, 23000, Annaba
Algeria
khaled.bouzenad@hotmail.fr, messaoud.ramdani@univ-annaba.org, {naili.narima, khaled.mendaci}@gmail.com

ABSTRACT: *Principal Components Analysis (PCA) has been intensively studied and is widely applied in industrial process monitoring. The main purpose of using PCA is the dimensionality reduction by extraction of the feature space that still contain the most information in the original data set. Despite its success in this field, the most important obstacle faced is the sensitivity to noise, also the fact that the majority of collected data from industrial processes are normally contaminated by noise makes it unreliable in some cases. To overcome these limitations, several strategies have been used. One of these has been interested to combine the robustness theory with PCA method, such theory sonsists in robustifying the existing algorithms against noise or outliers. Fuzzy Robust Principal Components Analysis (FRPCA) is one of the results for such combination that acheive better result compared with the classical method. In this work the RFPCA method is used and compared with the classical one to monitoring a biological nitrogen removal process. The obtained results demonstrate the performances superiority of this method compared with the conventional one.*

Keywords: Process Monitoring, Multivariate Statistical Process Control, Fault Diagnosis, NLPCA, Water Treatment Plant, Sensor Validity Index

Received: 7 May 2013, Revised 1 June 2013, Accepted 4 June 2013

© 2013 DLINE. All rights reserved

1. Introduction

With the increased demand on water supply over the last century due to population growth, the adoption of new technology to ensure water quality at lower cost is essential. In a typical process of water treatment for drinking purposes, raw water from various sources is chemically treated, filtered and disinfected. The type of treatment it then undergoes depends on the source and the quality of its water. In this field, the searched efficacy in terms of effluent quality and economies of treatment costs has made necessary the modeling, identification and monitoring of biological treatment processes. The drinking water treatment processes allow well and truly significantly improve the quality of raw water, the classical complete treatment of water is performed in several steps: oxidation, clarification and disinfection, whose some not necessary for the cleanest waters. These water treatment processes must meet specific conditions that will ensure their efficiency. Every element must be taken into account in the correct process operation of the treatment chain, the sensors in the heart of metrological process (to measure: the flow, pressure, level, temperature, pH, oxygen level, etc.) have become now ubiquitous, at the same time, the steps: quality and research zero fault, must ensure the efficiency of the process. Indeed, in order to operate correctly this control system, the fault diagnosis is an essential element for any proceeding of automation process. The objective of this paper, is to validate a

data collection delivered by the sensors useful to the control water treatment plant using a diagnosis procedure based on the NLPCA, we will interest by an application of sensor fault detection and localization particularly on the station of Azzaba, situated in the northeast of Algeria,the principal component analysis is then a modeling tool used to perform the monitoring. In order to construct a model, we use a data matrix constitutes the measures data collected by sensors available in the installation, the PCA is then a statistical technique which consists simultaneously to identify the linear relations between the variables of the process and to analyze and reduce the dimensionality of big size dataset. The PCA consists to replace variables set by new variables uncorrelated two to two, of a smaller size and maximum variance, these new variables called principal components.

2. Principal Component Analysis

The Principal Component Analysis (PCA) is a method of data analysis family and more generally of multivariate statistics, which consist to transform a variables interrelated (called "*correlated*" in statistics) into new linearly uncorrelated variables from each other. These new variables are called "*principal components*" or axes. It allows the practitioner to reduce the information to more limited number of components compared to the initial number of variables. It is both a geometric approach (representation of variables in a new geometric space according to maximum inertia direction) and statistical approach (search independent axes explaining the most variability "*variance*" of data). Then, when we want to compress a set of random variables, the first axes of the PCA are a best choice, in terms of inertia or explained variance.

We called principal axes the direction axes of the eigenvectors of the covariance matrix of the process variables, where the first axis that is associated with the biggest eigenvalue and the second axis orthogonal to the first, is associated with the second biggest eigenvalue...etc, and then the last axis is that associated with the lowest eigenvalue. The two or three first principal axes constitute the directions of the reduced space "*Principal space*" which belongs to the original data space.

This approach is based on the projection of the original data collection on the new lower-dimensional space, and from the projection matrix, we can estimate our original information while minimizing the estimation error, in this sense the PCA can be considered as a minimization technique of the estimation error, otherwise, the estimated data must be approximately near to the original values.

However, the principal component analysis is a method of reducing the number of variables necessary to represent geometrically a phenomenon. From a collection of n objects (individuals) in a space of m descriptors (variables), its goal is to find a representation in a reduced space of dimensions ($\ell << m$) which preserves "*the best summary*." the reduction is only possible if the original variables m are not independent. This notion of independence is measured using the correlation coefficients, it should that these coefficients are not null. PCA is a linear method called factorial, because the reduction is not a selection of the basic variables but by a definition of new variables (principals) obtained by combining the original variables. The mathematical tool associated with the PCA method is therefore based on linear algebra and calculation matrix.

Mathematically, we can interpret each variable as a vector of R^n. We denote $x_1, x_2, \ldots x_m$ these m vectors of R^n. We can also save the data in a matrix X of size ($n * m$) with a coefficient x_{ij} representing the ith position of the jth variable. The column vectors of X are $x_1, x_2, \ldots xm$ and for $j = 1 \ldots m$, $x_j \in R^n$ represents the jth variable measurements at all instants. Similarly, we denote $e_1, e_2, \ldots e_n$ the line vectors of X and $i = 1 \ldots n$, $e_i \in R^m$ represent the time measures at sample i of m variables. This data contains the measurements of variables (measuring temperature, flow,) at different instants (called events). When there are only ℓ variables that it is easy to represent the dataset on a graph to ℓ dimensions. The tracing can be deduced:

- The absence of correlation between the variables
- The existence of a strong link
- The emergence of sub-populations or groups of variables.

If there is m variables measured and n events by variable, we search a feasible graphical representation (that is to say, on a plan) the more faithful to reality as possible (that is to say, which minimizes the distortions). For this, we use an orthogonal projection on a selected plan in order to maximize the square mean of the distance between the variables.

2.1 Mathematical calculation

We suppose X, a matrix of m variables and n number of observations for each variable. Where, the number of lines represents

the data dimension. Beforehand, to make the independent result of the used units for each variable, a pretreatment is essential to center and reduce the variables. For this we assume, that we subtract from each column of the start set the mean μ_j and we divide on the standard deviation σ_j. We then obtain a new normalized matrix Y with its mean centered at zero. The calculation of the PCA is defined in this way: Each column x_j of the new matrix centered reduced is given by:

$$Y_j = \frac{X_j - M_j}{\sigma_j} \tag{1}$$

Once the data has been centered and reduced, the data correlation matrix, Σ, can be calculated as:

$$\Sigma = \frac{1}{N-1} Y^T Y \tag{2}$$

To calculate $Y\,Y^T$ returns to produce a matrix containing the sums of the deviations. $Y^T Y$ multiply by $1/(n-1)$ can obtain the matrix Σ where the elements located on the diagonal correspond to the variance σ_{ij}^2 and other to the covariance $\sigma_i\,\sigma_j$. Σ is often called the covariance matrix of X. The eigenvalues of the covariance matrix Σ represents the projections variances t of data on the directions shown by the eigenvectors pi (i =, ...m). In conclusion, the direction in which the data projection variance of the vector X is maximal, is represented by the eigenvector pi corresponding to the maximum eigenvalue. The sub-vector space of ℓ dimension which ensures the maximum dispersion of observations is defined by an orthonormal basis consisting ℓ eigenvectors corresponding to the biggest eigenvalues of the matrix Σ. It is therefore possible to reduce the dimension of the data representation by retaining the previous expression only the $t_j p_j$ (j =1,. . .) terms associated with the biggest eigenvalues . ℓ The PCA therefore determines an optimal transformation of the data matrix x:

$$T = XP \;\; And \;\; X = TP^T \tag{3}$$

Where $T \in R^{n\times m}$ et $P \in R^{m\times m}$ are the matrixes of the principal components and the corresponding eigenvectors outcome from the spectral decomposition of the covariance matrix Σ. the relations (3) find their interest when we reduce the dimension representation space. Once the number $\ell < m$ components to retain is determined, the data matrix X can be approximated. For this, the eigenvectors matrix is partitioned as:

$$P = \widetilde{P}\widehat{P} \;\; where \;\; \widehat{P}\;\Sigma\; R^{n\times\ell} \tag{4}$$

The first ℓ eigenvectors $\widehat{P}$ makes the principal space while the $(m-\ell)$ last eigenvectors $\widetilde{P}$ makes the residual space. From the equation (3), we then can explain the part of the data explained by the first ℓ eigenvectors and the residual part explained by the remaining components:

$$\widetilde{X} = XC \;\; where \;\; \widehat{C} = \widehat{P}\widehat{P}^T \tag{5}$$

and

$$E = X - \widehat{X} = X(1 - \widehat{C}^{(\ell)}) \tag{6}$$

The identification of the PCA model is therefore to estimate its parameters by the decomposition on eigenvectors and eigenvalues of the matrix Σ and to determine the number ℓ of principal components to retain. In the absence of noise on the measures, the null eigenvalues of Σ indicate the existence of linear relations between the components of X. In the presence of noise on the measures, the smaller eigenvalues compared to the other indicate the existence of linear or quasilinear relations between the different components of X. we see here the key role played by the ℓ number components in determining the relations of redundancy between variables and also the difficulty of determining this structural parameter.

3. Non Linear Principal Component Analysis

The principal component analysis has interesting properties for industrial processes monitoring. Unfortunately, in the industry, the most physical systems has a non-linear behavior and then the linearity property of the linear PCA, Figure (figure 2), pose always the problem of the inaptitude of this method to represent nonlinear data, since it is a linear projection and only the linear dependencies between the variables can be revealed.

Hastie [1] proposes an approach for a generalization of PCA in the nonlinear case based on the principle of principal curves, this generalization is performed by a projection of data on curves instead of lines. Kramer [2] proposes the extension of the non-linear principal component analysis (NLPCA) using a neural network with five layers whose the weights are calculated by learning by minimizing the squared error between the network inputs and outputs.

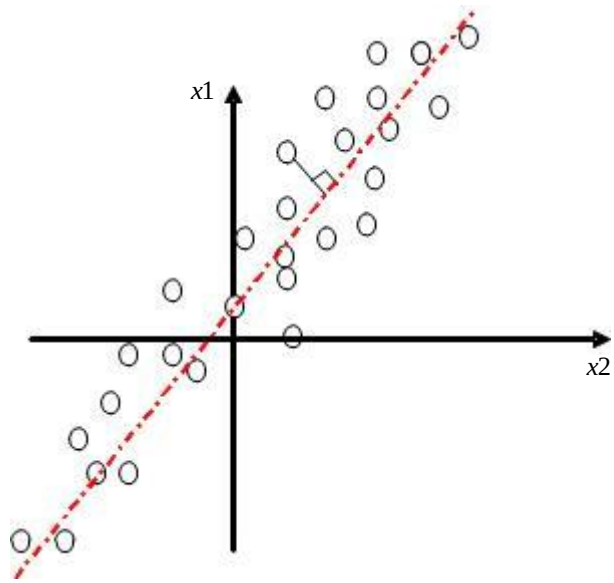


Figure 1. Principle of linear PCA

This section is dedicated to non-linear extension of PCA (NLPCA) which allows to extract the linear and nonlinear relations between variables. The figure (Figure 3) shows the principle of the general PCA, the overall model is composed of two sub-models: a sub-model of data compression projects the data of dimension m into principal components space of dimension ℓ and the second sub-model performs the reverse operation, namely a projection of $\Re^{\ell}$ to $\Re^{m}$. Thus, in the linear case these two sub-models are characterized by the eigenvectors orthogonal matrix of the data correlation matrix $\widehat{P}$ and the overall model is given by the projection matrix $C_{\ell} = \widehat{P}\widehat{P}^{T}$.

In the nonlinear case, the goal is to find two non-linear functions Ψ and Φ. Φ is the nonlinear model of compression to calculate nonlinear principal components from the data, and Ψ is the decompression nonlinear model for estimating the original variables from the nonlinear principal components given by the compression model, the projection model gives from the data matrix X the principal components T and the non-linear model allows to give the matrix $\widehat{X}$ an estimate of X, based on principal components T. In this case, we can write:

$$T = \Phi(X) \tag{7}$$

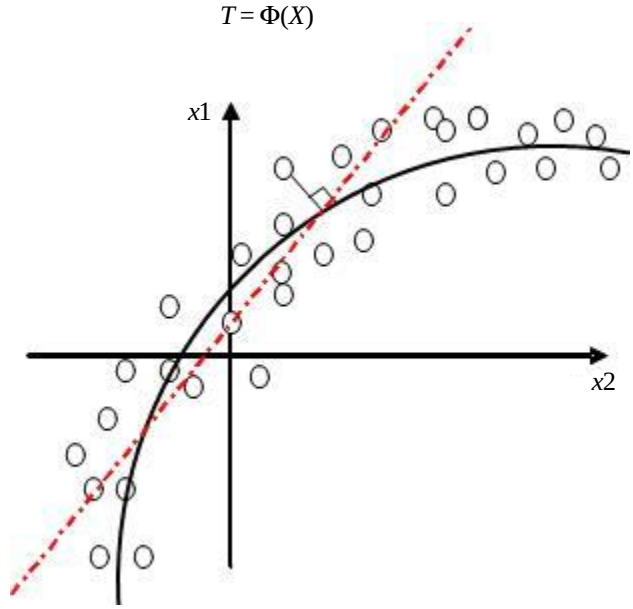


Figure 2. Difference between linear and nonlinear PCA

Where Φ is a non-linear function equivalent to the eigenvectors matrix "P" of the linear PCA. While:
$X \in R^{n\times m}$ the data matrix, and $T \in R^{n\times \ell}$ the principal components matrix. The decompression model provides an estimate $\hat{x}$ of x from the nonlinear components t (such as x and t are the lines of X and T, respectively):

$$\hat{x} = \Psi(t) \tag{8}$$

Thus, the data matrix X containing m variables can be expressed as a function of ℓ first nonlinear components.

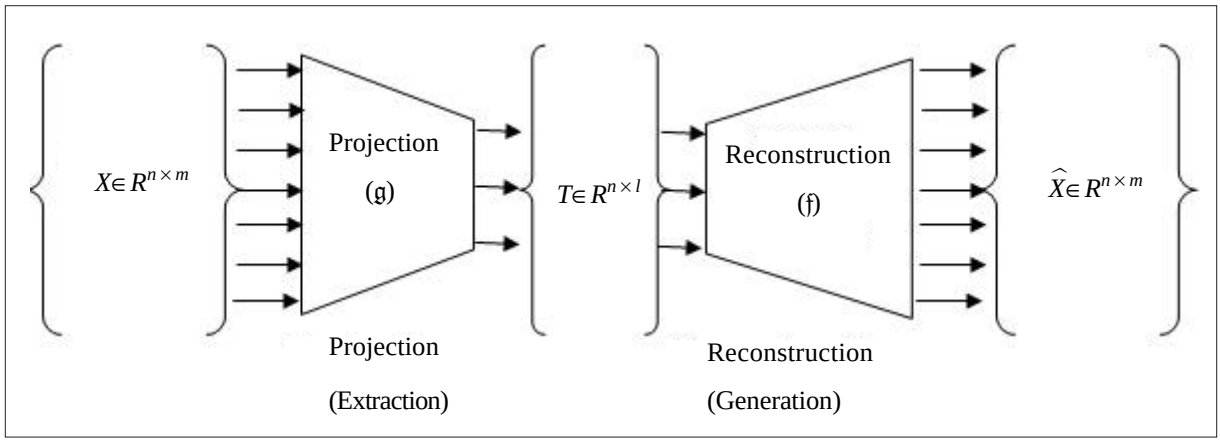


Figure 3. Basic principles of PCA model

And then the data matrix X can be represented by the estimate $\hat{X}$ plus the estimation error $\tilde{X}$ (residual matrix) :

$$X = \hat{X} + \tilde{X} = \Psi(T) + E \tag{9}$$

Where $T = \Phi(X)$ is the nonlinear principal components matrix such as $T = [T_1, ...T_\ell]$ and E the residuals matrix. The problem is then to identify the nonlinear projection functions Φ and Ψ. Where $\hat{X} \in R^{n\times m}$ the data reconstruction matrix with $\hat{X} = \Psi(t)$, Ψ represents the nonlinear function of reconstruction or generation. We note $E(k)$ the squared error obtained from the resulting errors on m output neurons, by a learning algorithm of neural network based on the principle of optimization, we search to minimize, by nonlinear optimization methods, the following cost function:

$$\min \sum_{k=1}^{N} \| X(k) - \hat{X}(k)^2 \| = \min \sum_{k=1}^{N} \| X(k) - \Psi(\Phi(X(k)) \|^2 \tag{10}$$

4. Neural Approach of NLPCA

The nonlinear principal component analysis (NLPCA) based on neural networks, has known a considerable progress and interest in recent years and has been widely used in the field of diagnosis. In this section, we will present a neural network with five layers for extracting nonlinear principal components.

4.1 Auto-associative neural network

Non linear principal component analysis (NLPCA) which is an extension of linear PCA, has a particular interest in the last years. Most of researchers use a neural approach to calculate the NLPCA model proposed by Kramer [2]. In the case of a single non-linear principal component, the structure of such a network is illustrated in figure (Figure 4). To make the NLPCA, the auto-associative network contains three layers between the input and output variables. A transfer function Ξ_1 makes a projection of the input column vector of dimension m, to the first hidden layer (coding layer), represented by $h_j^{(x)}(j = 1,..., r)$ a column vector of dimension r (r is the number of neurons in the first hidden layer):

$$h_j^{(x)} = \Xi_1 \left(\sum_{i=1}^{m} v_{ij}^{(x)} + b_j^{(x)} \right) \tag{11}$$

$V^{(x)}$ is the weight matrix of dimension ($r \times m$), is a vector containing the r bias parameters. The second transfer function Ξ_2 projects the outputs data of the first hidden layer (coding layer) to "*bottleneck layer*" containing a single neuron, which

represents the nonlinear principal component t. The transfer function Ξ_1 is generally nonlinear (using the hyperbolic tangent function or sigmoid function), while the Ξ_2 function is the identity function ($\Xi_2(x) = x$):

$$t = \Xi_2\left(\sum_{i=1}^{r} w_j^{(x)} h_j^{(x)} + b^{-(x)}\right) \tag{12}$$

Next, the transfer function Ξ_3, which is a nonlinear function, projects the data from t to the latest hidden layer (decoding layer): $h_j^{(t)}$ ($j = 1,..., r$) where r represents the number of neurons in the third hidden layer:

$$h_j^{(t)} = \Xi_3\left(w_j^{(t)} t + b_j^{(t)}\right) \tag{13}$$

The last transfer function Ξ_4 is the identity function which projects the outputs data from $h_j^{(t)}$ to $\widehat{x}$: the output column vector of dimension m:

$$x_i = \Xi_4\left(\sum_{i=1}^{r} v_{ij}^{(t)} h_j^{(t)} + b_j^{(t)}\right) \tag{14}$$

The cost function $E = \| X(k) - \widehat{X}(k)^2 \|$ is minimized to find the optimal values of the $V^{(x)}$, $b^{(x)}$, $w^{(x)}$, $b^{-(x)}$, $w^{(t)}$, $b^{(t)}$, $v^{(x)}$, and $b^{-(t)}$

It should be noted that the extraction of principal components can be done in two ways. The first is to extract the principal components sequentially with a single neuron in the middle layer "*bottleneck layer*" (sequential NLPCA) (Figure 4). The second is to extract the desired components simultaneously inserting the neurons in the middle layer (NLPCA parallel or simultaneous).

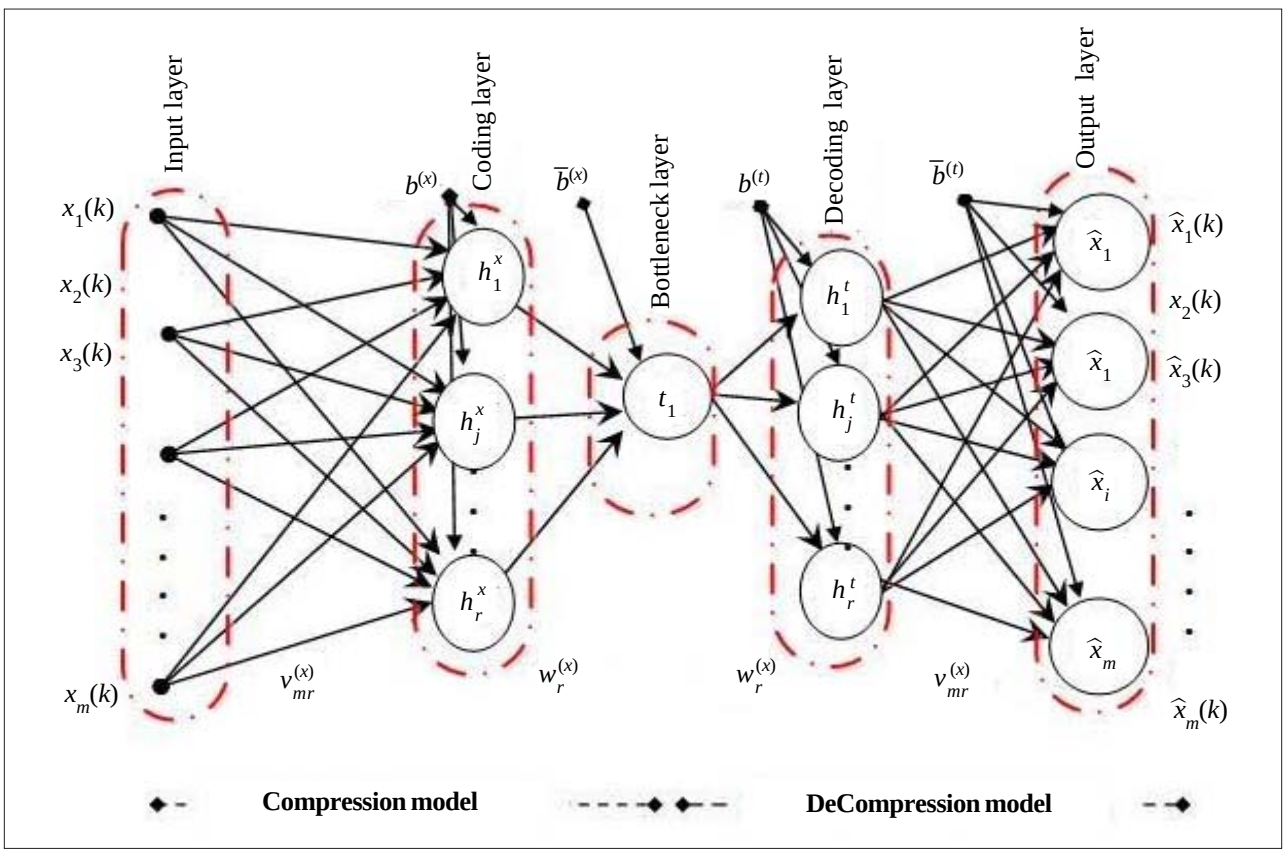


Figure 4. Auto-associative network with a five layer for extracting one nonlinear principal component

5. Sensors Fault Detection based on Spe Indicator

Once the NLPCA model is obtained, we present in this section its use in sensor fault detection. The indicator of detection *SPE* (Squared Prediction Error) performs the fault detection in the residual space. At sample k, is given by:

$$SPE(k) = e(k)\, e(k)^T \tag{15}$$

$$e(k) = x(k) - \hat{x}(k) \tag{16}$$

Where, $e\,(k)$ represents the vector of estimation errors. The process is considered abnormal operating (presence a default at the sample k) if:

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

Where α^2 is the upper control limit of $SPE(k)$, determined theoretically by BOX [3] , such that $\theta_i = \sum_{j=\ell+1}^{m} \lambda_j^i$, for $i = 1, 2, 3$ and λ_j is the jth eigenvalue of the matrix Σ to the power of i. The upper control limit theory, for the confidence threshold α given, is then $\delta_\alpha^2 = g\chi^2_{h,\,\alpha}$:

Where $g = \theta_2/\theta_1$, h = integer (θ_1^2/θ_1) (integer(z) is integer number of z) and $\chi^2_{h,\,\alpha}$ is the distribution of the χ^2 with h degree of liberty.

6. Localization

After detecting a fault, it is necessary to identify the faulty sensor or sensors, it is through the principle of fault localization, for this, among the methods used in fault localization within of NLPCA, we find the approach partial NLPCA which uses the benches of NLPCA models with different sets of variables, the contribution plots to the SPE of each variable it is an approach to localize the faults, such as the variable with the highest contribution is considered as a default variable. In this paper we present two methods to identify the faults:

6.1 *The first method*

The localization method based on *the contribution plots* $cont_j^{SPE}$ of jth variables to the SPE detection index In this case the contribution of jth variable at time k is defined by the equation:

$$cont_j^{SPE}(k) = (e_j(k))^2 = (x_j(k) - \hat{x}_j(k))^2 \tag{18}$$

6.2 Second method

The localization via a *sensor validity index* (*SVI*) [4] This method is based on the principle of reconstruction. Consists to suspect a faulty sensor and reconstruct the value of the measure based on the PCA model already calculated and the measurements of other sensors. The procedure is repeated for each sensor. The localization is performed by comparison of the detection index before and after reconstruction. The Sensor Validity Index is a measure of sensor performance. It should have a standarized range regardless of the number of principal components, noise, measurement variances or type of faults. The SVI should also distinguish the abnormal operational conditions from the sensor fault situation. The ratio of SPE_j and the SPE can provide these desired properties for the identification of a sensor fault:

$$\eta_j^2 = \frac{SPE}{SPE_j} \tag{19}$$

Where the SPE is the global squared prediction error calculated before reconstruction and SPE_j is calculated after the reconstruction of the jth sensor.

7. Application

In this section, we will particularly interest by an application of sensor fault diagnosis used in Water Treatment Plant of AZZABA which was built recently with a capacity of 300 *l*/*s*. This WTP takes charge the waters of these areas and contribute to improve the quality of water for 80000 subscribers. The monitoring is done continuously through sensors for process monitoring, measuring : turbidity, pH, pressure, level, temperature, oxygen level, etc. In order to model the process using the NLPCA, we collect dataset "*already available*" measured by sensors on line during normal functioning modes of the process for 300 days. After modeling, the next step of the study is to detect and locate the sensors fault, the following figures present the results simulation obtained during our studies about the diagnosis applied on the process of water treatment.

7.1 Simulation Results

In this context, we will give the results obtained from the developed procedure of diagnosis for fault affecting the sensors.

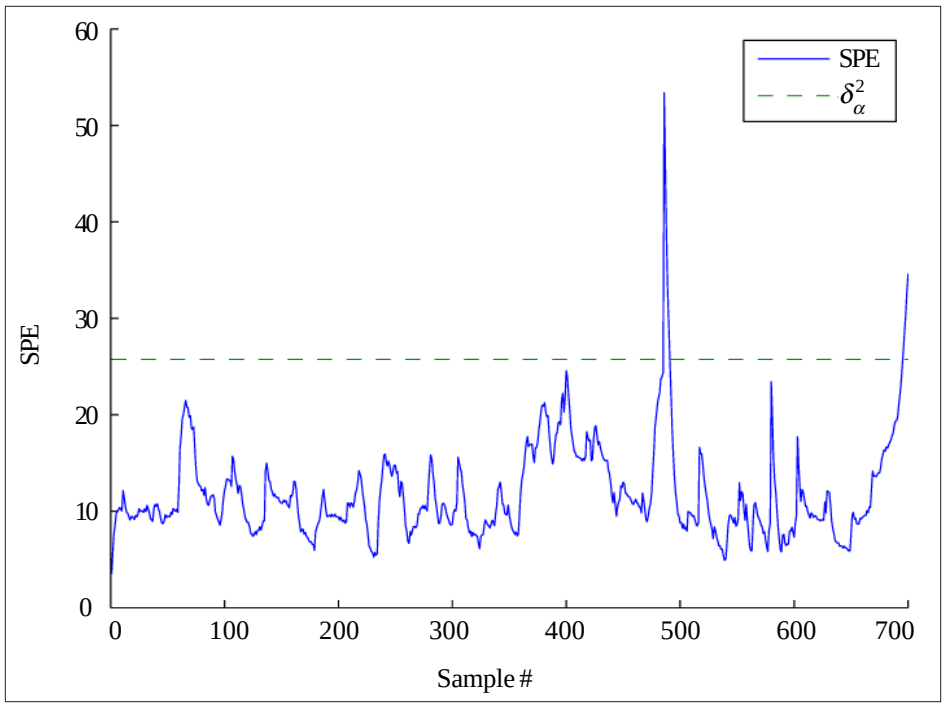


Figure 5 Evolution of the SPE index during normal stat (no fault)

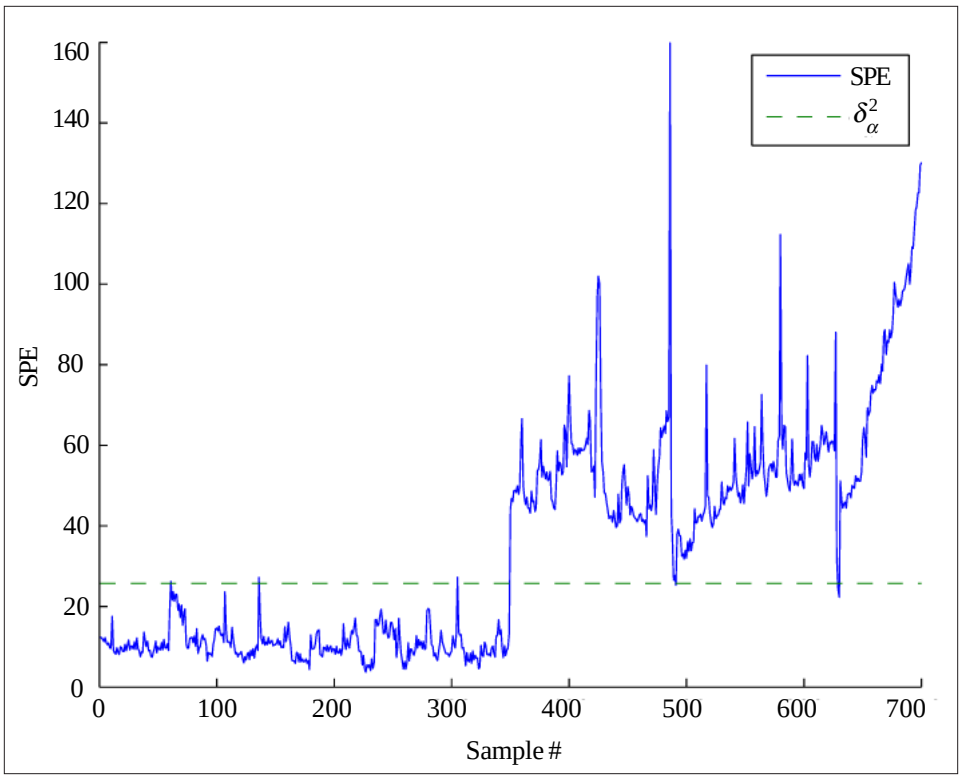


Figure 6. Evolution of the SPE index with a default in the variable *PH* at sample 350

The figure (Figure 6) Presents the evolution of *SPE* indicator, that exceeds the upper control limit, it is clear that we detect a fault, at sample 350. For localization anything we can observe in this figure, for this we want to locate the incriminated variable,

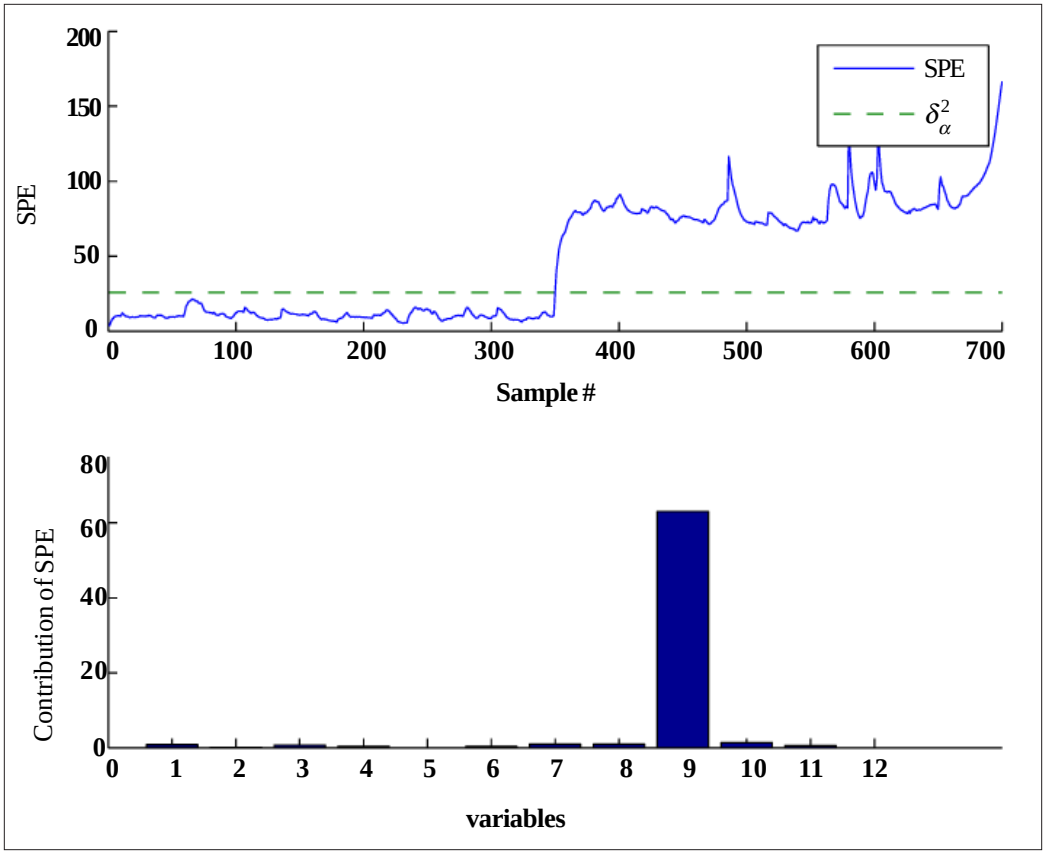
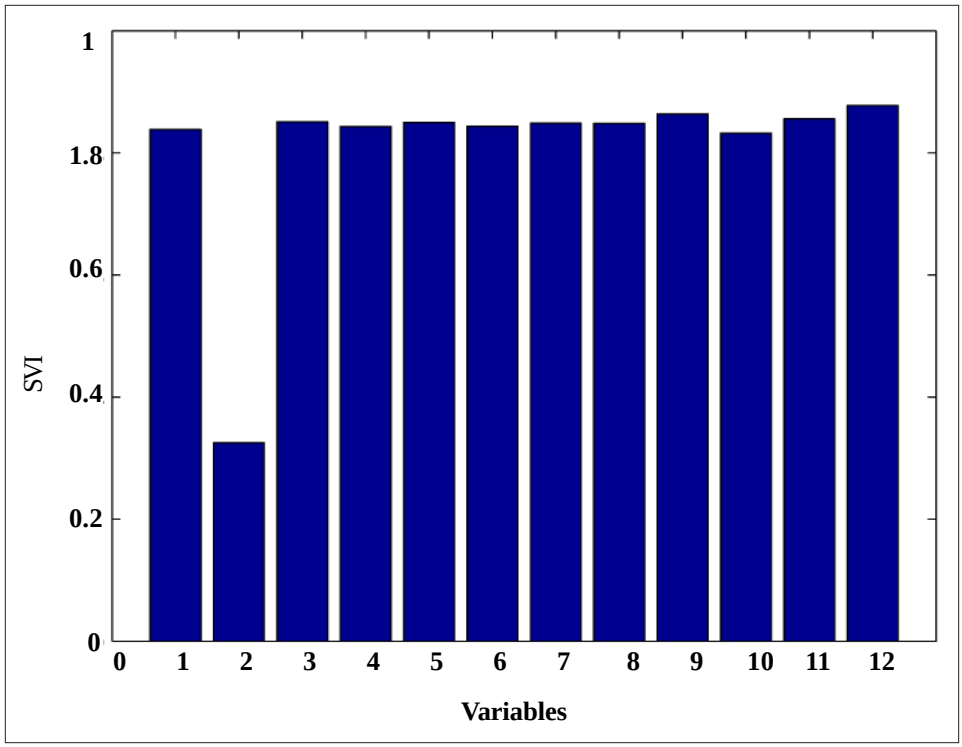
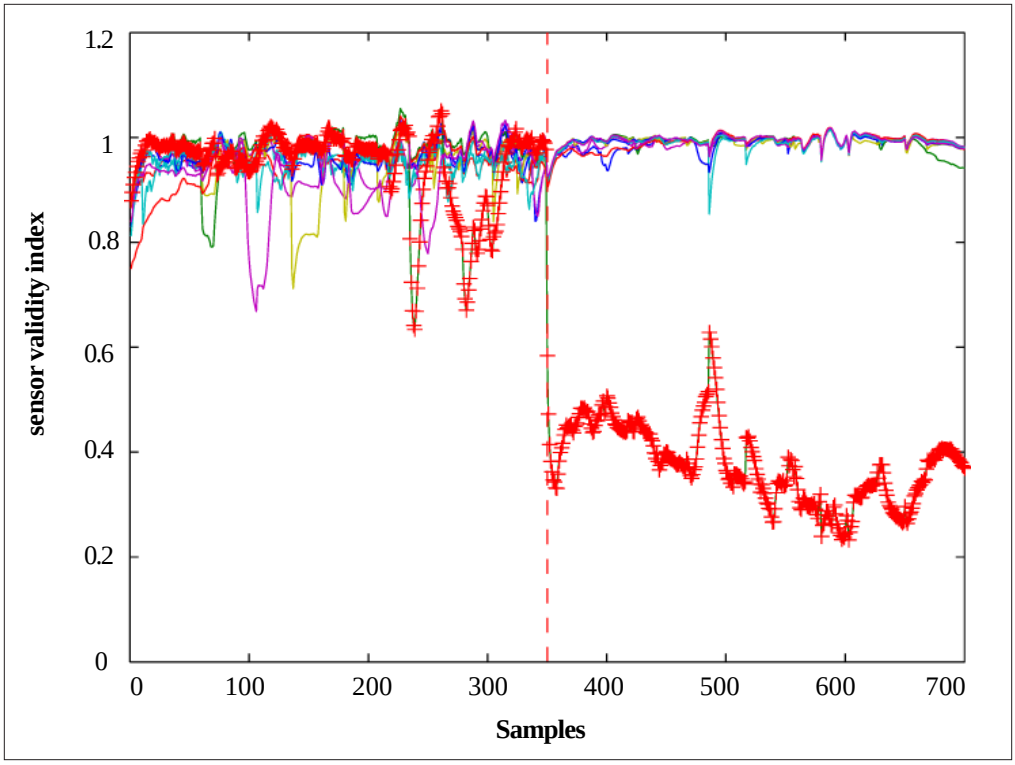


Figure 7. Evolution of filtered SPE and the fault localization affecting the 9*th* sensor measuring the *NTU* at sample 150



(a)



(b)

Figure 8. localization from the evolution of SVI with a fault affecting the 2th sensor measuring the *PH*

firstly; with the procedure of contribution plots, secondly; with the method based on reconstruction principle. To avoid false alarms we use the EWMA to filter the effect of outliers and noise, the figure (Figure 7) presents the evolution of the filtered SPE in the presence of a fault affecting the 9th sensor measuring the turbidity *NTU* and the fault localization affecting the same variable based on the contribution plots, the figures (Figure 8) present the evolution of SVI with a fault affecting at sample 350 the sensor measuring the variable *PH* these figures we can locate the fault from the decreases variable, which is the 2nd *SVI* corresponding the sensor measuring the *PH*.

8. Conclusions

In recent years, the fault detection and diagnosis methods have been widely developed and used to improve the process operation; particularly the fault detection based on Principal Components Analysis "*which does not require prior knowledge about the process mechanism*" had known a big progress and has been widely developed. The PCA is a modeling tool based on the selection of an optimal number of principal components. In this paper we present, the principle of linear PCA to introduce its nonlinear extension, the NLPCA model is obtained using a neural network with five layers in cascade. For monitoring a process, the statistical SPE is used to detect abnormalities, to identify the faulty variables, two diagnosis algorithms are used such as the localization based on the contribution plots and the localization based on reconstruction principle via sensor validity index (SVI). The filter applied to the SVI and SPE adds an important feature for sensor fault localization because reduces the effect of false alarms. The principal idea of this article is to apply the diagnosis of sensors operating state used in WTP, the simulation results obtained in this work show the effectiveness of the proposed approaches. Although the efficacy of this method, it can be improved for the better in the future.

References

[1] Hastie T. (1984). Principal curves and surfaces. PhD thesis, Stanford University.

[2] Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37 (2) 233-243.

[3] Box G. E. P. (1954). 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, 25, p. 290-302, 1954.

[4] Ricardo Dunia et al. (1996). Use of principal components analysis for sensor fault identification, *Computers chem. Engng*, 20, p. 713-718.

[5] Qin S. J. Dunia, R. (2000). Determining the number of principal components for best reconstruction. *Journal of Process Control*, 10 (2) 245250.

[6] George C. Ranger, Frank B. Alt. (1996). Choosing principal components for multivariate statistical process control, *Communication in Statistics -Theory and Methods*. 25 (5) 909-922.

[7] Jackson, J. E. (1991). A users guide to Principal Components, Wiley, New York.

[8] Jolliffe, LT. (1986). Principal Component Analysis, Springer Verlag, Berlin.

[9] Wold, S., Esbensen, K., Geladi, P. (1987). Principal Component Analysis, *Chemometrics and Intelligent Laboratory Systems*, 2, p. 37-52.

[10] E. E. Oja, The Nonlinear PCA Learning Rule and Signal Separation - Mathematical Analysis, Technical Report, Helsinki University of Technology, p. 26, 1995.

[11] Oja, E. (1982). A simplified neuron model as a principal component analyzer, *Journal of mathematical biology*, 15, p. 267-273.

[12] Oja, E., Karhunen, J. (1985). On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix, *Journal of mathematical analysis and applications*, 106, p. 69-84.

[13] Carlos Alcala, Joe Qui, S. (2009). Reconstruction-based contribution for process monitoring, *Automatica*, 45, p. 1593-1600.

[14] Qin, S. J., Weihua, L. (1999). Detection, identification and reconstruction of faulty sensors with maximized sensitivity. *AIChE Journal*, 45 (9) 1963-1976.

[15] William W. Hsieh. (2001). Nonlinear principal component analysis by neural networks. *Tellus A*, 53 (5) 599615.