ISSN : 2231-0711

A New Fault Diagnosis Method Using Fault Directions in Partial Least Square

Ouni Khaled, Dhouibi Hedi, Nabli Lotfi

UR ATSI D. Génie Electrique, ENIM
R. Ibn El Jazzar. 5000 Mounastir, Tunisie

***Abstract*— In this paper, we propose a new approach of fault detection and diagnosis combining a Neural Nonlinear Principal Component Analysis (NNLPCA) and Partial Least Square (PLS). We have made a comparative study between the Linear Principal Component Analysis (LPCA) and Nonlinear Principal Component Analysis (NLPCA) to monitor a manufacturing process. This study has shown the capability of NLPCA in explaining nonlinear correlations in the process data. The traditional LPCA is limited to complex nonlinear systems; therefore, an adaptive NLPCA based on an improved training auto-associative neural network is presented. The proposed approach is applied to fault detection of a manufacturing process. The performance of the proposed approach is then illustrated and compared to those of classic LPCA.**

***Keywords*— Neural Nonlinear Principal Component Analysis, Partial Least Square, Clustering, Pre-analysis, Fault visualization, Fault diagnosis.**

I. Introduction

In manufacturing industries, reliability is a costly event. It remains a product quality indicator of the greatest importance in competitive manufacturing operations [1]. Certainly, a number of different approaches to the failure detection and diagnosis have been evolved in many researches [2].

An important measure of quality for some products is the number of defects per manufactured item. The number of defects per unit area, volume or weight or on a single manufactured item is an important measure of quality for many products. A typical example is the system of transforming and producing cigarettes. The aim of a cigarette manufacturing process is to control and respect the interval constraint on the weight of the manufactured unit. In fact, from a quality point of view, a very heavy cigarette has a difficult pulling and a light cigarette gives consumers the impression to be harmed presenting less garnished ends which empty easily. From a cost point of view, tobacco excess in a cigarette is considered as a loss and can cause stops resulted from stuffing, as in making pudding.

It is known that each stop of the machine systematically generates a certain quantity of rejection (cigarettes during formation are discarded). Usually, a manufacturing system may have several kinds of sensors, but information processing and decision making using the data acquired by all the sensors are very difficult problems for a cigarette manufacturing process. Sensor defaults (biases or drifts) can affect the cigarette manufacturing process comportment until damaging the production. A cigarette manufacturing process is a non linear, non stationary and multi variable complex system. For this kind of industrial process, the precise mathematical model may be difficult to obtain due to the complexity and the high dimensionality of the process. To get this objective, it is possible to consider the modeling approaches based on the data driving technique such as the Principal Component Analysis (PCA), linear or non linear [3]. The PCA is a statistical method which uses linear correlations while reducing the variable dimension. The PCA, also known as empirical orthogonal function analysis, is used to reduce the dimensionality of the measured data. Linear PCA is a classic approach of reducing the dimension of multi-variable data with a minimum loss of information. Using the linear PCA, the approximation of a space with a large dimension to an under-space with a small dimension is realized by the dominant proper vectors of the covariance matrix. Applying the PCA for detecting and locating defects has particularly been important and widely used to monitor industrial processes [4], [5]. Most industrial processes have non linear behaviors.

During the use of the linear PCA, significant non linear information is lost. The non linear PCA is an expansion of the linear PCA. D. Dong and T.J. Mcavoy have shown that using the classic linear PCA in non linear systems is inadequate. If secondary or low-variance components are removed from the PCA model, an important piece of information will be rejected. And if its secondary components are kept, we can use an excess of components to solve this problem. The classical PCA has some limits; for example, it does not pick up the non linear relations between the variables. Also, if the dimension of the input variables is higher than 2, the projection into a linear plan gives a limited decrease in the dimension and visualization of data. The Non Linear Principal Component Analysis (NLPCA) aims at extracting both linear and non linear relations by projecting data on curves and surfaces. The NLPCA method based on artificial neural network has generated an important non linear part during the training procedure [6]. The monitoring approach described in this paper use non- linear Principal Component Analysis based on Neural Network. This method associates NNLPCA and PLS-DA, and it is based on the following assumptions.

We consider NLPCA, based on a multi-layer perceptron (MLP) with an auto-associative topology. The auto-associative neural network performs the identity mapping the output $\hat{x}$ which has to be equal to the input x by minimizing the mean square error (MSE).

- After determining the control limit or threshold, a statistical index is used for fault detection.
- We employ PLS-2 to visualize the number of classes in the data with and without faults and classify them into different classes using a silhouette cluster.

- To find fault direction in PLS that optimally separates each fault region of data from normal data. The weights in fault direction are used to generate contribution plots for fault diagnosis.

Our approach consists of three mean steps:

- Pre-analysis: The auto-associative neural network model has been adapted to perform nonlinear PCA and minimize the cost function $J = \|x - \hat{x}\|^2$.

- Fault detection and class visualization: They allow isolating normal and abnormal data with a silhouette cluster which is carried out by performing PLS-2 for an isolated class of faults.

- Fault diagnosis: Locating faults, therefore, is realized by calculating a contribution plot in a PLS direction.

In this paper, we will begin first by a presentation of some related works. Section 2 gives a process description. Section 3 is dedicated for a brief review of the non linear PCA based neural networks. The monitoring approach is presented in section 4 followed by the exploitation of this approach for diagnosis. In section 5, we discuss the applicability of our approach through a real data simulation of producing a cigarette manufacturing process. Finally, a conclusion is presented with some perspectives.

II. Related Works

The principal component analysis, if initially designed for compressing the number of characteristics in order to reduce the space dimension of data representation, is likewise an interesting tool to detect and locate measurement errors and process dysfunctions [7], [8]. However, the direct application of the PCA to data resulting from a dynamic system does not reveal the exact relations existing between the variables. Ku et al [9], proposed a development, called dynamic principal component analysis, which consists in applying the PCA to a matrix of extended data containing different variables of the staggered system. Recently, Li and Qin (2001) and then Wang and Qin (2002) proposed some extensions to the dynamic PCA close to the identification methods of sub-spaces to eliminate an eventual bias on the model parameters. Whereas the PCA is a linear method, most of physical systems have non linear behaviors. This has motivated a certain number of works to broaden the analysis range in a non linear frame [10]: principal curves, an auto-associative neural network, non linear principal analysis (NLPCA), and a radial-based function network (RBF).

The adaptive PCA proposed by Dayal and McGregor [11] updates the model parameters by applying an exponential sliding window to adjust the model to new conditions. Other extensions have been developed. Each one is based on different process aspects. Nomikos and Mac Gregor [12] have extended the use of multi-variable projection methods into packet processes (Batch processes) using the multiple PCA (multiway PCA). The hierarchical or multi-bloc PCA allows an easier modeling and interpretation of a big matrix by decomposing it into a small matrix or bloc [13].

The association of the conventional PCA with the wavelet transformation aroused a considerable interest these last years [14]–[17]. Recent studies relying on the RNAs have been proposed by R. Shao et al [18]. They use an online-optimization algorithm of neural network data (Input Training Networks), where non parametric thresholds are applied on the control graphs to detect faults. To determine the number of principal components, many rules have been proposed in literature [19]–[22]. Many neural network-based techniques have been developed in the case of non linear PCA such as the sequential and parallel extraction and the input training [23]. After the fault has been detected, it is important to identify the abnormal data.

Therefore, the variable reconstruction approach by Duna et al is in a linear case. Haraket in [24] proposed the reconstruction approach in the nonlinear case by examining the residuals given by the NLPCA model before and after reconstruction. So, in our proposed approach, we have used the PLS-2 Partial Last Square discriminant analysis and the classical algorithm (NIPLAS) to predict the number of classes to be taken in advance by the k-mean clustering algorithm. As soon as the k-mean clustering algorithm is applied, the silhouette technique calculates the mean width of the silhouette of each group as well as the width of the silhouette cluster for the data totality. The mean width of the silhouette can be applied to evaluate the validity of clusters and also be used to decide how the predictive number of determined clusters is suitable for PLS-2.

III. Process Description

One of the manufacturing systems where the respect of a produced item's weight arises is a cigarette making workshop. The process, as shown in Fig. 1, makes a regular and homogenous tobacco pudding (endless cigarette):

A beam of tobacco is enveloped by cigarette paper by means of an adhesive. The resulting pudding is cut up into segments corresponding to one cigarette in order to obtain a rough consumed unit (cigarette without filter). Within this process, a weight interval constraint must be respected. In fact, from a quality point of view, a too heavy cigarette is difficult to draw and a too light one does not satisfy requirements. The production of cigarettes consists of three steps:

- Preparation of a tobacco cut beam that will be setting to obtain a trimmer or modulus τ.
- Forming of a pudding with density υ by enveloping the beam with cigarette paper.
- Cutting up the pudding into ℓ long rough consumed units. Henceforth, such a unit is called simply cigarettes.

The system of transforming and producing cigarettes is described in a textual and functional manner [25]. The problem of respecting the interval constraint on the weight of the manufactured unit is then posed.

- Factor 1: The timer (modulus) τ.
- Factor 2: The density υ.
- Factor 3: The compactness α.
- Factor 4: The pulling resistance γ.
- Factor 5: The dampness rate ρ.
- Factor 6: The weight ω.

The density depends on the modulus value which is adjustable while acting on the modulus pieces of the process. The fluctuation of the modulus depends on the state of the piece shapes in the process. The compactness α depends on several parameters as cultural practices and certain treatments after harvest. The value of ρ depends on α, the weather and the storage conditions. Note that these parameters υ, τ, ω are inter-related. Obviously, the variation of one of these parameters provides a weight variation. When it is outside the validity range, the production has to be rejected or the machine will be blocked. For a normal functioning of the process, the value of each parameter should lie in a given validity interval :

- ω : The weight of the cigarette with $\omega \in [\omega_{\min}, \omega_{\max}]$ expressed in g.
- τ :The modulus of the cigarette with $\tau \in [\tau_{\min}, \tau_{\max}]$ expressed in g/m3
- ρ : The dampness rate of the cigarette with $\rho \in [\rho_{\min}, \rho_{\max}]$.
- α : The compactness of tobacco with $\alpha \in [\alpha_{\min}, \alpha_{\max}]$.
- γ : The pulling resistance $\gamma \in [\gamma_{\min}, \gamma_{\max}]$.
- υ : Density of tobacco $\upsilon \in [\upsilon_{\min}, \upsilon_{\max}]$.

Thus, we look for identifying and locating failing sources of parameter drifts to prevent the negative consequences which will affect all these factors.

IV. Non Linear PCA Based Neural Network

Let us consider $x(k) = [x_1(k), x_2(k), \ldots, x_m(k)]^T$ the vector containing m observed variables in instant k. Then, we can define the data matrix $X = [X_1(1), X_1(1), \ldots, X(N)]^T \in \Re^{N \times m}$ with N samples. So, the PCA concludes the linear combination of the data matrix X in terms of capturing the variation in the data set.

$$T = XP \; and \; X = TP^T \tag{1}$$

with $T = [t_1, t_2, \ldots, t_m] \in \Re^{N \times m}$, where the vector t_i are called the first principal components (PC); and with the matrix $P = [p_1, p_2, \ldots, p_m] \in \Re^{N \times m}$, where the orthogonal vectors p_i are called eigenvectors of the data covariance matrix Σ.

Whereas much has been found by the use of the PCA and other linked methods, being linear methods involves an excessive simplification of the potential of the analysed data sets. The appearance of neural networks (NN) models, a class of powerful non linear empirical methods stemming from the artificial intelligence field, increases the hope that the linear restriction in our analysis of the data sets can finally be solved. Various NN methods have been developed to carry out the PCA. The non linear principal component analysis (NLPCA) using the NN was first presented by Kramer (1991) and is now being used by researchers in many fields. Many NLPCA methods have been proposed. We distinguish here two methods. The first one is based on principal curves proposed by Hastie [26]. However, we cannot use this approach directly for diagnosis. The second NNLPCA method is based on five-layer neural networks proposed by Kramer [27] and which has recently been used for detecting faults. We have chosen to use the last method. The NNLPCA is applied by using a five-layer neural network composed of parallel series layers.

The output of a layer q is the input of the layer $q+1$. The network is composed of five layers [28]: an input layer and an output one with neurons. The first layer hidden for coding and the third one for decoding are based on a non linear transfer function (sigmoid). The second layer hidden inside a network is called a bottleneck layer. In the first hidden layer, the transfer function is the sigmoid function defined as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

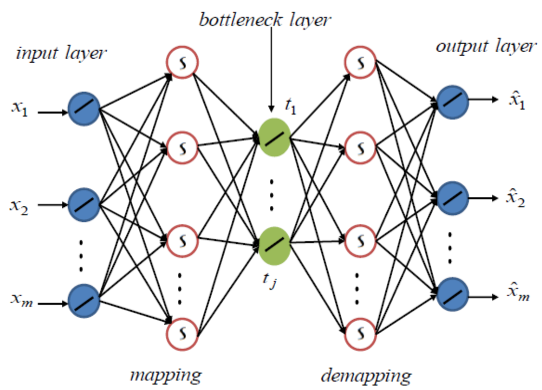

Fig. 1 Optimal sigmoid five-layer neural network for extracting the non linear component t_j.

The sigmoid neural network contains three hidden layers between the input and output variables. The first hidden layer is based on a non linear transfer function where you use the sigmoid function $\sigma(x)$ [29]. The function realizes a projection of input variables towards the first hidden layer (mapping layer), and is expressed by:

$$h_k^{(x)} = f_1((w^{(x)}x + b^{(x)})_k) \tag{3}$$

A transfer function f_1 maps from x, the input column vector of length m, to the mapping layer, presented by $h^{(x)}$, where $w^{(x)}$ and $b^{(x)}$ represent respectively the weights and biases, which are optimized using a conjugate gradient algorithm. The weights and biases are adjustable parameters.

A transfer function f_2 maps from the mapping layer to the bottleneck layer containing a reduced number of neurons, which represents the nonlinear principal component t.

$$t = f_2(w^{(x)}h^x + b^{(x)}) \tag{4}$$

The transfer function f_1 is a non linear function and f_2 is an identity function.

The transfer function f_3 maps from t to the final hidden layer $h_k^{(u)}$ (demapping layer).

$$h_k^{(u)} = f_3((w^{(u)}u + b^{(u)})_k) \tag{5}$$

The transfer function f_4 maps from $h^{(u)}$ to the output column vector $\hat{x}$.

$$\hat{x} = f_4((w^{(u)}h^{(u)} + b^{(u)})_i) \tag{6}$$

The weights and biases are optimized using the gradient propagation algorithm to minimize the cost function between the network input x and the output $\hat{x}$:

$$J = \|x - \hat{x}\|^2 \tag{7}$$

A desired minimum square error between the neural network output and the original data is thus minimized. The choice of the number of hidden neurons in a mapping and demapping layer follows a general principle of parsimony.

In this paper, we propose our new fault detection method which is based on three steps in Fig. 2: pre-analysis, fault detection and class visualization, and fault diagnosis.

The first step of the approach is to calculate $\hat{X}$ with the NNLPCA. The second step is to start with the comparison between the SPE and a detection threshold, which makes it possible to detect a default and to visualize the number of classes in the data using the NIPALS algorithm of PLS-2. The NIPALS or PLS-class algorithm has been performed by using the PLS algorithm implemented in the PLS-Toolbox [30] to predict a class of fault. The data are then classified into different classes using k-means clustering.

In this work, k-means clustering is used to isolate different classes of data to separate the data containing normal and abnormal classes of data. In the third step, fault direction in the PLS is applied to find a fault direction that optimally separates each fault of data from normal data. The weights in fault directions are used to generate contribution plots for fault diagnosis.

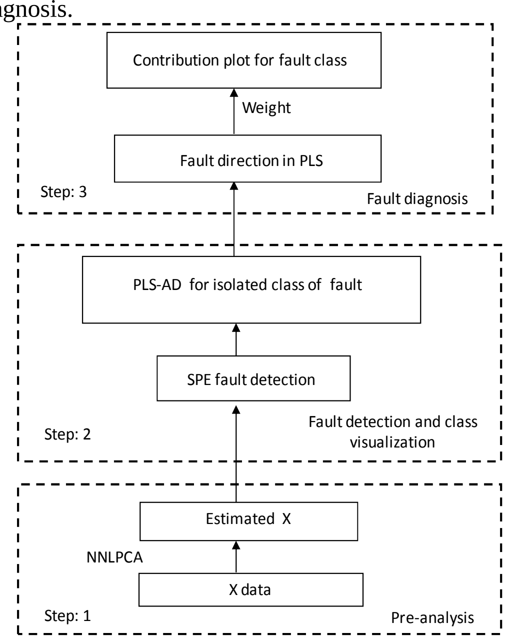

Fig. 2 Fualt visualization, and fault diagnosis method.

V. FAULT VISUALISATION AND DIAGNOSIS APPROACH

A. Pre-analysis

The NNLPCA is used to estimate the output of the neural network and to calculate the non linear principal components. The choice of the number of neurons j in the bottleneck layer is determined by estimating an index ε given as follows:

$$\varepsilon = \sqrt{\frac{\left\|\hat{X}-X\right\|^2}{\left\|X-\hat{X}\right\|^2}} \quad (8)$$

where $\bar{X}$ is a matrix whose vectors are composed of average vectors of the matrix X. We have used the NNLPCA to estimate the outputs of the neural networks and to calculate the non linear principal components t_j. The number j will be increased and ε will be surveilled at the same time. The monitoring approach presented in the experiments section uses these parameters.

B. Fault detection and class visualisation

1) Fault detection

It can be applied after training a five-layer neural network with a sigmoid transfer function. Using the NNLPCA, detecting faults can be realized by the quadratic error SPE(k) (square prediction error), also known as statistic Q.

$$SPE(k) = e^T(k)e(k) = \sum_{i=1}^{N}\left(x_i(k)-\hat{x}_i(k)\right)^2 \quad (9)$$

The process is considered wrong at the instant k if:

$$SPE(k) < \delta_\alpha^2 \quad (10)$$

where δ is the trust threshold of the SPE. When

$\theta_i = \sum_{j=\ell+1}^{m} \lambda_j^i, i = 1,2,\ldots,m$ and λ_j are the proper value of the covariance matrix Σ.

To improve the detection quality and to reduce the rate of false alarms, the Exponentially Weighted Moving Average (EWMA) filter is applied to the residues. The general expression of this filter applied to the residues is given by:

$$\bar{e}(k) = (I-\beta)\bar{e}(k-1) + \beta e(k) \quad (11)$$

where β is a diagonal matrix whose elements are the fault factors for the residues, I is an identity matrix and $\bar{e}(0) = 0$.

$$\overline{SPE}(k) = \left\|\bar{e}(k)\right\|^2 \quad (12)$$

2) Class visualisation

PLS-DA classification: PLS discriminant analysis (PLS-DA) is a linear regression method. PLS-DA is a classification technique that encompasses the properties of partial least square and the discriminant analysis. The PLS-DA applies partial least square (PLS) regression to designate the class of the sample.

When PLS is used in discrimination, the matrix of responses Y contains the information about class. PLS-DA is used to build a predictive model between the input data matrix X and the matrix Y of response variables [31]. In the case of two groups (with fault and without fault), a matrix Y can be created containing the value (0) which indicates the class without fault and the value (1) which designates the class with fault. In the literature we distinguish two proprieties of partial least square regression PLS1 and PLS2 with a single variable response and with several variable responses respectively. When the matrix Y is selected for the PLS-DA, p is the number of classes at fault.

For each class at fault there are $n_1, n_2, \ldots n_p$ observations respectively for each variable in the classes. The p classes are stocked in the data matrix $X \in \Re^{m\times N}$ where there are the two methods, PLS1 and PLS2, to predict the Y model. The PLS-DA is defined as the PLS-2 regression of X . The prediction of the block PLS-2 is defined as [32]:

$$Y = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

where each column of Y corresponds to a class. We can then consider the PLS-DA an approach modeling a set of binary variables from explanatory variables. Each Y element can take 1 or 0. The first element n_1 of column 1 of Y is attributed to 1, which indicates that the line 1 of the matrix is at fault.

k-means clustering: Industrial data usually contain both normal and abnormal data in high-dimensional space, making it difficult to separate manually. For instance, the k-means clustering method is detailed in [33] : K-means clustering is a partitioning method whose objective is to partition data into mutually exclusive k clusters and to return the index of the cluster to each designated observation. Among the most well known clustering methods is k-means, as introduced by Mac Queen [34], which is known for its efficacy in clustering data sets. Unlike hierarchical clustering, k-means clustering controls actual observations and creates a single level of clusters. The distinctions mean that k-means clustering is often more desirable than hierarchical clustering for total data. Each cluster in the partition is defined by its member objects and by its centroid. The centroid for each cluster is the point to which the sum of distances from all objects in that cluster is minimized. K-means computes cluster centroids differently for each distance measure to minimize the sum with respect to the measure that you specify. K-means uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters. The k-means algorithm tries to minimize iteratively the following criteria, by finding an appropriate set of centroids [35], [36].

The result is a set of clusters that are as compact and well-separated as possible. The main steps in k-means algorithm are as follows [37]:

- Select an initial partition and define the centers.
- Assign each entity (station) to the cluster that has the closest centre.
- When all points have been assigned to one cluster, reorder the positions of the centers.
- Repeat steps 2 and 3 until cluster membership does not change.

Finally, this algorithm minimizes the objective function; in this case, a squared error function can be expressed as: Euclidean distance:

$$d_{ij} = \left[\sum_{k=1}^{N} (X_{ij} - X_{ik})^N \right]^{\frac{1}{2}} \qquad (13)$$

To get an idea of how well-separated the resulting clusters are, you can make a silhouette plot using the cluster index output from k-means. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters. This measure ranges from +1, indicating points that are very distant from neighboring clusters, through 0, indicating points that are not distinctly in one cluster or another, to -1, indicating points that are probably assigned to the wrong cluster [38].

C. Fault diagnosis using fault direction in PLS

Fault direction in the PLS is applied to normal data and to each class of fault data to find a fault direction that optimally separates each fault of data from normal data. The weights in fault directions are used to generate contribution plots for fault diagnosis. The PLS-DA is a PLS-based model, where the general model form can be written as in equation (14):

$$Xb = y \qquad (14)$$

where y is a column vector of observations of a dependent variable and $X \in \Re^{N \times m}$ is a matrix that results from N observations of the m variables. The column vector b contains the m regression coefficients. For solving this equation, we have proposed [39]:

$$X^T Xb = X^T y \qquad (15)$$

y is an n-dimensional vector whose components take on the two considered classes c_1 and c_2, which are respectively the class of normal data and each class of fault data. Then to satisfy the normalization of the first step of the PLS algorithm, one must rescale y to have a zero mean $n_1 c_1 + n_2 c_2 = 0$, where n_i is the number of occurrence of c_i. Arbitrarily, we choose $c_1 = 1$, which implies $c_2 = -\frac{n_1}{n_2}$.

In this case, the right hand side of the equation (15) is reduced to:

$$X^T y = \sum_{n=1}^{n_1} X^{kj} - \frac{n_1}{n_2} \sum_{j=n_1+1}^{n_2} X^{kj} \qquad (16)$$

where X^{kj} is an element of X and $1 \le k \le m$. This can also be written as;

$$X^T y = n_1 (\mu_1 - \mu_2) \qquad (17)$$

where μ_i is the m-dimensional mean whose corresponding variables have a c_i class. Thus, $X^T y$ is in the same direction as the line connecting the means of the two classes of y . The final solution vector b obtained by the PLS is then given by:

$$X^T Xb = n_1 (\mu_1 - \mu_2) \qquad (18)$$

$$b \,\square\, \Sigma_t^{-1} (\mu_1 - \mu_2) \qquad (19)$$

where the covariance matrix is $\Sigma = \frac{1}{N} X^T X$.

So, we have a PLS direction, where $\phi_1 = (\mu_1 - \mu_2)$, and this is the first PLS direction. Therefore, we define this ϕ_i direction as the fault direction for class c_i . The weight in ϕ_i is used to generate the contribution plot for the class in fault.

$$\phi_i = \left[\phi_1, \phi_2, ..., \phi_j, ..., \phi_m \right]^T \qquad (20)$$

For a fault direction, the j^{th} element ϕ_j is the contribution from the j^{th} variable. The monitoring algorithm is described in Fig. 3. In the proposed algorithm, the NNLPCA is used to minimize the MSE, and the evolution of SPE has been employed for fault detection. If the first test is true, we must execute the second filtered test to distinguish between a fault and a false alarm.

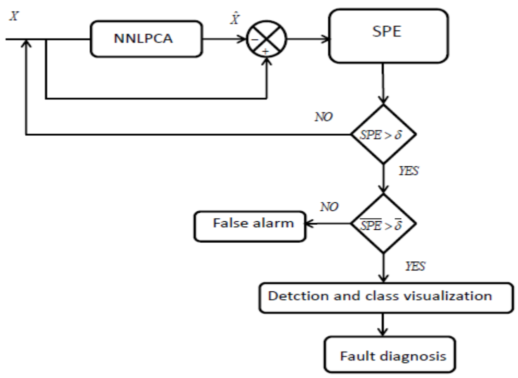

Fig. 1 Algorithm of monitoring method.

VI. Experiments

The diagnosis method, exposed previously, has been validated in simulation on a real system: a cigarette manufacturing process. Results presented below are organized in three parts. In the first section, we determine the number of non linear principal components. In the following section, these latter components are to be used in the bottleneck layer for neural network training. In the third section, we present the results obtained for the fault detection, visualization and diagnosis.

D. Determining the number of non linear principal components

For the cigarette manufacturing process, we have at our disposal a data-base corresponding to measurements carried out during three months. The measurements of the process variables are collected in a matrix $X \in \Re^{N \times m}$. If m is the number of variables, $\rho, \upsilon, \gamma, \tau, \omega$ and α are respectively the dampness rate of tobacco, density, the pulling resistance, the modulus of the cigarette, the weight and the compactness. If N is the number of observations for each variable, all data is centred and reduced and the new data matrix is standardized.

TABLE I
INDEX ε IN FUNCTION OF NEURON NUMBER OF THE BOTTLENECK LAYER

$\varepsilon(1)$	$\varepsilon(2)$	$\varepsilon(3)$	$\varepsilon(4)$	$\varepsilon(5)$	$\varepsilon(6)$
1.116	1.084	1.136	1.931	2.954	2.593

The number of non linear components is $j = 2$ and there are two neurons in the bottleneck layer.

E. Neural network training

The training set (400×6) has been performed and used to determine how many nonlinear component have been kept in the final model. For the network training, we have used the gradient descent backpropagation algorithm. After optimizing the number of neurons as 10 neurons in the first and third hidden layers, we have had good results of the training of the five-layer neural network.

The performance has been the mean square error (MSE), that is the average squared error between the input and output data. The performance goal has been 1.23 at epoch 5,587. We do the training of a five-layer neural network until capturing a variance of the input X, which is the estimated output $\hat{X}$. After that, the training has been achieved after 5,587 training epochs as shown in Fig. 4.

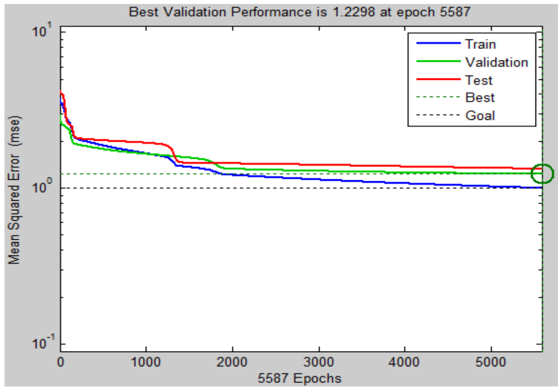

Fig. 4 Mean squared error as a function of the number of epochs for the training set when applying ANN.

Fig. 5 and Fig. 6 present measurement and estimation of compactness and modulus, respectively. The estimation is given by an NNLPCA model.

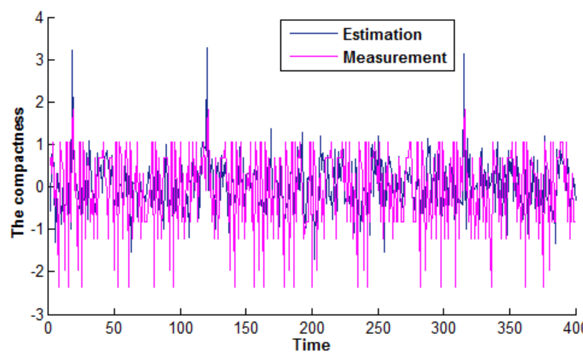

Fig. 5 Measurements and estimation of compactness α.

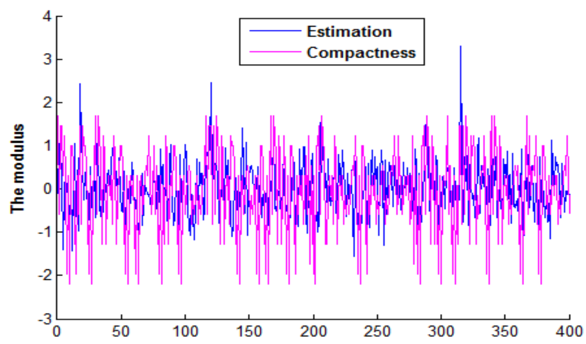

Fig. 6 Measurements and estimation of compactness τ.

VII. Simulation Example

In this application, a fault bias is simulated in the parameters of the modulus τ during a fixed time interval (fault bias 5%). The SPE chart is applied to detect faults. The contribution plot based on both PLS fault direction and scores in LPCA is used to diagnose the faults and compare their performances. The evolution of the SPE allows showing the existence of three operation regions presented in Fig.7. We see clearly that there are three operation regions where A and C are the normal regions, and B is the fault region.

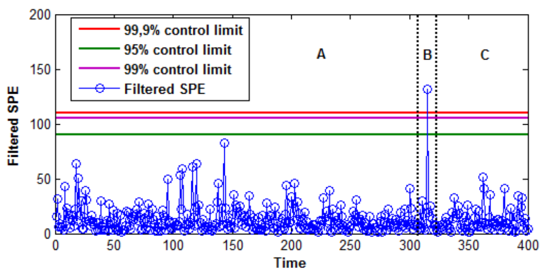

Fig. 7 The evolution of filtered square prediction error.

As for the evolution of the SPE chart, we make out five regions; and to predict the number of classes, we are going to use the partial least square discriminant analysis by applying the PLS-2 algorithm. Fig. 8 presents the contribution of classes A, B and C which correspond to the regions A, B and C.

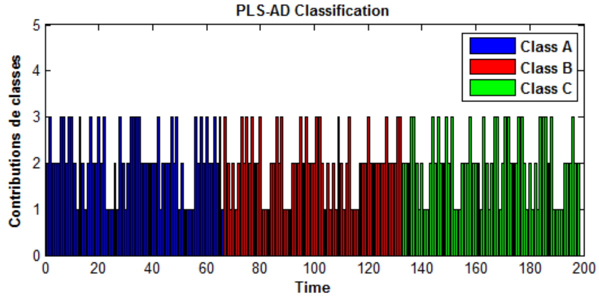

Fig. 8 Class prediction by PLS-DA.

After removing ambiguous points from the transitional regions, we perform the PLS direction to get an overall view of the three classes A, B and C. Fig. 9 shows clusters in PLS directions, where the classes A, B and C correspond to the operation regions A, B and C. Based on our assessment, we have noticed that group B is the region at fault.

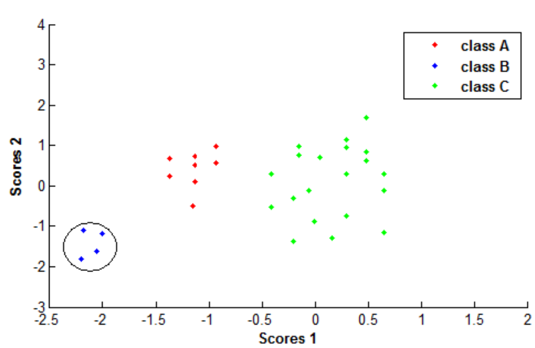

Fig. 9 2-D PLS score plot.

After the process, data are classified into disjoint classes, the contribution plots based on the PLS fault direction for fault B is given in Fig. 10. To make a comparison, the contribution plot based on the LPCA model is also given in Fig. 11. The contribution plot based on the PLS in class B indicates that variable τ which has a higher contribution is a variable at fault. Whereas, in the contribution plot based on the LPCA there are two variables having higher contributions, but variable ω is not a cause of faults.

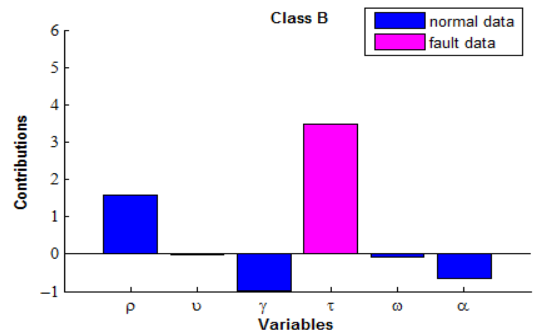

Fig. 10 Contribution plot based on PLS direction.

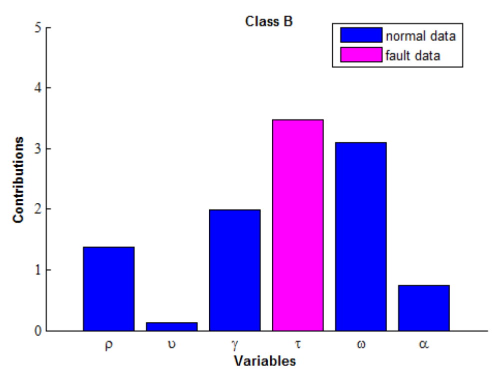

Fig. 11 Contribution plot based on LPCA.

VIII. Conclusions

The linear PCA is not adapted to the study of the non linear system. So, we have proposed an approach based on the NNLPCA. The NLPCA is associated to the neural network. Our contribution is to extend this method to data processing which presents non linear behavior.

The SPE chart is used to isolate data in different regions that correspond to fault and normal regions. A clear classes visualisation is obtained by applying the discriminant analysis PLS-2. A contribution plot based on the fault directions in PLS is applied to real production data of an existing workshop. The proposed methodology is therefore validated by a large set of data, and it provides an interesting industrial efficiency for the considered case study. However, we are going to improve the quality of the detection by improving the modeling quality. The results obtained make it possible to validate the method which is applied for detecting and localizing the considered sensor faults. The industrial application of our approach has shown the limit of the LPCA in exploiting nonlinear correlations of data measurements.

REFERENCES

[1] J. G. Besseris, “A methodology for product reliability enhancement via saturated–unreplicated fractional factorial designs,” *Reliability Engineering and System Safety.*, vol. 95(7), pp.742–749, 2010.

[2] Y. Papadopoulos, “Model-based system monitoring and diagnosis of failures using state charts and fault trees,” *Reliability Engineering and System Safety.*, vol. 81(3), pp.325–341, 2003.

[3] Q. Peter, SJ. Qin, J. Wang, “A new fault diagnosis using fault direction in fisher discriminant analysis,” *American Institute of Chemical Engineers Journal* ., vol. 51(2), pp.555-571, 2005.

[4] L. Erickson, J. Ohansson, N, Kettaneh and S. Wold, *Multi-and Megavariate Data Analysis Part I*: *Principles and Applications*., Umetrics, 1991.

[5] M. F. Haraket, G. Mourot and J.Ragot, “Nonlinear PCA Combining Principal Curves and RBF-Networks for Process Monitoring,” in *Conference on Decision and Control 2003*, *Hawaii USA*, pp.1956–1961.

[6] K. Ouni, H. Dhouibi, L. Nabli, M. Hassani and Z. Simeu-Abazi, “Fault Detction and Localization with Neural Principal Component Analysis,” in *18th Mediterranean Conference on Control and Automation (MED)*, *Marrakech, Morocco*, pp. 880 – 885, june 2010.

[7] D. W. Shifei, X. Jia Xu, and L. Zhang, “PCA-Based Elman Neural Network Algorithm,” *Springer-Verlag Berlin Heidelberg.*, vol. 5370 pp.315-321, 2008.

[8] R. Dunia, S. J. Qin, T. F. Edgar and T. J. McAvoy, “Identification of faulty sensors using principal component analysis,” *AIChE Journal.*, vol. 42(10), pp. 2797–2812, 1996.

[9] J. Gertler, “Fault detection and isolation using parity relations,” *Control Engineering Practice.*, vol. 5(5), pp. 653–661, 1997.

[10] W. F. Ku, R.H Storer and C. Georgakis, “Disturbance detction and isolation by dynamic principal component analysis,” *Chemometrics and Intelligent Laboratory System*., vol. 30(1), pp. 179–196, 1995.

[11] A. R Webb, “An approach to non-linear principal component analysis using radially symmetric kernel functions,” *Statics and computing*., vol. 6(2), pp. 159–168, 1996.

[12] B. S. Dayal and J. F. MacGregor, “Recursive exponentially weighted pls and its applications to adaptive control and prediction,” *Journal of Process Control*., vol. 7(3), 169–179, 1997.

[13] P. Nomikos and J. F. MacGregor, “Monitoring batch processes using multiway principal component analysis,” *AIChE Journal*., vol. 40(8), pp. 1361–1375, 1994.

[14] S. Wold, N. Kettaneh and K. Tjessem, “Hierarchical multiblock pls and pc model for easier model interpretation and as an alternative to variable selection,” *Journal Chemometrics*., vol. 10(5), pp. 463–482, 1996.

[15] B. R Bakshi, “Multiscale pca with application to multivariate statistical process monitoring,” *AICHE Journal*., vol, 44(7), pp. 1596–1610, 1998.

[16] K. A Kosanovich and M. J Piovoso, “Pca for wavelet transformed process data for monitoring,” *Intelligent Journal of Data Analysis*., vol. 1(1-4), pp. 85–99, 1997.

[17] M. Misra, H. H. Yue, S. J. Qin and C. Ling, “Multivariate process monitoring and fault diagnosis by multiscale pca,” *Computer Chemical Engineering*., vol. 26(9), pp.1281–1293, 2002.

[18] Rosen C and J. A Lennox, “Multivariate and multiscale monitoring of wastewater treatment operation,” *Water Science and Technology*., vol. 35(14), pp. 3402–3410, 2001.

[19] R. Shao, F. Jia, E. B. Martin and A. J. Morris, “Wavelets and non-linear principal components analysis for process monitoring,” *Control Engineering Practice*., vol. 7(7), pp. 865–879, 1999.

[20] P. Besse and L. Ferre, “Sur l’usage de la validation croisée en analyse en composantes principales,” *Revue de Statistique Appliquée*., vol, 41(1), pp. 71–76, 1993.

[21] B. T. Himes, M. E Goldberger and A. Tessler, “Grafts of fetal central nervous system tissue rescue axotomized Clarke’s nucleus neurons in adult and neonatal operates,” *The Journal of Comparative Neurology*., vol. 339(1), pp. 117–131, 1994.

[22] L. Ferré, “Selection of components in principal component analysis: A comparison of methods,” *Computational Statistics and Data Analysis*., vol, 19(6), pp.669–682, 1995.

[23] S. L Valle, L. Weihua and 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*., vol. 38(11), pp. 4389–4401, 1999.

[24] M. Kirby, R. Miranda, “Nonlinear reduction of high-dimensional dynamical systems via neural networks,” *Physical Reviewer Letters*., vol. 72(12), pp. 1822–1825, 1994.

[25] M. F Haraket, Y. Tharrault, G. Mourot and J. Ragot, “Multiple sensor fault detection and isolation of air quality monitoring network using RBF-NLPCA model,” *International Journal of Adaptive and Innovative Systems*., vol. 1(3-4), pp. 267–284, 2010.

[26] H. Dhouibi, D. S Collart, L. Nabli and E. Craye, “Using Interval Constrained Petri Nets for Regulation of Quality,” in *12th symposuim large scale systems (LSS): theory and application*, 2010, Villeneuve d'Ascq, France.

[27] J. Zhang, E. B. Martin and A. J. Morris, “Process Monitoring Using Non-Linear Statistical Techniques,” *Chemical engineering journal*., vol. 67(3), pp. 181–189, 1997.

[28] A. K Mark, “Nonlinear Principal Component Analysis Using Autoassociative Neural Networks,” *Aiche journal*., vol. 37(2), pp. 233–243, 1991.

[29] C. Zang and M. Imregun, “Structural Damage Detection using Artificial Neural Network and Mesured FRF Data Rduced Via Principal Component Projection,” *Journal of Sound and Vibration.*, vol. 242(5), pp. 813–827, 2001.

[30] Z. Haslinda and T. D Thao, “A Principal Component Approach in Diagnosing poor Control loop performance,” in *Proceedings of the World Congress on Engineering and Computer Science*, *WCECS* , San Francisco, October 24-26, 2007, pp. 194–199.

[31] L. H. Chiang, E.L. Russell and R.D Braatz, *Fault Detection and Diagnosis in industrial Systems*, 2nd ed., Paris: Springer, 2001.

[32] S. Masouma, B. D. Jouan-Rimbaud, J. M. Vercauteren and J. H. Rutledge, “Discrimination of wines based on 2D NMR spectra using learning vector quantization neural networks and partial least squares discriminant analysis,” *Analytica Chimica Acta.*, vol. 558(1-2), pp. 144–149, 2003.

[33] Y. Yücel and C. Demir, “Principal component analysis and cluster analysis for the characterisation of marbles by capillary electrophoresis,” *Journal Elsevier.*, vol. 63(2), pp. 451–9, 2004.

[34] J. M. Queen, “Some methods for classification and analysis of multivariate observations,” in *Berkeley symposium on mathematical statistics and probability*, 1967, University of California Press, vol, 1, pp. 281–297.

[35] J. Chiquet, M. Eid and N. Limnios, “Modelling and estimating the reliability of stochastic dynamical systems with Markovian switching,” *Reliability Engineering and System Safety.*, vol. 93(12), pp.1801–1808, 2008.

[36] H. A. Taboada, F. Baheranwala D.W. Coit and N. Wattanapongsakorn, “Practical solutions for multi-objective optimization: An application to system reliability design problems,” *Reliability Engineering and System Safety.*, vol. 92(3), pp. 314–322, 2007.

[37] M. C. Demirel, “Cluster Analysis of Stream flow Data over Turkey,” M.Sc. Thesis, Technical University, Istanbul, 2004.

[38] Z. Zhang, Z. Zhao and G. Bai, “3D Representative Face and Clustering Based Illumination Estimation for Face Recognition and Expression Recognition,” *Springer-Verlag Berlin Heidelberg.*, vol. 5553, pp. 416–422, 2009.

[39] F. Vance, “Partial Least Squares, Conjugate Gradient And The Fisher Discriminant,” *Los Alamos National Laboratory, Los Alamos New Mexico.*, vol. 3958, pp. 01–17, 1996.