ECG signal monitoring using linear PCA

H.Chaouch
Assistant In The Superior Institute Of Applied Maths And Informatics In Kairouan (ISMAIK)

K.Ouni
Assistant In The Superior Institute Of Applied Sciences And Technology In Kairouan (ISSATK)

L.Nabli
Professor at the Superior Institute of Applied Sciences and Technology in Sousse (ISSATS)

ABSTRACT

In this paper, we propose an approach to study the biomedical signal of the electro-cardiographic ECG. Our contribution consists in applying the principal component analysis (PCA) to help diagnose the cardiovascular system. Many researches have recently approached the same theme by integrating many tools of signal processing, but the most interesting method is the PCA.Starting from the ECG signal, we define the characteristic parameters of the electrical activity of the heart. The PCA allows detecting and then localizing the defective parameters of the ECG and facilitating the diagnosis of the existing heart disorders. The results we get at the end of this paper show that this approach is reliable compared with data given by the medical expert.

Keywords: Principal component analysis, ECG, Diagnosis, Detection of defects, Calculation of contributions.

A	proper vector matrix
C	model projection matrix
c	Row vector of matrix C
E	projection of X to $\Re^{m-\ell}$
e	projection of x to $\Re^{m-\ell}$
I	identity matrix
ℓ	number of principal components
M	number of sensors
m	number of variables
n	number of sample measurements
P	proper value matrix
SPE	square prediction error
T^2	Hotelling statistic
T	score matrix
X	measurement matrix, or data matrix
$\hat{X}$	projection of X to $\Re^{\ell}$
x	sample vector
$\hat{x}$	projection of x to $\Re^{\ell}$
δ_α	Threshold of T^2
χ_α	Threshold of SPE
PA	P wave amplitude
RA	R wave amplitude
QS	duration between Q and S waves
QP	duration between Q and P waves
RR	duration between two successive R waves
ECG	electrocardiographic signal

1. INTRODUCTION

The analysis of the electrocardiographic signal (ECG) uses the tools of signal processing. These tools help analyzing and approaching the cardiovascular problems while integrating the physiological side. The interpretation of the electrocardiographic signal, which comes under the practioner field, uses the analysis or even the modeling of the ECG signals while taking into account the myocardium physiology.The prognosis and diagnosis of the cardiovascular system based on the ECG characteristics have been the focus of many recent researches in the field of signal processing and indirect supervision [1, 2, 3, 4, 5, 6, 7, 8 and 9]. Among the tools used in these researches we have the PCA, the fuzzy logic and the neural network. In this paper, we will integrate the PCA to help diagnose the ECG. PCA is a method of reducing a classic linear dimension that consists in projecting the samples on the axes of maximum variance of the data. Recently, the methods of detecting and localizing defects based on PCA have received particular attention and have been broadly used in the supervision of industrial and biologic processes [10, 11, 12 13, 14 and 15]. The PCA is based on three parts. The first step is to determine the number to keep in the PCA model by introducing various techniques [15, 16 and 17]. The choice of the method of determining the principal components has a significant impact on the rest of the PCA steps. In this work, we use the cumulative variance percent to determine the number of principal components. The second step is the detection of defects that uses the PCA to follow the process behavior at a normal state. On comparing the observed behavior with that given by the PCA, the defects are detected. Many methods have been used to detect the defect, and the frequent ones are the square prediction error (SPE) and Hotteling statistic T^2 [18, 19, 20 and 21], which

will be introduced also in this approach. To localize the defected variables, many methods have been introduced [22, 23, 24 and 25]. In this paper, the defects are localized by the method of calculating contributions.

2. THE ELECTROCARDIOGRAM

2.1 ECG waves

Each cycle of depolarization or repolarization of the heart corresponds to the passage of electric current in the atria to the ventricles that contract in order. This means recording that the ECG is still in the same order of different waves: P, Q, R, S and T.

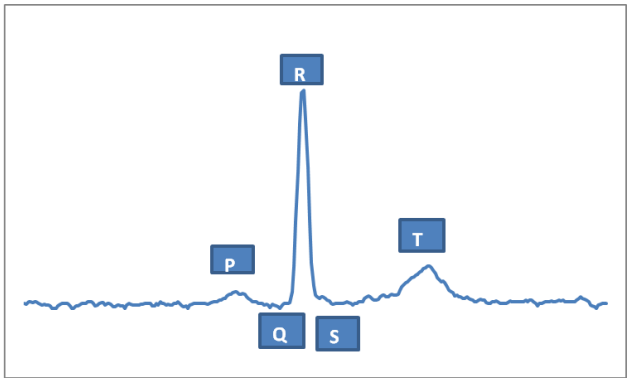

Fig.1. ECG waves

•The P wave: depolarization of the atria from the sinus node to the atrioventricular node resulting in a deflection on the ECG.

•The complex QRS wave: The ventricular depolarization is indicated by the QRS or QRS complex. By definition, the Q wave is the first negative wave; R wave is the first positive wave of the complex and the S wave the first negative wave after wave R.

•The Twave correspondsto ventricular repolarization current. Th is wave follows the QRS complex after returning to the isoelectric line.

2.2 ECG segments

Intervals and segments of ECG parameters are important for assessing the normality or otherwise of the space between two electrical events.

• The PQ interval: is the time between depolarization of the atrium and the ventricle. It's time to wave propagation of depolarization through the atria, the atrioventricular node, the bundle of His and the Purkinje fibers, until the ventricular-myocardial cells.

• The RR interval: between the peaks of two successive R waves and represents the instantaneous frequency.

• QT interval: is the time of ventricular systole from the beginning the excitement of the ventricles until the end of their relaxation.

• The ST segment: is the phase of ventricular depolarization phase during which the ventricular cells are depolarized all: there is no a priori power extension.

We define, therefore, three other parameters related to the monitoring of ECG waves; namely: the P wave amplitude *PA*, the R wave amplitude *RA* and the widths of complex QRS wave, *QS*.

3. MODEL OF THE PCA APPROACH

For an automated approach of detecting and localizing defective parameters on an ECG signal, we have resorted to go through the steps described in the following figure.

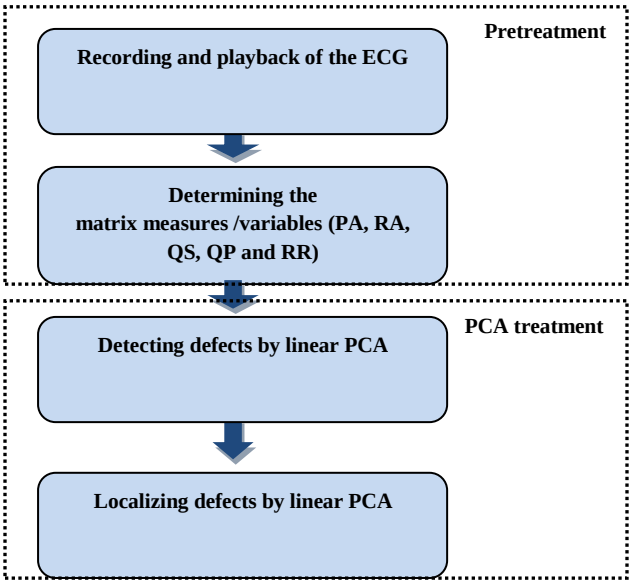

Fig.2. Model approach of ECG signal based on linear PCA

The first step consists in recording and reading the signal stemming from the MIT/BIH data base in order to convert the biological data into computing data. To apply the PCA method, we need to determine a matrix of measurement of variables. Here, the parameters characterizing an ECG signal, taken as variables, are: the wave amplitude P (PA), the wave amplitude R (RA), the width of the complex QRS (QS), the interval separating the wave P and the wave Q (PQ), and the interval between the two successive R waves (RR). The detection step is based on the PCA principle: the signal parts exceeding the chosen threshold are the defective parts which will be localized to determine the defected parameters or variables: this is the localization of defects.

4. LINEAR PRINCIPAL COMPONENT ANALYSIS (LPCA)

The methods for detecting defects and localizing parameters based on PCA have received a particular attention and been broadly used to the supervision of biologic processes. The principle of this approach is using the analysis in principal components to establish a model for the behavior of a normal working process, and thus detecting the defects while comparing

the observed behavior to the given one. In principal components, a characteristic vector $t \in \Re^{\ell}$ is associated to each vector of data so that it can optimize the representation to the sense of minimizing the error of estimation of X or maximizing the variance of the score matrix T . The vector t and the x letter are bound by a linear transformation $t = P^T x$, where the matrix of transformation $P \in \Re^{m \times l}$ meets the condition $P^T P = I$.

Using PCA, an $n \times m$ normalized data matrix can be decomposed as follows:

$$X = \hat{X} + E \tag{1}$$

In this way, the columns of X are normalized to zero mean and unit variance. The matrices $\hat{X}$ and E represent the modeled and unmodeled variations of X respectively;

$$\hat{X} = TP^T \tag{2}$$

$$E = \tilde{T}\tilde{P} \tag{3}$$

X is composed in a way that the composed matrix $[P\tilde{P}]$ is orthonormal, and $[T\tilde{T}]$ is orthogonal. After the PCA model is built, a new sample vector $X \in \Re^m$ can be decomposed in two parts;

$$x = \tilde{x} + e \tag{4}$$

Where $\tilde{x} = PP^T x \equiv Cx \in \Re^{\ell}$ is the projection of the sample vector to the principal component subspace (PCS), $\Re^{\ell}$, and $e = (I - C)x \equiv \tilde{C}x \in \Re^{m-\ell}$ is the projection of the sample vector to the residual subspace (RS), $\Re^{m-\ell}$.

The prediction $\hat{x}_i$ is used as a recovery of x_i, which can be expressed as in (5).

$$\hat{x}_i = xc_i = [x_1 x_2 .... x_i ..... x_m][c_{1i} c_{2i} .... c_{ii} ..... c_{mi}]^T \tag{5}$$

$$= [c_{-i}^T 0 c_{+i}^T]x + c_{ii} x_i$$

Where $C = PP^T = [c_1 c_2 ..... c_m]$ (6)

4.1 Determining the number of principal components

The first step in this approach is to determine the number of principal components. This step is a research procedure of the model structure of the PCA. Many criteria using a lot of approaches have been used in this context

- Cumulative percentage of the total variance (CPV)

In this method, the proper values of the correlation matrix are the measurements of this variance. To choose ℓ, we have to choose the percentage of the total variance that we can keep. The number of principal components is then the smallest number taken in a way that the percentage is attained or overrun; the components are chosen successively in descending order of variances.

$$PCV(\ell) = 100 \times \left(\frac{\sum_{j=1}^{\ell} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \right) \% \tag{7}$$

- Crossed-validation procedure

This method is based on minimizing the PRESS quantity. This quantity is presented as the sum of the errors squares between the observed data and the ones predicted by the model starting from a different identification action.

$$PRESS(\ell) = \frac{1}{Nm} \times \left(\sum_{k=1}^{N} \sum_{i=1}^{m} (\hat{x}_i^{\ell}(k) - x_i(k))^2 \right) \tag{8}$$

- Proper value mean

This method consists in taking into account only the components for which the proper value is higher than the arithmetic mean of all the proper values.

- The Xu and Kailath approach

Xu and Kailath [27] determine the number of components based on the highest proper value of the covariance matrix. This approach implies also that the small proper values of the covariance matrix are equal.

For our approach, the Xu and Kailath method is more reliable in terms of results. Thus, we will integrate it in what follows the PCA procedure.

4.2 Defect Detection

In the approach to PCA-based detection, two statistics T^2 and SPE are used for surveillance.

For the instant k we have:

$$T^2(k) = \sum_{i=1}^{\ell} \frac{t_i^2(k)}{\lambda_i} \quad (9)$$

$$SPE(k) = \sum_{j=1}^{m} (e_j(k))^2 \quad (10)$$

where $e_j(k)$ is the j[th] residue which is given by:

$$e_j(k) = x_j(k) - \hat{x}_j(k) \quad (11)$$

Where $x_j(k)$ is the j[th] element of measuring vector, $x(k)$ is the i[th] principal component and t_i is the i[th] proper value of the correlation matrix $\sum$, which represents the variance of t_i. m is the number of quality indictors and ℓ is the number of components, with $\ell <$ m which is the estimation of x by the PCA model given by:

$$\hat{x} = Cx \quad (12)$$

Where $C = P_\ell P_\ell^T$. P_ℓ is the matrix formed by the ℓ first proper vectors of the matrix $\sum$. The process will be considered functioning abnormally (presence of a default) if one of these following inequalities at least is true:

$$SPE \prec \delta_\alpha^2 \quad (13)$$

$$T^2 \prec \chi_\alpha^2(\ell) \quad (14)$$

Where δ_α^2 and $\chi_\alpha^2(\ell)$ are respectively the thresholds of T^2 and SPE.

4.3 Localization of defects

The calculation of contributions is an approach used for localizing defects. The variable having the highest contribution is considered at fault. In the case of SPE, the contribution $cont_j^{SPE}(k)$ of the j^{th} variable at the instant k is defined by the following equation:

$$cont_j^{SPE}(k) = (x_j(k) - \hat{x}_j(k))^2 \quad (15)$$

The contribution of a variable by the statistic T² is defined by Qin [26].

$$T_i^2 = \left\| A^{-1/2} P(:,i) x_i \right\|^2 \quad (16)$$

5. APPLICATION ON ECG SIGNAL

5.1 Data base

The data base used is the MIT-BIH Arythmie base which contains 48 thirthy-minute extracts of two recording channels of mobile ECG, obtained from 47 subjects studied by the BIH Aythmie Laboratory between 1975 an 1979. A 24-hour recording was taken at Beth Israil Hospital of Boston. This recording is digitalized at 360 samples per second. The data matrix (500x5) is taken from an ECG signal of the MIT-BIH data base. The variables are the characteristic parameters of this ECG: amplitudes of P and R waves, the width of the QRS complex and the length QR and the heart frequency RR. The progress of these parameters is presented in the following figure.

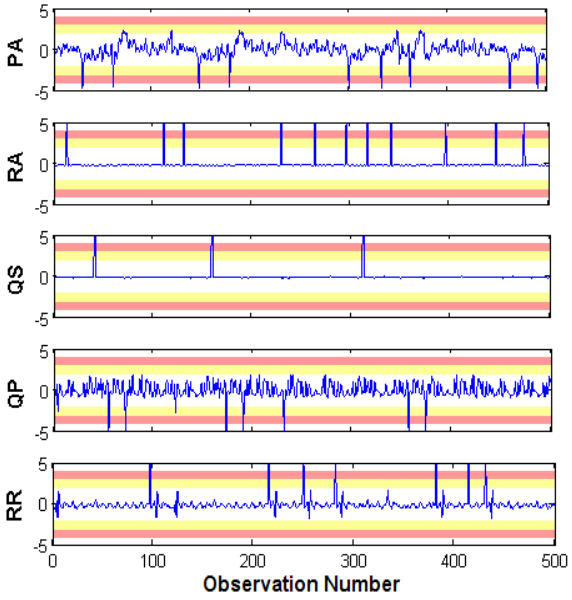

Fig.3 Progress of ECG parameters

5.2 Data Pre-analyse

Beforehand, to make the results independent from the used units for each variable, an essential pretreatment should consist in centering and reducing the variables. The following figure presents the variables of reduced and centered data.

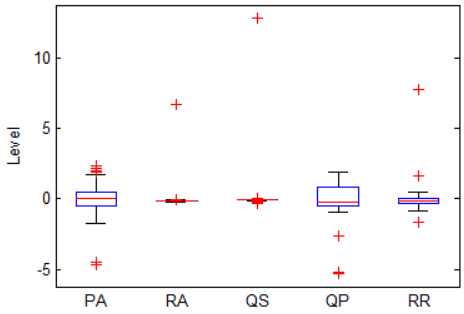

Fig.4 Variables of reduced and centered data

5.3 Determining the number of principal components:

The results of this step of PCA procedure are presented in the following table:

PA	***RA***	***QRS***	***QP***	***RR***
ℓ =1	ℓ =2	ℓ =3	ℓ =4	ℓ =5
1,36	1.07	0.98	0.85	0.71

Table 1: Variances of different indicators

The following figure gives the speed of the progress of the principale components for each varaible.

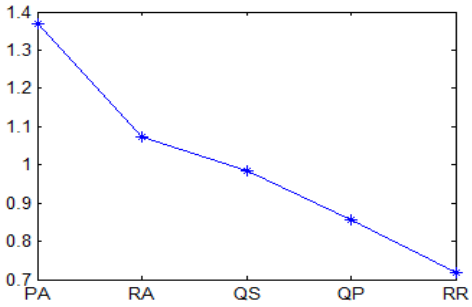

Fig.5 Progress of component selection criterion.

5.4 Detecting defects using SPE and T²

Fig.6 presents the defect detection using the Hoteling T² method; these defects are the signal parts exceeding the thresholds.

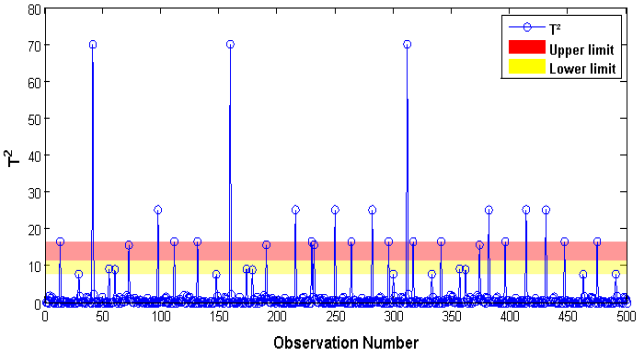

Fig.6 Detecting defects by T²

Observing Fig.7 which presents the defect detection by the SPE method we notice that there is not any difference at the level of results compared with the T² method.

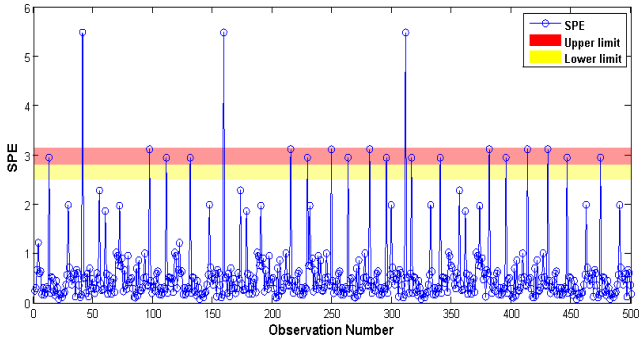

Fig.7 Detecting defects by SPE

5.5 Localizing defected variables

After detecting defects by the two methods T² and SPE, a localization procedure is applied to determine the defected variable or variables using the same techniques (Fig.8 and Fig.9).

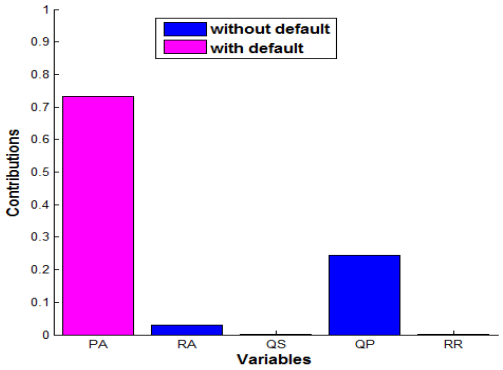

Fig.8 Localizing defects par T²

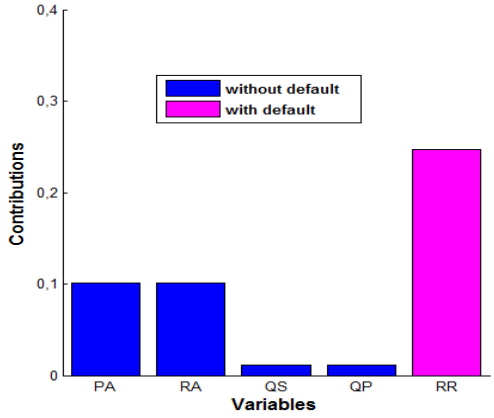

Fig.9 Localizing defects par SPE

According to Fig.8, the PA indicator has the strongest contribution in relation to the other variables; that is why, it is considered the variable in defect. According to Fig.9, the RR indicator has the strongest contribution in relation to other variables and is considered the defected variable. Comparing

localization results of T^2 and SPE methods, we notice that the defected variable is not the same, which requires an intervention by an expert in the ECG field to hold the closest result to real data.

6. CONCLUSION

In this paper we have presented an approach to help diagnose the cardiovascular system starting from an ECG signal. A linear PCA is introduced to detect and then localize the faulty parameters of the ECG. The variables of this analysis are the characteristic parameters of the ECG, the amplitudes of the P and R waves, the width of the QRS complex, the RR heart frequency and the QP interval. The PCA puts into evidence the defected parameters of the ECG signal, using the SPE and Hotelling statistics. After applying the localization procedure, we have noticed a difference at the level of defected variables. Our expert (cardiologist) intervention has shown that the result of the SPE approach is more reliable. Accordingly, the RR variable is in defect due to the irregular heart frequency of the introduced ECG signal. To sum up, the linear PCA application to diagnose the ECG has given reliable results. Whereas, this approach is not enough to cover all defects on an ECG signal because it is not always linear. An integration of the non linear PCA is considered to improve the results of detection and localization and make them closer to the expert data.

7. REFERENCES

[1] S.D.Pamar, Bhuvan Unhelkar, "On the performance analysis of the ICA Alghorithms with maternal and fetal ECG inputs and contaminated noises", IJCA, vol 2, n°7, June 2010.

[2] M. M Sheikh Algunaidi, M. A. Mohd Ali and Md. Fokhrul Islam. "Evaluation of an improved algorithm for fetal QRS detection", International Journal of the Physical Sciences Vol. 6(2), pp. 213-220, 18 January, 2011.

[3] T.M.Nazmy, H.El-Messiry, B.Al-Bokhity, « Classification of cardiac Arrhythmia based on hybrid system », IJCA,vol 2-N°4, june 2010.

[4] H.Chaouch, « Détection du complexe QRS par analyse multi-échelle », mémoire de mastère, INSAT Tunisie,2009.

[5] R.D.Mali, M.S.Khadtare, Dr.U.L Bombale, « Removal of 50 Hz PLI using discrete wavelet Transform for quality diagnosis of biomedical ECG signal, IJCA, vol 23-n°7, June 2011.

[6] J.P.Coudrc, "Analyse Quantitative des composantes temps-échelle de l'ECG à haute résolution moyenne pour l'évaluation du risque du tachycardies ventriculaires et de la mort subite après un infarctus du myocarde", Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 2007.

[7] S .Graja, "segmentation et classification de l'onde P d'un électrocardiogramme :détection du risque de fibrillation auriculaire », Thèse de doctorat, L'école nationale supérieure des télécommunications de Bretagne, 2008.

[8] A. Cabasson, "Analyse des périodes P-P et P-R dans les électrocardiogrammes". Mémoire de Mastère, Université Nice Sophia Antipolis, Septembre 2005.

[9] L. Clavier, J.M. Boucher, R. Lepage, J.J Blanc and J.C Cornily. Automatic P-wave analysis of patients prone to atrial fibrillation. Med bioeng Comp, vol.40 n°1, Jan, 2002.

[10] L. Nabli and K. Ouni., "The indirect supervision of a system of production by the principal components analysis and the average dynamics of the metrics", International Review of Automatic Control (IREACO), pp. 560-567, November 2008.

[11] L. Nabli, K. Ouni, and H. Messaoud., " Méthode de surveillance indirecte d'un système de production par l'analyse en composantes principales", La 5ème conférence internationale JTEA 2008, Journées Tunisiennes de l'Electrotechnique et de l'Automatique, 2-4 Mai.

[12] L. Nabli, K. Ouni, and H.Haj. Salem., " Approche Multi agents pour la surveillance indirecte d'un système de production par l'analyse en composantes principales", la Conférence Internationale Francophone d'Automatique CIFA 2008, pp. 128-136, 3-5 Septembre.

[13] D.Pata, M.Kumardas, S.Pradhan, " Integration of FCM, PCA and Neural Networks for classification of ECG arrhythmias",IJCS 36-3-05, 2010.

[14] M. F. Harakat, G. Mourot, and J. Ragot., " Détection de défauts de capteurs d'un réseau de surveillance de la qualité de l'air", Journal Européen des Systèmes Automatisés. JESA, pp. 417-436, 2005.

[15] Besse P et Ferré L, « Sur l'usage de la validation croisée en analyse en composantes principales. Reveue de statistique appliquée, XLI (1), pp.71-76, 1993.

[16] Ferré L, « Selection of component analysis : A comparition methods", Computional statics and Data Analysis,pp.669-682, 1995.

[17] Valles S, Weihua L, and Qin S.J, "Selection of number of principal components: The variance of reconstruction error criterion with a comparision to other methods", Industrial &Engineering Chemisty Research,vol 38, PP.4389-4401, 1999.

[18] T. Kourti, "Recent Developments in Multivariate SPC Methods for Monitoring and Diagnosing Process and Product Performance", Journal of Quality Technology, vol.28, N° 4, pp. 409-428, 1996.

[19] M.Gregor, Kourti T and Nomikos P, "Analysis, monitoring and fault diagnosis of industrial processes using multivariate statistical projection methods, 13th Triennial Word Congress, pp.145-150, San Francisco, 1996.

[20] Gertler J, "Fault detection and diagnosis systems", Marcel Dekker edition, New York, 1998.

[21] Dunia R and Qin S, "A subspace approach to multidimentional identification and reconstruction", AICHE journal, vol.44,pp1813-1831, 1998.

[22] Gertler J. and McAvoy T, "Principal component analysis and parity relations - Astrong duality". IFAC Conference SAFEPROCESS, Hull, UK, pp. 837-842, 1997.

[23] Gertler J., Weihua L., Yunbing H. and McAvoy T. (1998). Isolation enhanced principal component analysis. 3rd IFAC Workshop on On-line Fault Détection and Supervision in then Chemical Process Industries, Lyon, June 4-5, France.

[24] Dunia R.,Qin S.J. and Edgar T.F,"Identification of faulty sensors using principal component analysis", AIChE Journal, vol. 42, N±. 10, pp. 2797-2812, 1996.

[25] Kourti T., and MacGregor J. F, "Recent Developments in Multivariate SPC Methods for Monitoring and Diagnosing Process and Product Performance", Journal Of Quality Technology, vol. 28, N±. 4, pp. 409-428, 1996.

[26] S. J. Qin, S. Valle, et M. Piovoso. "On unifying multi-block analysis with applications to decentralized process monitoring", Journal of Chemometrics, Vol. 15, N°9, pp. 715-742, 2001.

[27] G. Xu et T. Kailath. "Fast estimation of principal eigenspace using lanczos algorithm", SIAM Journal on Matrix Analysis and Applications, Vol. 15, N°3, pp. 975-994, 1994.

8. AUTHORS PROFILE

Chaouch Hanen received his enigineering diploma of industrual instrumentation and maintenance from l'INSAT-Tunis in 2007. In 2009, she obtained her mastery in measurment and instrumentation from the same university.She is currently Assistant in the Superior Institute of Applied Maths and Informatics in Kairouan (ISMAIK).His research interestsinclude Modelling, Control, Monitoring and Command biologic Systems.

Ouni Khaled received his License of Automatically Industrial Data Processing from ESST-Tunis. In 2007, he obtained his Mastery in Safety and Security of the Industrial Systems from ISSAT- Sousse.

He is currently Assistant in the Superior Institute of Applied Sciences and Technology in Kairouan (ISSATK). His research interests include Modelling, Control, Monitoring and Command Manufactory Systems.

Nabli Lotfi received his Mastery of Sciences and DEA from ENSET- Tunis - Tunisia in 1989 and 1991 respectively. In 2000, he obtained his doctorate degree in Industrial Automation: Automatic and Industrial Computing from the University of Sciences and Technologies of Lille France. He is currently Assistant professor of Electrical Engineering at the National School of Engineers of Monastir (ENIM)- Tunisia and a Master of conference candidate. His research interests include Modeling, Control, Monitoring and Command Manufactory Systems.