ECHAHID HAMMA LAKHDAR UNIVERSITY - EL-OUED
Under the Supervision of the DGRSDT and in collaboration with the CRTI
International Pluridisciplinary PhD Meeting (IPPM'20)
23-26, 2020 1st Edition, February
Theme: Modern Technology and Fineness Life

Improved Process Monitoring Based on Sparse Principal Component Analysis (SPCA)

Riadh Toumi*, Yahia Kourd**

* Laboratory of Electrical Engineering and Renewable Energy, Faculty of Science and Technology, Mohamed-Cherif Messaadia University, Algeria
** Faculty of Science and Technology, Mohamed-Cherif Messaadia University, Algeria
E-mail: r.toumi@univ-soukahras.dz, said-briki@univ-eloued.dz

Abstract

The principal components analysis was (PCA) intensively developed and widely applied in industrial processes for monitoring. The purpose of using the PCA is to reduce the extractable dimension of the still valid feature space to the most information in the primary dataset. Sparse Principal Component Analysis (SPCA) is a relatively new mechanism proposed for the creation of Principal Components (PCs) with sparse loads via a variance tradeoff. Using the SPCA, some of the loads on the PCs can be limited to zero. In this article, we applied this method for the diagnosis and monitoring of a biological process. The results obtained are discussed.

Keywords: Process Monitoring, Principal Component Analysis (PCA), Sparse Principal Component Analysis (SPCA), fault Detection and Isolation.

1. Introduction

Principal Component Analysis (PCA) is a statistical technique for identifying linear relationships between process variables and for analyzing and reducing the dimensionality of a large dataset. It consists of replacing the variables defined by new uncorrelated variables called Principal Components (PCs), with a smaller size and maximum variance. But the PCs are general linear sets of all input variables. Moreover, the coefficients of these sets are generally non-zero. Therefore, each major component obtained may have no practical significance. Thus, Sparse Principal Component Analysis (PCA) which is an extension of the PCA can solve this problem. It proposes is to produce PCs with sparse loads via a compromise of variances and some of the loadings on the PCs can be moderated to zero. Many methodologies have been proposed in the literature and among them: Sparse PCA via regularized SVD (SPCA-RSVD) was proposed [1] [2] to obtain sparse PCs, the equalization penalties favouring the scarcity of PC charges. Thus, validate and compare, as a priority, the choice of the degree of contrast as a synthesis parameter [5]. Also, another method that studies the rare PCA with LASSO or cardinality constraints to manufacture sparse load vectors called generalized power (GPower) has been proposed, this approach helps to surpass the other algorithms in computational speed [4]. Also, another technique was proposed which is Sparse Principal Component Analysis (SPCA) [3], its goal is to solve the regression problem by imposing LASSO (elastic mesh) constraints on the L1norm of the regression coefficients (scattered loads), so the optimization problem is determined with the cost a simple square fit and can be used in the case where m is much wider than the sample size and the desired NNZL can be specified for each component.

The goal of this paper is to use the Sparse PCA method for the diagnosis and monitoring of a biological process, using as data the measurements collected by the sensors available in the process installation in order to build our model. Section 2 describes Principal Component Analysis (PCA). In section 3, the analysis of sparse main components is developed. Sections 4 describe the detection and isolation of faults. And section 5, a simulation case study using a biological plant, where the results obtained are discussed.

2. Principal Component Analysis

Principal Component Analysis is a means of family data analysis on multivariate statistics. It is a geometric and statistical approach (explaining variability variance of data by independent research axes). So, when we have to compress a set of random variables, the best choice is the first axes of the PCA, under conditions of inertia or variance. It allows the expert to reduce the information to a small number of components in relation to the initial number of variables. PCA converts a correlated variable into a new variable that is linearly uncorrelated with each other [2] [6]. The converted variables are called Main Components (PCs) or Axes. The principal axes are the direction axes of the eigenvectors of the variable co-variance matrix, where the first axis associated with the enormous eigenvalue and the second axis orthogonal to the first is combined with the second eigenvalue, etc., then the least axis is that combined with the minimum eigenvalue. The first two or three main axes form the directions of the reduced space. The main space is part of the original data space.

However, the PCA is a method that reduces the number of variables essential to represent geometrically a phenomenon. From a group of *n* objects (individuals) in a space of *m* descriptors (variables), its objective is to discover a representation in a decreased space of dimensions *(l<<m)* which save"the best summary". The reduction is only possible if the initial variables m are not independent. This notion of independence is calculated by the correlation coefficients, it must that these coefficients are not nulls. The reduction is not a selection of the basic variables but by a definition of new variables (principals) obtained by combining the original variables that's why we call the PCA a linear method factorial. The mathematical tool connected with the PCA method is therefore established on linear algebra and matrix calculations.

Mathematically, we can define each variable, where *m* is number of variables. We can also keep the data in a matrix X of size *(n * m)* with x_{ij} represents the i^{th} measurement of the j^{th} variable.

The column vectors of *X* are $x_{i1}, x_{i2} \ldots\ldots x_{im}$ for *i=1. . . n* represent the j^{th} calculations at all instants of each variable. Similarly, the measurements at time *i* of *m* variables is represented by $e_{1j}, e_{2j} \ldots\ldots e_{nj}$ the line vectors of *X* for *j=1. . . m*. This data includes the calculations of variables (PH, temperature, dissolved oxygen...etc) at different instants. When we have only variables it's easy to define the dataset on a graph to *l* dimensions.

2.1. Mathematical calculation

We suppose *X*, a matrix of *n* number of observations for each variable and *m* variables. Where, the number of lines defines the data dimension. Before that, to make the independent result of the used units for each variable, a pretreatment is important to center and decrease the variables. That's why we believe, that we deduct from each column of the start set the mean μ_j and we divide on the standard deviation σ_j . We then get a new normalized matrix *Y* with its mean centered at zero. The measuring of the PCA is described in this way: Each column x_j of the new matrix centered reduced is given by:

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

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

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

In order to calculate YY^T returns to create a matrix have the sums of the deviations. YY^T multiply by *1/(n−1)* can get the matrix Σ where the elements placed on the diagonal correspond to the variance σ_{ij}^2 and other to the co-variance σ_i σ_j j. Σ is usually 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 way in which the data projection variance of the vector X is maximal, is described by the eigenvector p_i matching to the maximum eigenvalue.\\

The sub-vector space of dimension *l* which guarantee the maximum dispersion of observations is characterized by an orthonormal basis include by eigenvectors corresponding to the greatest eigenvalues of the matrix Σ. It is therefore probable reduce the dimension of the data representation by retaining from the previous expression only the $t_j p_j$ (*j*=1,…*l*) terms combined with the greatest eigenvalues. Hence The PCA decided an optimal transformation of the data matrix X:

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

Where $T \epsilon R^{n*m}$ T et $P \epsilon R^{m*m}$ are the matrixes of the principal components and the corresponding eigenvectors consequence from the spectral decomposition of the covariance matrix Σ. The relations (3) find their benefit when we reduce the dimension representation space. Once the number $l < m$ components to preserve is determined, the data matrix *X* can be approximated. That's why; the eigenvectors matrix is partitioned as:

$$P = \tilde{P}\,\hat{P} \quad where \quad \hat{C}^{(l)} = \hat{P}\,\hat{P}^T \tag{4}$$

The first *l* eigenvectors $\hat{P}$ create the principal space while the *(m - l)* last eigenvectors $\tilde{P}$ create the residual space. From the equation (3), we can make clear the part of the data explained by the first *l* eigenvectors and the residual part explained by the remaining components:

$$\hat{X} = \hat{C}^{(l)} X \quad where \quad \hat{C}^{(l)} = \hat{P}\,\hat{P}^T \tag{5}$$

and

$$E = X - \hat{X} = X\,(1 - \hat{C}^{(l)}) \tag{6}$$

The identification of the PCA model is subsequently to estimate its parameters by the decomposition on eigenvectors and eigenvalues of the matrix Σ and to decide *l* the number of principal components to conserve. In the absence of noise on the measurement, the null eigenvalues of Σ point to the presence of linear relations between the components of *X*. In the presence of noise on the measurement, the least eigenvalues compared to the other point to the

presence of linear or quasilinear relations between the different components of X. we look here the crucial role represented by the *l* number components in deciding the relations of redundancy between variables and also the difficulty of deciding this structural parameter.

3. Sparse PCA

Sparse principal component analysis (sparse PCA) is a specialized mechanism applied in statistical analysis, especially in the analysis of multivariate data sets. It's also known as development of the traditional method of PCA (principal component analysis) for the reduction of dimensionality of data by inserting sparsity constraint on the input variables [3]. Classic principal component analysis (PCA) employ a vector space convert to reduce multidimensional data sets to least dimensions. It finds linear combinations of input variables, and converts them into new variables know as principal components that coincide to directions of maximal variance in the data. The number of principal components generated by these linear combinations is generally much less than the number of input variables in the premier data set, nevertheless explaining most of the variance existing in the data [7]. A specific disadvantage of classic PCA is that the PC (principal components) are generally linear combinations of all input variables. SPCA (Sparse PCA) overcomes this disadvantage by achieving linear combinations that consist just a some input variables.

We summarized different popular SPCA as:

ScoTALASS [8]: for K-sparse PCs, the coefficient vector is the solution of the following optimization problem,

$$\max u^T \Sigma u \ , \|u\|_1, \quad u \perp u_j \ , \quad 1 \leq j \leq k-1 \quad , u \in R^p$$
$$s.t. \ \|u\|_2 = 1$$

Where u_j is the coefficient vector of the j^{th} sparse principal component PCs and t is the tune factor.

SPCA [3]: solve the regression-type optimization problem as follows:

$$\min_{A,B} \left\{ \sum_{i=1}^{n} \|x_i - A\Gamma^T x_i\|_2^2 + \lambda \sum_{j=1}^{k} \|\beta_j\|_2^2 + \sum_{j=1}^{k} \lambda_{1,j} \|\beta_j\|_1 \right\}$$
$$s.t \ A^T A = I$$

Where$A \in R^{NxK}$, x_i=1….n is sampling data.

$$\Gamma = (\beta_1, \ldots\ldots, \beta_k)$$

With I is the identity matrix, $\beta_i \in R_p$, *i = 1,… k.*

If A^* and B^* are the optimal solutions, then the columns of B^* are estimated coefficient vector for the first PCs.

sPCA-rSVD [9]: solve the following optimization problem

$$\min_{A,B} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{P} (X_{ij} - u_i v_j)^2 + P_\lambda(v) \right\}$$

Where X is the data matrix, $u = (u_1 \ldots u_n)^T$, $v = (v1 \ldots v_p)$ and is the regulation penalty. If u^* and v^* are the optimal solutions, then v^* is the coefficient vector of the first sparse PCs.

Sparse principal component analysis SPCA by the choice of norm [10]: solve the following optimization problem

$$\max u^T \Sigma u \ , u \in R^p \ , \quad \forall v \in R^p$$
$$s.t \ [(1-\lambda)\|v\|_2^2 + \lambda \|v\|_1^2]^{1/2} \leq 1$$

Where $\sum$ is the covariance matrix of X, u= u_1 ….u_p and u_j is the coefficient vector of the j^{th} sparse PCs, and λ∈[0, 1].

4. Fault Detection and Isolation

4.1. Fault Detection

The generality traditional methods of SPC (Statistical process control) provide control charts explainable only a reduced number of variables. Also, univariate charts specify quantitative information by disregard the impact of the correlation between variables. Accordingly purpose many of multivariate extensions of univariate approaches (CUSUM, Shewhart charts and EWMA) have been suggested in literature. The first multivariate charts have been developed explain the $\chi 2$ and the Hotelling T^2 maps. In this context, the typical index utilized for the detection of unusual operation is SPE index conjunction with the SPCA (Sparse PCA) model. The SPE index provides fault detection in the residual subspace [11]. Its term at the instant k is given by:

$$SPE\ (k) = e(k) - e(k)^T \ ) \qquad (7)$$

Where the vector of estimation errors is represented by:

$$e(k) = \tilde{x}(k) = x(k) - \widehat{x(k)} \quad (8)$$

Where the estimated data at the instant k has the following formula:

$$\widehat{x(k)} = \varphi\big(t(k)\big) = \varphi\left(\phi\big(x(k)\big)\right) \quad (9)$$

With that indicator, the process is see in normal operation to the k^{th} observation if:

$$SPE\ (k) < \delta_\alpha^2$$
$$\delta_\alpha^2 = g\ \chi_{h,\alpha}^2$$

δ_α^2 : Represents a detection threshold of SPE, the coefficient g can be calculated by the variance v and the average m, h is the degree of freedom:

$$g = {}^{v}/_{2m}$$
$$h = {}^{2m^2}/_{v}$$
$$\delta_\alpha^2 = {}^{v}/_{2m} * \chi_{h,\alpha}^2$$

α: is the *predefined* significance level.

4.2. Fault localization

The generality traditional methods utilized in the context of PCA for fault location approaches are established on the account of contributions. The essence of these methods is to account the contributions of the several variables to the indicators utilized for the detection of defects. Therefore, the contributions principle is mostly, established on the share quantification of every variable in the account of the detection index SPE [12]. The contribution $Cont_j^{SPE}$ of j^{th} variable at the instant k is given by:

$$Cont_j^{SPE}(k) = \tilde{x}_j(k) \quad = (x_j - \hat{x}_j)^2 \quad (10)$$

5. Application study

5.1. Process description

With a view to prove the effectiveness of the proposed scheme, a biological system, which contains of a continuous aerated bioreactor for WWT waste water treatment in pulp and paper industry is chosen as an exemplary nonlinear process, Fig. 1 . The bioreactor consists a mixed microbial population growing in a mixed of two types of substrates, an xenobiotic one and energetic [13]. The prime pollutant is the xenobiotic substrate because the energetic substrate is facilely biodegradable.

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

c(t)[g/l]: the biomass concentration.

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

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

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

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

μ_{sm} and μ_{em}: the growth rates of the substrates. $Y_{c/s}$ and $Y_{c/e}$: yield coefficients.

K_s and K_e: Michaelis-Menten constants. a_s and a_e : Inhibiting coefficients.

The differential equations describing the performance of the process are:

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

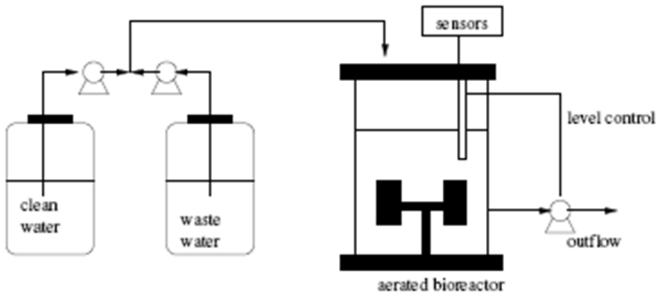

Fig.1. Schematic representation of the aerated bioreactor for wastewater treatment

To generate data in an open loop mode. The collected data for the normal condition is utilized to construct the monitoring model of competitive principal component analyzers. The data vector of the k^{th} sample is:

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

5.1. Simulations results

In this context, we will give the results acquired from the developed process of diagnosis for defect affecting the sensors. To demonstrate the method and illustrate its benefits in fault detection and isolation, we use dataset from biological processes containing five variables and 450 samples collected. They are used to build our model. We observe the evolution of the SPE in a normal state after having built our model as illustrated in figure 2. A fault is injected at a precise instant, we observe the evolution of SPE in Figure 3, we notice the appearance of a fault from 120 to 240 samples exactly at the time of the injection of the fault. EWMA was also used for filter impact of outliers and noise as shown in Figure 3.

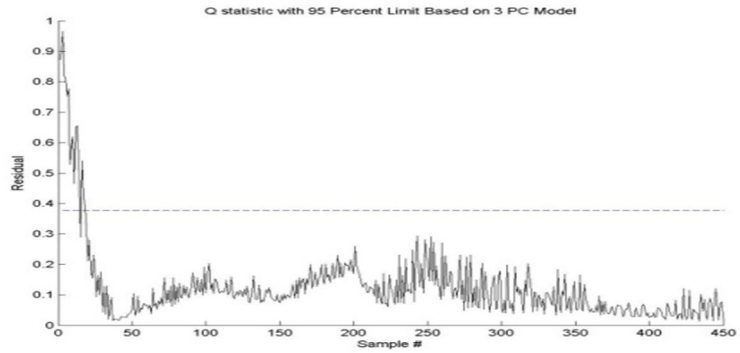

Fig.2. Evolution of the SPE index during normal stat

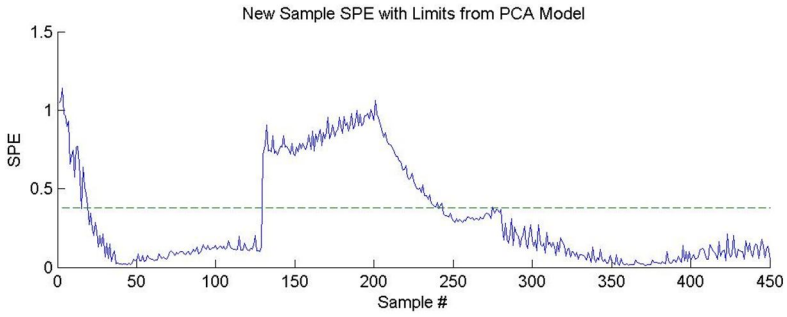

Fig.3. Evolution of the SPE index with a default in the 3th variable at sample 130

After detecting a defect, it is necessary to know which sensor is defective. In our case it is the third sensor that represents the concentration of the energetic substrate as a motive in Figure4.

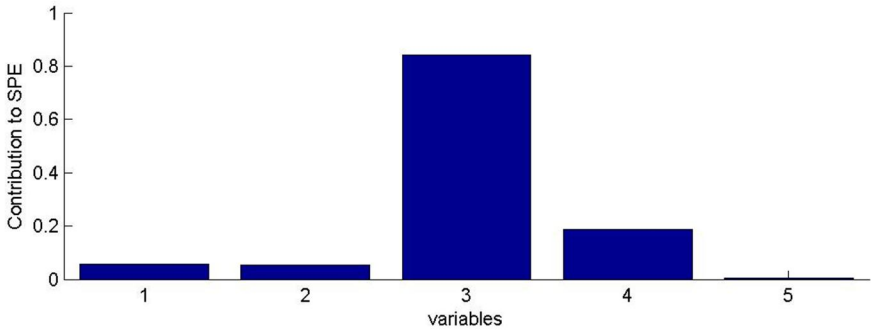

Fig.4. Evolution of filtered SPE and the fault localization affecting the 3th sensor measuring the concentration of the energetic substrate at sample 120

6. Conclusion

In late years, the fault detection and diagnosis approaches have been widely developed and utilized to Improved process monitoring; especially the fault detection based on PCA Principal Components Analysis which does not demand previous knowledge about the process mechanism had known a great progress and has been vastly developed. The PCA is a linear modelling tool based on the chosen of an optimal number of principal components . We insert techniques to derive PCs with sparse loadings and the interpretability of PC loadings and we used to improve process monitoring based on these principal components .

For monitoring a procedure, the statistical SPE is applied to detect anomaly, to identify the defect variables, we utilized diagnosis algorithms like the localization based on the contribution plots. We applied a filter to the SPE, therefore adds an important feature for sensor fault localization because decrease the false alarms.

The main idea of this work is to improve process monitoring based on sparse principal component analysis (SPCA) and used in biological processes, the simulation results gained in this paper display the effectiveness of the proposed method. Although the efficacy of this approaches, it can be improved for the better in future research.

References

1. Martin, E. B., Morris, A. J., & Zhang, J. (1996). Process performance monitoring using multivariate statistical process control. IEE Proceedings-Control Theory and Applications, 143(2), 132-144.
2. Jolliffe, I. T. (2002). Principal components in regression analysis. Principal component analysis, 167-198.
3. Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. Journal of computational and graphical statistics, 15(2), 265-286.
4. Shen, H., & Huang, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. Journal of multivariate analysis, 99(6), 1015-1034.
5. Journée, M., Nesterov, Y., Richtárik, P., & Sepulchre, R. (2010). Generalized power method for sparse principal component analysis. Journal of Machine Learning Research, 11(Feb), 517-553.
6. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.
7. Nomikos, P., & MacGregor, J. F. (1995). Multivariate SPC charts for monitoring batch processes. Technometrics, 37(1), 41-59.
8. Jolliffe, I. T., Trendafilov, N. T., & Uddin, M. (2003). A modified principal component technique based on the LASSO. Journal of computational and Graphical Statistics, 12(3), 531-547.
9. Witten, D. M., Tibshirani, R., & Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics, 10(3), 515-534.
10. Xie, L., Lin, X., & Zeng, J. (2013). Shrinking principal component analysis for enhanced process monitoring and fault isolation. Industrial & Engineering Chemistry Research, 52(49), 17475-17486.
11. Mendaci, K., Ramdani, M., & Benzaraa, T. (2013, April). Nonlinear multivariate statistical process monitoring of a water treatment plant. In 2013 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO) (pp. 1-6). IEEE.
12. Chaouch, A., Bouzenad, K., & Ramdani, M. (2014). Enhanced Multivariate Process Monitoring for Biological Wastewater Treatment Plants. Int. J. Electr. Energy, 2, 131-137.
13. Ramdani, M., & Mendaci, K. (2013, August). Nonlinear Statistical Process Monitoring based on Competitive Principal Component Analysis. In 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13). Atlantis Press.