FAULT DETECTION USING PRINCIPAL COMPONENT ANALYSIS (PCA) IN A WASTEWATER TREATMENT PLANT (WWTP)

Diego Garcia-Alvarez

University of Valladolid
Valladolid, Spain
dieggar@cta.uva.es

Abstract

In this paper Principal Components Analysis (PCA) is used for detecting faults in a simulated wastewater treatment plant (WWTP). PCA is a multivariate statistical technique used in multivariate statistical process control (MSPC) and fault detection and isolation (FDI) perspectives. PCA reduces the dimensionality of the original historical data by projecting it onto a lower dimensionality space. It obtains the principal causes of variability in a process. If some of these causes changes, it can be due to a fault in the process. False detected alarms due to measured disturbances are treated using Switch-PCA.

Index Terms: Fault detection, Principal Component Analysis (PCA), Wastewater treatment, Multivariate Statistical Process Control (MSP).

I. INTRODUCTION

Actually there are several multivariate statistical methods for the analysis of process. Some of this methods have recently been used successfully for monitoring and fault detection. These methods are useful because the safe operation and the production of high quality products are same of the main objectives in the industry. Classical and advanced control techniques have resolved a large number of problems, but when a special cause occurs in a process, it can not operate under control. The development of an industrially reliable online scheme for such processes would be a step toward effectiveness and robustness.

Classical Statistical Process Control (SPC) uses typical control charts, such as Shewhart charts, cumulative sum (CUSUM) charts, and exponentially weighted moving average (EWMA) charts for monitoring a single variable. When univariate control charts are applied to multivariate systems, with a lot of variables, the results are improper when a fault or an abnormality in the operation occurs, some of these charts alarm in a short period of time or simultaneously. This situation is produced because the process variables are correlated, and a special cause can affect more than one variable at the same time. Multivariate Statistical Process Control (MSPC) uses latent variables instead of every measured variable. All these methods use historical databases to calculate empirical models that describe the trend of the whole system. They are able to extract useful information inside the historical data, calculating the relationship between the variables. When a problem appears, it changes the covariance structure of the model and it can be detected.

Multivariate statistical process control (MSPC) approach, and principal component analysis (PCA) in particular, have been investigated to face this problem. Jackson and Mudholkar investigated PCA as a tool of MSPC [8] two decades ago. PCA can be described as a method to project a high-dimensional measurement space onto a space with significantly fewer dimensions [2]. PCA finds linear combinations of variables that describe major trends in data set. Mathematically, PCA is based on an orthogonal decomposition of the covariance matrix of the process variables along the directions that explain the maximum variation of the data.

PCA has been studied from two perspectives, one of these is the cited MSPC, and the other is the fault detection and isolation (FDI) perspective, this perspective is discussed by Venkatasubramanian [17]. The author divides the fault detection and diagnosis techniques in three parts: quantitative model-based methods, qualitative models and search strategies and process history based methods. PCA falls in the third category because it uses historical databases to derive the statistical model (PCA model).

The charts most commonly used with PCA techniques are Hotelling statistics, T2, and the sum of squared residuals, SPE, or *Q* statistic. The *T2* statistic is a measure of the variation in the PCA model and the Q statistic is a measure of the amount of variation not captured by the PCA model.

The purpose of this article is to implement a method for fault detection using principal component analysis method and to apply it in wastewater treatment plant (WWTP). Theoretical aspects of PCA will be presented and the wasterwater treatment plant, the considered faults and the results obtained will be explained and discussed.

There are several groups work in fault detection in waste-water treatment plants using PCA [14] or using another fault detection approaches [5].

II. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a vector space transformation often used to transform multivariable space into a subspace which preserves maximum variance of the original space in minimum number of dimensions. The measured process variables are usually correlated to each other. PCA can be defined as a linear transformation of the original correlated data into a new set of uncorrelated data, so that, PCA is a good technique to transform the set of original process variables in a new set of uncorrelated variables that explain the trend of the process.

Consider a data matrix $X \in R^{n\times m}$ containing *n* samples of *m* process variables collected under normal operation. This matrix must be normalized to zero mean and unit variance with the scale parameter vectors *x* and *s* as the mean and variance vectors respectively. Next step to calculate PCA is to construct the covariance matrix R *:*

$$R = \frac{1}{n-1} X^t X \qquad (1)$$

and performing the SVD decomposition on R :

$$R = V\Lambda V^T, \qquad \textbf{(2)}$$

where Λ is a diagonal matrix that contains in its diagonal the eigenvalues of *R* sorted in decreasing order ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0$). Columns of matrix V are the eigenvectors of R. The transformation matrix $P \in R^{m\times n}$ is generated choosing *a* eigenvectors or columns of V corresponding to *a* principal eigenvalues. Matrix P transforms the space of the measured variables into the reduced dimension space.

$$T = XP \qquad (3)$$

Columns of matrix P are called *loadings* and elements of T are called *scores.* Scores are the values of the original measured variables that have been transformed into the reduced dimension space.

Operating in equation (3), the scores can be transformed into the original space.

$$\hat{X} = TP^T \qquad (4)$$

The residual matrix E is calculated as:

$$E = X - \hat{X} \qquad (5)$$

Finally the original data space can be calculated as:

$$X = TP^T + E \qquad (6)$$

It is very important to choose the number of principal components a, because TP^T represents the principal sources of variability in the process and E represents the variability corresponding to process noise. There are several proposed procedures for determining the number of components to be retained in a PCA model as [7, 18]:

a) The SCREE procedure [7]. It is a graphical method in which one constructs a plot of the eigenvalues in descending order and looks for the *knee* in the curve. The number of selected components are the components between the high component and the *knee.* An example of this graph is shown in fig. 2.

b) Cumulative Percent Variance (CPV) approach [18]. It is a measure of the percent variance ($CPV(a) \geq 90\%$) captured by the first *a* principal components is adopted:

$$CPV(a) = \frac{\sum_{i=1}^{a} \lambda_i}{trace(R)} 100 \qquad \textbf{(7)}$$

c) Cross validation.

A. Statistics for monitoring

Having established a PCA model based on historical data collected when only common cause variation are present, multivariate control charts based on Hotelling's T^2 and square prediction error (SPE) or Q can be plotted. The monitoring can be reduced to this two variables (T^2 and Q) characterizing two orthogonal subsets of the original space. T^2 represents the major variation in the data and Q represents the random noise in the data. T^2 can be calculated as the sum of squares of a new process data vector *x:*

$$T^2 = x^T P \Lambda_a^{-1} P^T x, \qquad (8)$$

where Λ_a is a squared matrix formed by the first *a* rows and columns of Λ.

The process is considered *normal* for a given significance level α if:

$$T^2 \leq T_a^2 = \frac{(n^2-1)a}{n(n-a)} F_a(a, n-a), \qquad (9)$$

where $F_a(a, n-a)$ is the critic value of the Fisher-Snedecor distribution with n and $n-a$ degrees of freedom and a the level of significance. a takes values between 90% and 95%.

T^2 is based on the first *a* principal components so that it provides a test for derivations in the latent variables that are of greatest importance to the variance of the process. This statistic will only detect an event if the variation in the latent variables is greater than the variation explained by common causes.

New events can be detected by calculating the squared prediction error *SPE* or Q of the residuals of a new observation. Q statistic [8, 7], is calculated as the sum of squares of the residuals. The scalar value Q is a measurement of *goodness of fit* of the sample to the model and is directly associated with the noise:

$$Q = r^T r \tag{10}$$

with $r = \left(I - PP^T\right)x$.

The upper limit of this statistic can be computed as the next form:

$$Q_a = \theta_1 \left[\frac{h_0 c_a \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \tag{11}$$

with $\theta_i = \sum_{j=a+1}^{m} X_j^i$, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$,

where c_a is the value of the normal distribution with a the level of significance.

When an unusual event occurs and it produces a change in the covariance structure of the model, it will be detected by a high value of Q.

B. PCA Monitoring

To implement a monitoring and fault detection system based on PCA, it is necessary to consider two tasks:

1. OFF-LINE. Acquire training data which represents normal process operations. Scale the training data and obtain the scale parameter vectors *x* and s. Carry out SVD to obtain PCA model. Determine the number of principal components and the upper control limits for T^2 and Q statistics.

2. ON-LINE.

a) Obtain the next testing sample *x,* and scale it using the scale parameter vectors *x* and *s.*

b) Evaluate the T^2 and Q statistics using the obtained PCA model. If one of these exceeds the upper limit, this measurement is considered an alarm. If there are some consecutive established number of alarms, an uncommon event has occurred.

c) Repeat from step 2.

III. APPLICATION

The approach presented in this paper has been tested in a simulated wastewater treatment plant (WWTP). This plant is based on the COST benchmark [1, 3]. This benchmark was development for the evaluation and comparison of different activated sludge wastewater treatment control strategies. The model is implemented using MATLAB© and SIMULINK©.

This model plant utilizes a dynamic model of activated sludge process which is known as activated sludge model no. 1 or ASM1.

Fig. 1 shows an overview of this plant. It is composed of two-compartment activated sludge reactor consisting of two anoxic tanks followed by three aerated tanks. This type of plants combine nitrification with predenitrification in a configuration that is usually built for achieving biological nitrogen removal in full-scale plants. The reactor is followed by a secondary settler. The settler is modeled as a 10 layers non-reactive unit. The 6th layer is the feed layer. Table I shows the physical parameters of the plant.

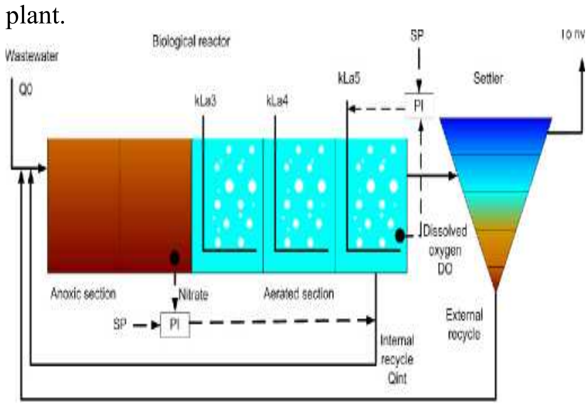

Fig. 1. General overview of the waste water treatment plant (WWTP)

The used influent was the dry influent data file [3]. In this file, the variation of influent flow is between 15000 − 3500 m^3/d. The plant, as Fig. 1 shows, has two reflux:

1. External reflux, from settler to input, it is approximately equal to influent flow.

Table 1

Physical parameters

Elements	Values	Units
Volume - Anoxic section	2000 (2×1000)	m^3
Volume - Aerated tank	4000 (3×1333)	m^3
Volume - Settler (10 layers)	6000	m^3
Area - Settler	1500	m^2
Height - Settler	4	m

2. Internal reflux, from the last aerated tank to input, it is approximately equal to three times the influent flow, but it is a controlable variable.

The objective of the control strategy is to control the dissolved oxygen level in the aerated reactor by manipulating of the oxygen transfer coefficient (K_La_5) and to control the nitrate level in the anoxic tank by manipulating of the internal recycle flow rate. Controllers are PI type. Tab. 2 shows the principal controllers settings.

Table 2

Controllers settings

Variables	Oxygen loop	Nitrate loop
Controller type	PI	PI
Controlled variable	DO $[g/m^3]$	S_{NO} $[gN/m^3]$
Manipulated variable	K_La5 $[1/hr]$	Q_{int} $[m^3/d]$
Setpoint	2 $[g/m^3]$	1 gN/m^3

The model of the plant is formed by 13 state variables. The involved variables are concentrations of:

1. Alkalinity *(S_{ALK})*.
2. Soluble biodegradable organic nitrogen *(S_{ND})*.
3. Ammonia nitrogen *(S_{NH})*.
4. Nitrate *(S_{NO})*.
5. Dissolved oxygen *(S_O)*.
6. Readily biodegradable substrate *(S_S)*.
7. Active autotrophic biomass *($X_{B,A}$)*.
8. Active heterotrophic biomass *($X_{B,H}$)*.

9. Particulate biodegradable organic nitrogen (X_{ND}).

10. Particulate products from biomass decay (X_P).

11. Slowly biodegradable substrate (X_S).

12. Particulate inert organic matter (X_I).

13. Soluble inert organic matter (S_I).

In this case, three faults have been considered. They are not sensors or actuators faults, they are faults in the process. The faults considered are:

– Toxicity shock. This fault is due to the reduction of the normal growth of heterotrophic organisms. This type of fault can be produced by toxic substances into the water coming from textile industries or pesticides. This fault is simulated by reducing the maximum heterotrophic growth rate (μ_H).

– Inhabitation. This fault can be produced by hospital waste that can contain bactericides, or metallurgical waste that can contain cyanide. This type of fault is due to the reduction of normal growth of the heterotrophic organisms and the increase in the decay factor of this type of organisms. This fault is similar to toxicity shock but it is more drastic. In this case, the fault is caused by reducing the maximum heterotrophic growth rate (μ_H) and by increasing the heterotrophic decay rate (b_H).

– Bulking. This type of fault is produced by the growth of filamentous microorganisms in the active sludge. This phenomenon causes impossibility of decantation in the settler. To simulate this fault the settling velocity in layer (v_{sj}) is reduced.

More information about these parameters and mathematical models can be consulted in [3].

Using this dynamic model the results were obtaining in steady state. For this, the plant model has to simulate $100-150$ days in open-loop configuration and determines this steady state. Then, the simulation in close-loop is simulated 14 days and faults are caused in the 7^{th} day. The samples for monitoring experiments were taken 100 times per day.

The selected variables to calculate principal components analysis (PCA) are the first eleven state variables and the effluent flow rate (Q_0). The concentration of particulate inert organic matter (X_I) and soluble inert organic matter (S_I) are not relevant to this study [16].

The number of principal components, calculated using CPV approach with 95% maximum variance level, are five, but fig. 2 shows that seven principal components can be a best option because they capture more variability of process.

The process monitoring under toxicity shock fault can be seen in fig. 3. Both statistics, T^2 and Q, arise their thresholds when the fault occurs. In this case, the Q statistic detects this fault better than T^2 statistic as this figure shows.

The inhabitation fault detection is more effective than the detection of toxicity shock fault because this type of fault is more drastic as it is possible to see in fig. 4. Finally, the bulking fault detection using PCA is shown in fig. 5.

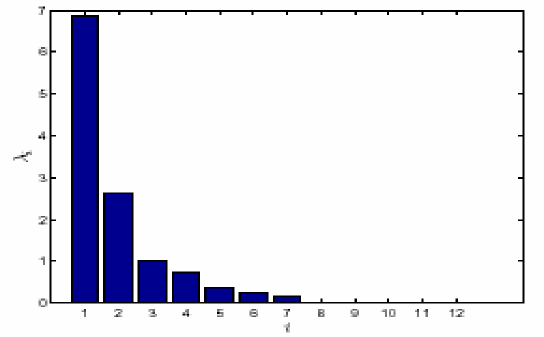

Fig. 2. The SCREE graph for principal component selection

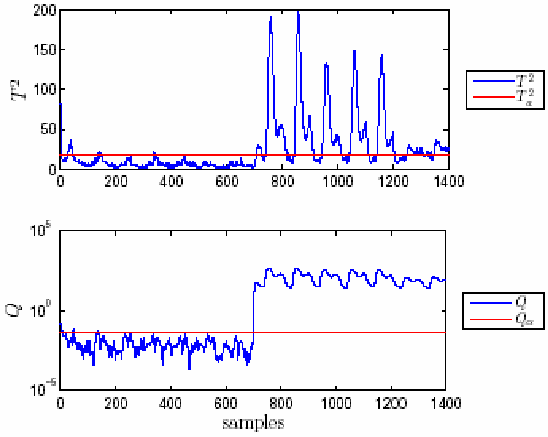

Fig. 3. Toxicity shock fault detection.
Logarithmic scale for Q statistic

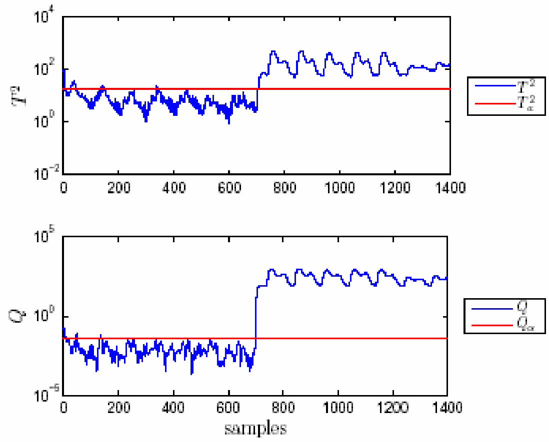

Fig. 4. Inhabitation detection.
Logarithmic scale for T^2 and Q statistics

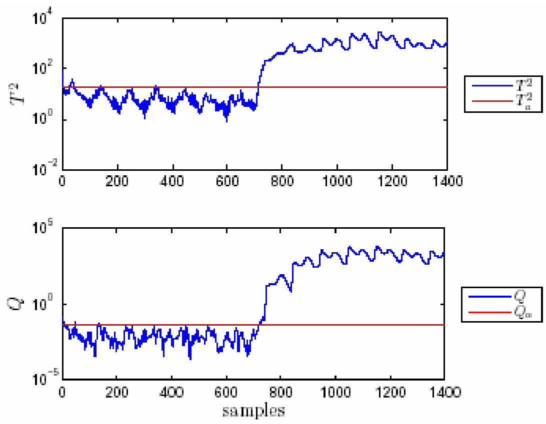

Fig. 5. Bulking fault detection.
Logarithmic scale for T^2 and Q statistics

IV. DISTURBANCES

Principal Component Analysis (PCA) is one of the most popular MSPC monitoring methods. However, it has some shortcomings, one of these is that PCA is not suited for monitoring processes that display non-stationary behavior. Another limitation of PCA is that most processes run in different conditions and modes. Using conventional PCA approach in this type of processes can produce an excessive number of false alarms or missed faults, because these grade transitions from one to another operation mode can break the correlation between the variables. Also measured disturbances can be detected as faults.

There are several proposed solutions that deal with this open problem, such as multi-scale PCA (MSPCA) [13], adaptive PCA (APCA) [18], recursive PCA [12], exponentially weighted PCA (EWPCA) [11], dynamic PCA [10] and Nonlinear PCA using autoassociative neural networks [9]. These proposed solutions fall in three different categories [6, 15]:

1. Build a PCA model for each operation mode.
2. Update the model to reflect the changes in the operation modes.
3. Develop a conventional PCA model to account for all such changes.

The proposed application can run under three different weather conditions: dry, rain and storm weather [3]. The example treated above was simulated under dry weather condition. When the conditions are rainy or stormy, the disturbances due to big volume of influent flow are detected as a fault by both monitoring statistics. Fig. 6 shows a false fault detection when the weather are rainy.

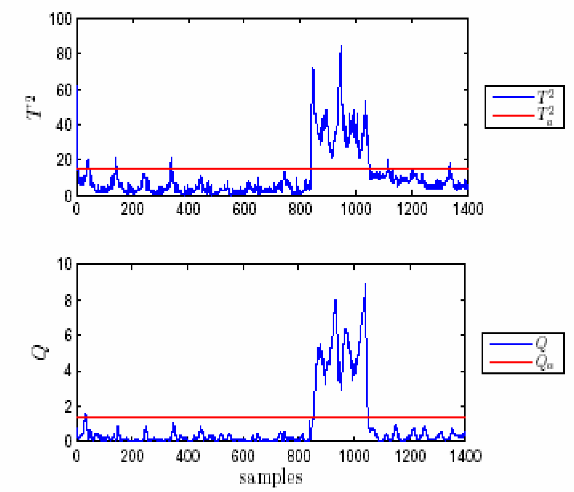

Fig. 6. False fault detected due to rain weather

In this case the used method to face with this limitation falls in the first direction: *build a PCA model for each operation mode.* Two PCA models are built: one PCA model for dry weather and another PCA model for rainy and stormy weather.

A switch structure can be used to decide what PCA model has to be used to monitor the process. In the monitoring phase for each new sample the switch structure checks the measured influent flow and applies the corresponding local PCA model and upper limits of T^2 and Q statistics. This technique is called Switch-PCA [4].

In fig. 7 the plant under rain weather is monitored using Switch-PCA. In this case when the volume of influent flow is greater than a threshold the switch structure changes the current PCA model for a PCA model corresponding to a rain weather conditions. In this case, the false fault due to rain weather is not detected.

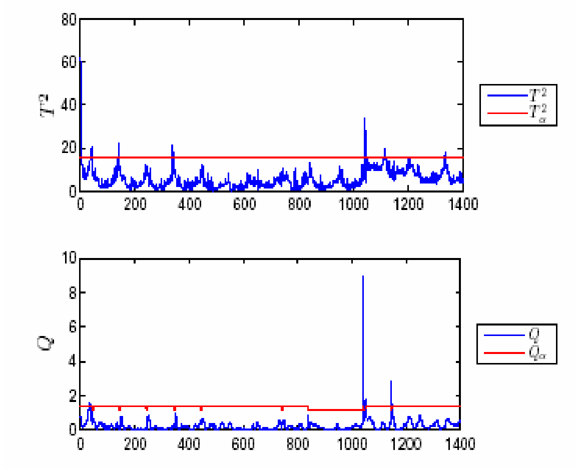

Fig. 7. Rain weather monitoring using Switch-PCA

V. CONCLUSIONS AND ACKNOWLEDGEMENTS

This work proposes an approach to face the fault detection using statistical techniques, concretely, the principal component analysis (PCA).

The approach has been proved in a simulated waterwaste treatment plat (WWTP) based on the COST benchmark. The considered faults are critical process faults that affect some plant parameters. Data are collected from the plant for normal conditions in order to calculate the PCA model and the thresholds of the T^2 and Q statistics, used in order to detect the faults.

Finally, the false faults due to measured disturbances are faced using a technique based in a switching structure. Off-line, different weather conditions are identified and a PCA model is built for each weather condition. On-line, the current weather condition is detected and it is monitored using the corresponding local PCA model.

I would like to thank Maria Jesus de la Fuente for his help and support to carry out this article.

REFERENCES

[1] J. Alex, L. Benedetti, J. Copp, K. Gernaey, U. Jeppsson, I. Nopens, M. Pons, L. Rieger, C. Rosen, J. Steyer, P. Vanrolleghem, and S. Winkler, *Benchmark Simulation Model no. 1 (BSM1),* Dept. of Industrial Electrical Engineering and Automation. Lund University, 2008.

[2] L. Chiang, E. Russell, and R. Braatz, *Fault Detection and Diagnosis in Industrial Systems.* Nueva York: Springer, 2000.

[3] J. Copp, *The COST Simulation Benchmark: Description and Simulator Manual (a product of COST Action 624 & COST Action 682),* European Cooperation in the field of Scientific and Technical Research,

[4] D. Garcia-Alvarez, M. Fuente, and P. Vega, "Fault detection in processes with multiple operation modes using switch-pca and analysis of grade transitions," *Proceding of the European Control Conference 2009, Budapest, Hungary (Acepted),* 2009.

[5] A. Genovesi, J. Harmand, and J. Steyer, "Integrated fault detection and isolation: Application to a winery's wastewater treatment plant," *Applied Intelligence,* vol. 13, pp. 59-76, 2000

[6] D. Hwang and C. Han, "Real-time monitoring for a process with multiple operating modes," *Control Engineering Practice,* vol. 7, pp. 891-902, 1999.

[7] J. Jackson, *A user's guide to principal components.* Wiley, 1991.

[8] J. Jackson and G. Mudholkar, "Control procedures for residuals associated with principal component analysis," *Technometrics,* vol. 21, pp. 341-349, 1979.

[9] M. Kramer, "Nonlinear principal component analysis using autoasso-ciative neural network," *The American Institute of Chemical Engineers Journal,* vol. 37, pp. 233-243, 1991.

[10] W. Ku, R. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and intelligent laboratory systems,* vol. 30, pp. 179-196, November 1995.

[11] S. Lane, E. Martin, A. Morris, and P. Gower, "Application of exponentially weighted principal component analysis for the monitoring of a polymer film manufacturing process," *Transactions of the Institute of Measurement and Control,* vol. 25, pp. 17-35, 2003.

[12] W. Li, H. Yue, S. Valle-Cervantes, and S. Qin, "Recursive pca for adaptative process monitoring," *Journal of Process Control,* vol. 10, pp. 471–486, 2000.

[13] M. Misra, H. Yue, S. Qin, and C. Ling, "Multivariable process monitoring and fault diagnosis by multi-scale pca," *Computers & Chemical Engineering,* vol. 26, pp. 1281-1293, Septiembre 2002.

[14] C. Rosen and J. Lennox, "Multivariate and multiscale monitoring of wastewater treatment operation," *Water research,* vol. 35, pp. 3402-3410, 2001.

[15] D. Tien, K. Lim, and L. Jun, "Compartive study of pca approaches in process monitoring and fault detection," *The 30th annual conference of the IEEE industrial electronics society,* pp. 2594-2599, November 2-6 2004.

[16] R. Tomita, S. Park, and O. Sotomayor, "Analysis of activated sludge process using multivariate statistical tools - a pca approach," *Chemical Engineering Journal,* vol. 90, pp. 283-290, 2002.

[17] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, and K. Yin, "A review of process fault detection and diagnosis. part iii: Process history based methods," *Computers & Chemical Engineering,* vol. 27, pp. 327-346, 2003.

[18] D. Zumoffen and M. Basualdo, "From large chemical plant data to fault diagnosis integrated to decentralized fault tolerant control: pulp mill process application," *Industrial & Engineering Chemistry Research,* vol. 47, pp. 1201-1220, 2007.