Fault Detection and Diagnosis using Multivariate Statistical Techniques in a Wastewater Treatment Plant. ⋆

D. Garcia-Alvarez * **M.J. Fuente** * **P. Vega** ** **G. Sainz** *

* *Department of Systems Engineering and Automatic Control, University of Valladolid, 47011 Valladolid, Spain*
[dieggar@cta,maria@autom,gresai@eis].uva.es
** *Department of Computer Science and Control, University of Salamanca, ETSII, 37700 Bejar (Salamanca), Spain* pvega@usal.es

Abstract: In this paper Principal Components Analysis (PCA) is used for detecting faults in a simulated wastewater treatment plant (WWTP). Diagnosis tasks are treated using Fisher discriminant analysis (FDA). Both techniques are multivariate statistical techniques 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 change, it can be due to a fault in the process. FDA provides an optimal lower dimensional representation in terms of a discriminant between classes of data, where, in this context of fault diagnosis, each class corresponds to data collected during a specific and known fault. A discriminant function is applied to diagnose faults using data collected from the plant.

Keywords: Fault detection, Fault diagnosis, Statistical process control, Wastewater treatment plant, Discriminant analysis

1. INTRODUCTION

Multivariate statistical methods for the analysis of process data have recently been used successfully for monitoring and fault detection. The safe operation and the production of high quality products are two of the main objectives in industry. Modern control techniques have resolved many problems, but when a special cause occurs in a process, it cannot operate under control. The development of an industrially reliable online scheme for such processes would be a step toward effectiveness and robustness.

Conventional univariate Statistical Process Control (SPC) uses typical control charts, such as Shewhart charts, for monitoring a single variable. When univariate control charts are applied to multivariate systems, with hundreds of variables, the results are improper because, when there is a fault or an abnormality in the operation, several of these charts set off an alarm in a short period of time or simultaneously. This situation is 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 system's trend. They are able to extract useful information from the historical data, calculating the relationship between the variables. When a problem appears, it changes the covariance structure of the model and can be detected.

Multivariate statistical approaches, and principal component analysis (PCA) in particular, have been investigated to deal with this problem. Jackson and Mudholkar investigated PCA as a tool of MSPC (Jackson and Mudholkar, 1979) two decades ago. The objective of this approach is to reduce the dimensionality of the original historical data by projecting it onto a lower dimensionality space. PCA finds linear combinations of variables that describe major trends in a 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 can be studied from two perspectives, one is the cited MSPC, and other is the fault detection and isolation (FDI) perspective, which is discussed by Venkatasubramanian (Venkatasubramanian et al., 2003a,b,c). The author divides the fault detection and diagnosis techniques into three parts: quantitative model-based methods, qualitative models and search strategies and process history-based methods. PCA falls into the third category because it uses historical databases to derive the statistical model (PCA model) (Hwang and Han, 1999; Kourti, 2002; Tien et al., 2004).

The charts most commonly used with PCA techniques are Hotelling statistics, T^2, and the sum of squared residuals, SPE, or Q statistic. The T^2 statistic is a measure of the variation in the PCA model and the Q statistic is a

⋆ This work was supported in part by the national research agency of Spain (CICYT) through the project DPI2006-15716-C02-02 and the regional government of Castilla y Leon through the project VA052A07

measure of the amount of variation not captured by the PCA model.

Once the fault is detected using monitoring techniques, it can be diagnosed by determining the fault region in which the observations are located. The approach used in this paper for fault diagnosis is pattern classification. When the data collected during the *out-of-control* operations have been previously diagnosed, the data can be categorized into separate classes when each class pertains to a particular fault (Chiang et al., 2000).

Fisher discriminant analysis (FDA) is a linear pattern classification method used to find the linear combination of features which best separate two or more classes. It is an empirical method based on observed attributes over the collected examples. FDA provides an optimal lower dimensionality representation in terms of a discriminant between classes of data, where, for fault diagnosis, each class corresponds to data collected during a specific, known fault. FDA has been studied in detail in the pattern classification literature (Duda et al., 2001), but its use for analyzing chemical process data had not been explored until recently (Chiang et al., 2000; He et al., 2005; Fuente et al., 2008).

The purpose of this article is to implement a method for fault detection and diagnosis using multivariate statistical methods and to apply it to a wastewater treatment plant (WWTP). Theoretical aspects of PCA and FDA will be presented and finally the wasterwater treatment plant, the considered faults and the results obtained will be explained and discussed.

2. 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 a 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, PCA is a good technique to transform the set of original process variables into a new set of uncorrelated variables that explain the trend of the process.

Consider a data matrix $X \in \Re^{n\times m}$ 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 $\bar{x}$ and s as the mean and variance vectors respectively. Then next step to calculate the PCA is to construct the covariance matrix R:

$$R = \frac{1}{n-1} X^T X \tag{1}$$

and to perform the SVD decomposition on R:

$$R = V \Lambda V^T \tag{2}$$

where Λ is a diagonal matrix that contains the eigenvalues of R in its diagonal 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 $P \in \Re^{m\times a}$ is generated by 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.

$$T = XP \tag{3}$$

The columns of matrix P are called *loadings* and the 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.

$$\hat{X} = TP^T \tag{4}$$

The residual matrix E is calculated as:

$$E = X - \widehat{X} \tag{5}$$

Finally the original data space can be calculated as:

$$X = TP^T + E \tag{6}$$

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, such as Zumoffen and Basualdo (2007) and Jackson (1991):

a) The SCREE procedure (Jackson, 1991): 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 Zumoffen and Basualdo (2007). A measure of the percent variance ($CPV(a) \geq 90\%$) captured by the first a principal components is adopted:

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

c) Cross validation.

Having established a PCA model based on historical data collected when only common cause variations 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 these 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 the squares of a new process data vector x:

$$T^2 = x^T P \Lambda_a^{-1} P^T x \tag{8}$$

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:

$$T^2 \leq T_\alpha^2 = \frac{(n^2-1)a}{n(n-a)} F_\alpha(a, n-a) \tag{9}$$

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

T^2 is based on the first a principal components so that it provides a test for deviations in the latent variables that are of the 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. The Q statistic (Jackson and Mudholkar (1979), Jackson (1991)) is calculated as the sum of the 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:

$$r = (I - PP^T)x$$

The upper limit of this statistic can be computed as follows:

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

with:

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

where c_α is the value of the normal distribution, with α being 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.

3. FISHER DISCRIMINANT ANALYSIS

For fault diagnosis, data collected from the plant during specific faults are categorized into classes, where each class contains data representing a particular fault. Define n as the number of observations, m as the number of measurement variables, p as the number of classes and n_j as the number of observations in the j^{th} class. The training data for all classes have been stacked into the matrix $X \in \Re^{n \times m}$. The total-scatter matrix is:

$$S_t = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \tag{12}$$

where $\bar{x}$ is the total mean vector whose elements correspond to the means of the columns of X. Let the matrix X_j be defined as the set of vectors x_j which belong to class j, then the within-scatter matrix for class j is given by:

$$S_j = \sum_{x_i \in X_j} (x_i - \bar{x}_j)(x_i - \bar{x}_j)^T \tag{13}$$

where $\bar{x}_j$ is the mean vector for class j:

$$\bar{x}_j = \frac{1}{n_j} \sum_{x_i \in X_j} x_i \tag{14}$$

The within-class-scatter matrix is:

$$S_w = \sum_{j=1}^{p} S_j \tag{15}$$

and the between-class-scatter matrix is:

$$S_b = \sum_{j=1}^{p} n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T \tag{16}$$

The total-scatter matrix is equal to the sum of the between-scatter matrix and the within-scatter matrix: $S_t = S_b + S_w$. The objective of the first FDA vector, w_1, is to maximize the scatter between classes while minimizing the scatter within classes:

$$\max_{w_1 \neq 0} \frac{w_1^T S_b w_1}{w_1^T S_w w_1} \tag{17}$$

with $w_1 \in \Re^m$. The second FDA vector, w_2, is computed so as to maximize the scatter between classes while minimizing the scatter within classes on all axes perpendicular to the first FDA vector, and so on for the remaining FDA vectors. These vectors are equal to the eigenvectors w_k of the generalized eigenvalue problem:

$$S_b w_k = \lambda_k S_w w_k \tag{18}$$

where the eigenvalues λ_k indicate the degree of separability between the classes. As it is the direction and not the magnitude of w_k which is important, the norm is usually chosen to be $\|w_k\| = 1$. The first FDA vector is the eigenvector associated with the largest eigenvalue and so on.

Then, the linear transformation of the data x from the m-dimensional space to the reduced space a-dimensional generated by the FDA vectors is:

$$z_i = W_a^T x_i \tag{19}$$

where $W_a \in \Re^a$ has the a FDA vectors as columns, and $z_i \in \Re^a$. FDA computes the matrix W_a that as the data $x_1, \ldots, x_n$ for the p classes are optimally separated when projected into the a-dimensional space.

There are several methods to choose the number of FDA vectors. These methods are very similar to PCA selection methods, cited in section 2. For example, cross validation or the SCREE procedure.

In order to diagnose the faults, FDA takes into account data collected during different faulty conditions, and uses a discriminant function that takes into account the similarity between the actual data and the data belonging to each class. An observation is assigned to the class i when the maximum discriminant function value, g_i, satisfies:

$$g_i(x) > g_j(x) \ \forall j \neq i \tag{20}$$

where $g_i(x)$ is the discriminant function for class i given a measured vector $x \in \Re^m$. The discriminant function that minimizes the error rate, when the event v_i occurs (for example, the fault i), is (Duda et al., 2001):

$$g_i(x) = P(v_i|x) \tag{21}$$

where $P(v_i|x)$ is the *a posteriori* probability of x belonging to class i. It can be shown that identical classification occurs when the equation (21) is replaced by:

$$g_i(x) = \ln p(x|v_i) + \ln P(v_i) \tag{22}$$

Using the Bayes' rule, considering that the data for each class are normally distributed and characterizing the data to this case, i.e., considering $W_a \in \Re^{m \times a}$ containing the eigenvectors $w_1, w_2, \ldots, w_a$ computed from equation (18), the discriminant function for each class can be derived as:

$$g_j(x) = -\frac{1}{2}(x - \bar{x}_j)^T W_a (\frac{1}{n_j - 1} W_a^T S_j W_a)^{-1} W_a^T (x - \bar{x}_j) + \ln(p_j) - \frac{1}{2}\ln[\det(\frac{1}{n_j - 1} W_a^T S_j W_a)] \quad (23)$$

where S_j, $\bar{x}_j$ and n_j are defined in equations (13) and (14) respectively.

4. 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 (Copp, -; Alex et al., 2008). This benchmark was developed for the evaluation and comparison of different activated sludge wastewater treatment control strategies. The model is implemented using MATLAB© and SIMULINK©.

Fig. 1 shows an overview of this plant. It is composed of a two-compartment activated sludge reactor consisting of two anoxic tanks followed by three aerated tanks. This type of plant combines 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 layer non-reactive unit. The 6^{th} layer is the feed layer. Table 1 shows the physical parameters of the plant.

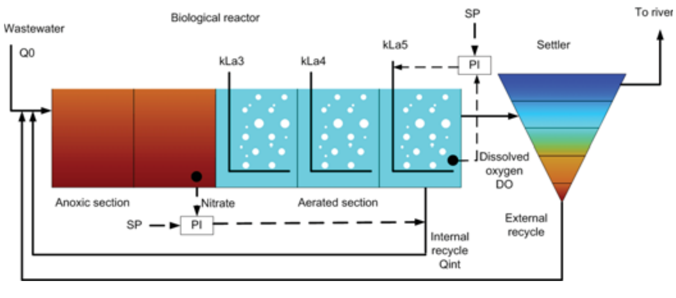

Fig. 1. General overview of the wastewater treatment plant (WWTP)

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

The influent used was the dry influent data file Copp (-). In this file, the variation of influent flow is between $15000 - 35000\ m^3/d$. The plant, as Fig. 1 shows, has two reflux: external refluxes, from settler to input, which is approximately equal to the influent flow, and internal reflux, from the last aerated tank to input, which is approximately equal to three times the influent flow, but which is a controlable variable.

The objective of the control strategy is to control the dissolved oxygen level in the aerated reactor by manipulation of the oxygen transfer coefficient (K_La_5) and to control the nitrate level in the anoxic tank by manipulation of the internal recycle flow rate. The controllers are of PI type. Tab. 2 shows the principal controller 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 variables involved 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 a reduction in the normal growth of heterotrophic organisms. This type of fault can be produced by toxic substances in 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 a reduction in the normal growth of the heterotrophic organisms and an increase in the decay factor of this type of organisms. This fault is similar to toxicity shock but is more drastic. In this case, the fault is caused by reducing the maximum heterotrophic growth rate (μ_H) and 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 the 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 Copp (-). But in this example the benchmark has been modified in order to introduce the fault parameters.

There are several groups working on fault detection in wastewater treatment plants using PCA (Rosen and Lennox, 2001) or using other fault detection approaches (Genovesi et al., 2000).

Using this dynamic model, the results were obtained 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 closed-loop is simulated for 14 days and faults are caused on the 7^{th} day. The samples for monitoring experiments were taken 100 times per day.

The selected variables to calculate principal components analysis (PCA) and Fisher discriminant analysis (FDA) 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 (Tomita et al., 2002).

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

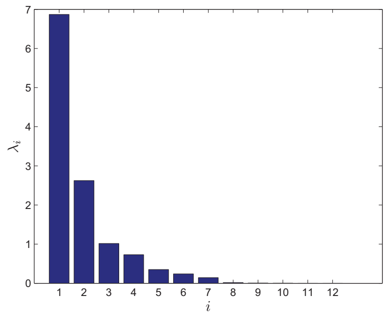

Fig. 2. SCREE graph for principal component selection

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

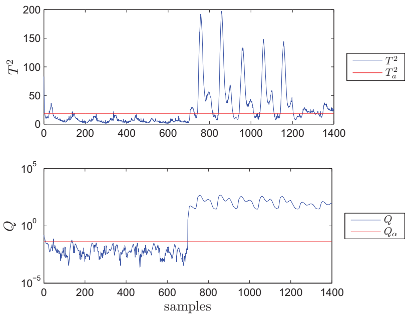

Fig. 3. Toxicity shock fault detection. Logarithmic scale for Q statistic.

The inhabitation fault detection is more effective than the detection of the toxicity shock fault because this type of fault is more drastic, as can be seen in Fig. 4. Finally, the bulking fault detection using PCA is shown in Fig. 5.

The number of selected FDA vectors for fault diagnosis tasks was two using the the SCREE graph method. Fig. 6 shows the discriminant functions (g_i, eq. 23) when a toxicity shock fault has been caused. The solid line corresponds

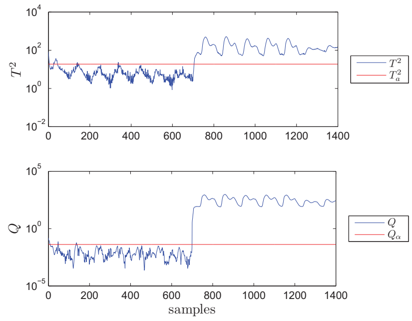

Fig. 4. Inhabitation detection. Logarithmic scale for T^2 and Q statistics.

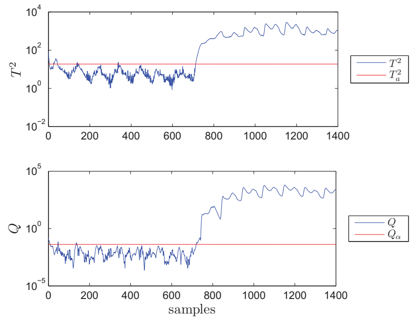

Fig. 5. Bulking fault detection. Logarithmic scale for T^2 and Q statistics.

to the discriminant function for toxicity shock fault, the dotted line corresponds to the discriminant function for inhabitation fault and the dashed line corresponds to the discriminant function for the bulking fault. In this case, once the fault has been detected (7^{th} day) the discriminant function for the toxicity shock fault is greater than the rest of the discriminant functions, so the fault is correctly diagnosed. The experimented faults used to find results are different from the considered faults used in the training data.

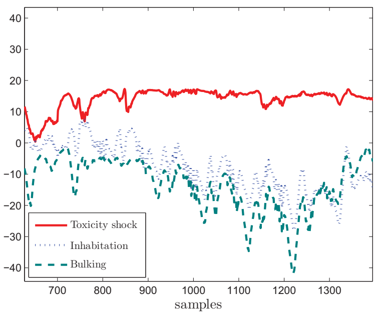

Fig. 6. Toxicity shock fault diagnosis

Fig. 7 shows the discriminant function graphs in the case where the inhabitation fault has occurred. In this situation, the discriminant function for the inhabitation fault is the greatest, so the fault is correctly diagnosed.

Finally, Fig. 8 shows the evolution of the discriminant function for the bulking fault. In this case, the evolution for the bulking fault is always greater than for the rest of the discriminant faults.

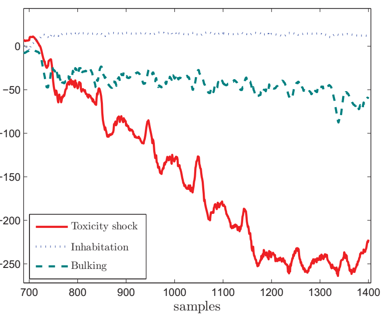

Fig. 7. Inhabitation fault diagnosis

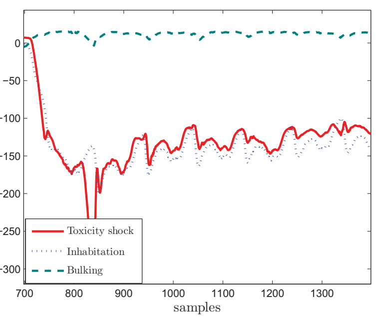

Fig. 8. Bulking fault diagnosis

5. CONCLUSIONS

This paper proposes an approach to deal with the fault detection and diagnosis using statistical techniques, concretely, the principal component analysis (PCA) is used in detection tasks and the Fisher discriminant analysis (FDA) is implemented in diagnosis tasks.

The approach has been proved in a simulated wastewater treatment plant (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 to detect the faults. Data for different classes (parameter faults) are also collected to calculate the FDA models for diagnosis. The used approach shows good results because the faults was detected and correctly diagnosed.

A useful update to this work can be to obtain data when two or more faults occur simultaneously. New discriminant functions can be calculated using this data and these new situations could be diagnosed.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the department of industrial electrical engineering and automation of Lund University for the implementation of BSM1 model.

REFERENCES

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

Chiang, L., Russell, E., and Braatz, R. (2000). *Fault Detection and Diagnosis in Industrial Systems*. Springer, Nueva York.

Copp, J. (-). *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.

Duda, R., Hart, P., and Stork, D. (2001). *Pattern Clasification*. Wiley, New York, 2nd edition.

Fuente, M., Garcia, G., and Sainz, G. (2008). Fault diagnosis in a plant using fisher discriminant analysis. *16th Mediterranean Conference on Control and Automation Congress Centre, Ajaccio, France*, 53–58.

Genovesi, A., Harmand, J., and Steyer, J. (2000). Integrated fault detection and isolation: Application to a winery's wastewater treatment plant. *Applied Intelligence*, 13, 59–76.

He, Q., Qin, S., and Wang, J. (2005). A new fault diagnosis method using fault directions in fisher discriminant analysis. *AIChE Journal*, 51(2), 555–571.

Hwang, D. and Han, C. (1999). Real-time monitoring for a process with multiple operating modes. *Control Engineering Practice*, 7, 891–902.

Jackson, J. (1991). *A user's guide to principal components.* Wiley.

Jackson, J. and Mudholkar, G. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, 21, 341–349.

Kourti, T. (2002). Multivariable dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. *Journal of Chemometrics*, 17, 93–109.

Rosen, C. and Lennox, J. (2001). Multivariate and multiscale monitoring of wastewater treatment operation. *Water research*, 35, 3402–3410.

Tien, D., Lim, K., and Jun, L. (2004). Compartive study of pca approaches in process monitoring and fault detection. *The 30th annual conference of the IEEE industrial electronics society*, 2594–2599.

Tomita, R., Park, S., and Sotomayor, O. (2002). Analysis of activated sludge process using multivariate statistical tools - a pca approach. *Chemical Engineering Journal*, 90, 283–290.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S., and Yin, K. (2003a). A review of process fault detection and diagnosis. part i: Quantitative model-based methods. *Computers & Chemical Eng.*, 27, 291–311.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S., and Yin, K. (2003b). A review of process fault detection and diagnosis. part ii: Qualitative models and search strategies. *Computers & Chemical Eng.*, 27, 313–326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S., and Yin, K. (2003c). A review of process fault detection and diagnosis. part iii: Process history based methods. *Computers & Chemical Eng.*, 27, 327–346.

Zumoffen, D. and Basualdo, M. (2007). From large chemical plant data to fault diagnosis integrated to decentralized fault tolerant control. *Industrial & Eng. Chemistry Research*, 47, 1201–1220.