2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Anomaly Detection in Power Generation Plants using Similarity-based Modeling and Multivariate Analysis

Felipe A. Tobar, Luis Yacher, Rodrigo Paredes and Marcos E. Orchard, *Member IEEE*

***Abstract*—This paper introduces an anomaly detection method based on a combination of nonparametric models of the process and multivariate analysis of residuals. This method basically intends to recognize abnormal conditions in the operation of a monitored system, considering for this purpose the definition of "baseline" operation through historical datasets. In particular, the proposed anomaly detector utilizes similarity–based modeling (SBM) techniques to represent the process behavior and principal component analysis (PCA) for the study of model residuals. The methodology not only helps to detect changes in the operation of the system, but also provides a structured algorithm for the inclusion of representative samples in the data set that is used to define the baseline of the system. The method is validated using data from a power generation plant.**

I. INTRODUCTION

THE implementation of adequate monitoring systems have become a key issue in a world where the economic impact of system reliability and cost-effective operation of critical assets is steadily increasing. In this sense, anomaly detectors are one of the first and most important steps needed to ensure operational continuity of the process, plant safety, as well as high production quality standards.

An anomaly detector [1] is a module that basically intends to recognize abnormal conditions in the operation of a monitored system. In most real applications, the anomaly detector is required to perform this task while minimizing both the probability of false alarms and the detection time (time between the initiation of a fault and its detection), given a fixed threshold. Conventional anomaly detection and fault diagnosis algorithms [1]-[2] provide a solution to the problem of monitoring a finite number of fault modes that are deemed to be severe, frequent and "testable" on the basis of a Failure Modes, Effects, and Criticality Analysis (FMECA).

Classical fault detection and identification (FDI) methods rely on an accurate model of the system under consideration and the utility of an innovation or "discrepancy" between the actual plant output and the model output, for all possible operating conditions, to detect an unanticipated fault [2][3]. The innovation (or residual) method captures the fault signature, and suggests which residuals are normal or which ones result from fault conditions. A variety of techniques have been proposed based on estimation theory, failure sensitive filters, multiple hypothesis filter detection, generalized likelihood ratio tests, model-based approach, statistical analysis, and information theory [3]-[11].

Availability of historical data is always assumed for purposes of defining an appropriate baseline. In fact, whenever parametric models are used to describe the system under analysis, this information helps to determine both the most appropriate structure for these models and to estimate adequate initial conditions for their parameters. However, if the actual system dynamics are not well understood, then verification, calibration, and validation of parametric models represent a difficult challenge. In contrast to the aforementioned problem, nonparametric models offer a direct representation of nonlinear systems based on historical data. In this manner, the definition of "baseline" operation is done only by selecting of a number of samples where the process behaved accordingly to a particular set of requirements or standards, avoiding the need of a particular structure or linear/Gaussian assumptions.

This work implements a system monitoring scheme to identify different operation conditions utilizing a nonparametric modeling approach known as *Similarity-based Modeling* (SBM). The supporting concept to implement this approach is to estimate the system output with SBM and compare it with the actual, measured, output when available; if this comparison reveals that the system is not performing according to a known, historical, database; or even if the modeling structure is not capable to replicate the system behavior, it is possible that the database being considered does not represent all of the possible operation conditions of the system under study. If this is the case, it will be necessary to complement the database with new samples in order to reduce the error associated to the system modeling, which is also a measure of how deviated is the system w.r.t a known database.

On the other hand, if the modeling structure properly replicates the system behavior, (under a given criteria) in multivariate systems it is also necessary to verify that the known relationships between variables still hold. Due to the large number of variables to estimate in most of the systems, the assessment of the system behavior cannot be performed purely considering each variable estimation error, consequently, multivariate analysis techniques such as *Principal Component Analysis* (PCA) must be employed in order to reduce the space dimension. Additionally, once the PCA has been applied, hypothesis testing resources such as the Hotelling's Test can be

Manuscript received Sept 26, 2010. This work was supported by CONTAC Ingenieros Ltda. and Endesa Chile.

F. A. Tobar is a graduate student at the Electrical and Electronic Engineering Department, Imperial College, London, SW7 2AZ, UK. (e-mail: felipe.tobar@gmail.com).

L. Yacher is the Manager of CONTAC Ingenieros Ltda., Avda. Américo Vespucio Sur 315, Ñuñoa, Santiago, Chile (corresponding author; Phone: 56-2-2700900; e-mail: lyacher@contac.cl).

Rodrigo Paredes is with Endesa Chile, Avda. Santa Rosa 77, Santiago, Chile (Phone: 56-2- 6309089).

M. E. Orchard is with the Department of Electrical Engineering, Universidad de Chile, Santiago, Chile (e-mail: morchard@ing.uchile.cl).

978-1-4577-0079-8/11/$26.00 ©2011 AACC

considered to ensure that the modeling errors remain in a statistically acceptable region.

An extension of the proposed methodology is to determine faulty conditions, since in case that the system could not be modeled with the extended database, the system could be operating under an abnormal condition. In this regard there must be noted that, in case of implementing a fault detection scheme, the addition of new samples to the database must be done with special attention of not incorporating samples corresponding to these abnormal conditions, since if this is done, the SBM algorithm will consider faulty conditions as known, and hence, normal. On the order hand, if the objective is to replicate system behavior given any operation condition, all of the samples should be considered. As a final remark, it is important to note that the proposed anomaly detection scheme also provides a structured algorithm for the inclusion of representative samples in the data set that is used to define the baseline of the system, a critical matter in complex time-varying/nonlinear systems.

This paper is organized as follows, Section II presents the necessary theoretical resources for the implementation of our system monitoring scheme this is the fundamentals of SBM, principal component analysis, and the Hotelling's test. Section III explains the considerations regarding the data preprocessing, the justification for the implementation of the proposed techniques, and the results of the estimation routine with a two different databases associated to power generation plants. Finally, Section IV states the concluding remarks of the works and suggests the guidelines of future research work in this field.

II. Theoretical Background

The proposed anomaly detector uses a combination of several well known methods to achieve its objective. This section focuses on presenting the theoretical framework for each one of these tools.

A. *Similarity-based Modeling for System Monitoring*

One advantage of the nonparametric modeling techniques is that they do not require an *a priori* knowledge of the system, since its implementation is based on the identification of similarities and relationships between a given data set and online observations, instead of the construction of algebraic structures based upon these observed data. A particular case of such structures is the Similarity-based Model (SBM), which estimates the system output by comparing online measurements and a historical data base which represents the system under study. SBM has proven to be a successful estimator when used in high dimension systems using considerably low number of training samples [13].

In order to understand the SBM basic concept for systems modeling, consider the static system defined by (1):

$$y = f(x), x \in R^m, y \in R^p \qquad (1)$$

where x and y are the system input and output respectively, and $f(.)$ is an unknown function.

When input and output measurements are available for the system in (1), it is possible to define the following training matrices (input and output matrices respectively) according to (2):

$$D_i = [x_1 x_2 \dots x_n] \in R^{m\times n}, \qquad (2.a)$$
$$D_o = [y_1 y_2 \dots y_n] \in R^{p\times n}, \qquad (2.b)$$

where $y_i = f(x_i), \forall i = 1..n$, and the pairs $[x_i, y_i]_{i=1..n}$ accurately represent the system behavior; i.e., they span the regions containing the system operations points.

Hence, SBM assumes that for a given an input x^*, it is possible to estimate $y^* = f(x^*)$ by a linear combination of the columns of D_o denoted by $\hat{y}^*$. Consequently, the problem of estimating $y^* = f(x^*)$ can be regarded as the determination of a vector $w \in R_+^n$ such that $\hat{y}^* = D_o w$.

This vector can be found as done in (3).

$$w = \frac{\widehat{w}}{\vec{1}^T \cdot \widehat{w}}, \qquad (3.a)$$
$$\widehat{w} = (D_i^T \Delta D_i)^{-1}(D_i^T \Delta x^*), \qquad (3.b)$$

where Δ is a *similarity* operator [13][14].

SBM is not restricted to any particular similarity operator; however, according to the literature it must have certain features. For two elements $A, B \in R^u$, $A\Delta B \in R^+$ must be symmetric, reach its maximum in $A = B$, and monotonically decay with $\|A - B\|$.

The literature does not provide a framework for choosing a suitable similarity operator based on the available measurements. In this regard, the designer has to consider the performance of several operators in order to implement an SBM scheme that truly captures the data variations and similarities. In this work, the operators that were considered are based on exponential functions, saturated linear operators, or the Epanechnikov kernel.

As presented above, SBM is a nonparametric modeling technique that allows the estimation of static systems, however, the vast majority of the industrial systems are dynamic ones, and hence, they go beyond of the scope of the SBM algorithm if the implementation limits to what it has been presented so far. This problem can be surmounted from two different standpoints provided that online measurements are available; the first one involves the incorporation of past observations (both inputs and outputs) as regressors to estimate the system response, by doing so not only the static input-output relationships are captured, but also the variables sequential dependence. The second approach is to merely neglect the dynamic properties of the system, regarding it as a static one. This concept can only be applied when the data is acquired at a very high frequency w.r.t. the system variations, such as thermo-dynamical or mechanical systems.

Given that a nonparametric modeling structure for dynamic systems is available (SBM), as well as a historical database representing some of the operation conditions, it is possible to evaluate whether the system behavior is within these known operation conditions or not. This can be performed by contrasting the online measurements, and the SBM estimates, specifically, as SBM emulates the plant behavior in normal operation conditions, if the estimates differ considerably from the actual measurements (w.r.t. a given criteria), it could be inferred that the system is not operating in a known operating point, and consequently the database must be extended with samples representing the unknown condition.

After the process of incorporating samples to the database is complete, i.e. once for every input x^* the estimation error given by

$$e = y^* - \hat{y}^* = f(x^*) - D_o \frac{(D_i^T \Delta D_i)^{-1}(D_i^T \Delta x^*)}{\vec{1}^T \cdot (D_i^T \Delta D_i)^{-1}(D_i^T \Delta x^*)} \in R^p$$

is acceptable under a specified criteria, the relationships between the measured variables should be assessed to ensure consistency with the operation conditions represented in the database. Due to the large number of variables that are present in industrial systems, multivariate processing algorithms should be implemented to verify these relationships. In the following, a broadly used example of these algorithms is presented.

B. Principal Component Analysis

Principal Component Analysis (PCA) is a dimensionality reduction technique for correlated variables, i.e. for a given a set of correlated variables, it aims to find a set of uncorrelated variables of smaller dimension. PCA performs a linear transformation of the data, which is optimal in terms of capturing its variability, and determines a new data set ordered by the level of representation of the entire process variance.

Theoretically, for the following data matrix:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix},$$

which comprises n observations for each one of the m variables, PCA finds a *loading matrix* $P \in R^{m \times a}, a \leq n$, which relates X to the first a principal components being contained in the *score matrix*

$$T = XP.$$

Denoting the i^{th} column of T by t_i, the transformation performed by PCA holds [15]:

1. $\mathrm{Var}(t_1) \geq \mathrm{Var}(t_2) \geq \cdots \geq \mathrm{Var}(t_a)$.
2. $\mathrm{Mean}(t_i) = 0, \forall i$.
3. $t_i^T t_k = 0, \forall i \neq k$.
4. There is no other transformation of a components that captures more variations of the data.

Additionally, the projection back in m-dimensional space is given by [16]:

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

and hence, the difference between X and $\hat{X}$ is the residual matrix E:

$$E = X - \hat{X},$$

which captures the variations of space generated by the remaining $m - a$ components, and has theoretically low signal-to-noise ratio. It has been formally justified [17] that, when a is properly chosen, these remaining components represent the random noise of the measurements, whereas the first a components describe dynamic variations.

The application of PCA in our system monitoring framework is to reduce the dimension of the error e (which is p) in order to analyze a smaller number of variables when identifying that the known relationships between the measured variables do hold. Once the PCA linear transformation has been performed, one resource to recognize if the system is performing as expected is to run a hypothesis test on the deviation of the error principal components.

C. Hotelling's Test

This test characterizes the variability of multivariate data by a scalar threshold, which is associated to a given level of significance [18]. In our case this allows to evaluate, for a given level of significance, if the estimation error (through its principal components) relies within an acceptable region for a selected subset of variables of interest.

To properly introduce the Hotelling's test, consider the sample covariance of the data matrix X given by

$$S = \frac{1}{n-1} X^T X.$$

The Hotelling's test states that a particular observation $x \in R^m$, could be represented by the training set, if the statistic

$$T^2 = x^T S^{-1} x,$$

is below the threshold

$$T_\alpha^2 = \frac{m(n-1)(n+1)}{n(n-m)} F_\alpha(m, n-m),$$

where $F_\alpha(g, k)$ is the α level of confidence of the F-distribution with g and k degrees of freedom.

III. Anomaly Detection in Power Generation Plant Using Similarity-Based Modeling and PCA

The proposed monitoring scheme was implemented in a Chilean natural gas power generation plant, which measurements were composed by 14000 observations for each one of the variables such as pressures, temperatures, valves positions, voltages, speed of rotating parts, and Boolean states. The data, acquired using OSIsoft PI system, were grouped in the input matrix $X \in R^{14000 \times 42}$, and the output matrix $Y \in R^{14000 \times 53}$; being 42 and 53 the number of input and output variables respectively. The i^{th} rows of the matrices X and Y were respectively denoted by $x_i \in R^{42}$ and $y_i \in R^{53}$, and the matrix containing all the measurements was denoted by the concatenated matrix $M = [X\ Y] \in R^{14000 \times 95}$. Also, these data were processed sequentially in order to emulate online observations, where a normalization step was included in order to avoid biased results due to the different variables magnitude. All the numerical implementations of this work were performed in MATLAB.

In order to implement a nonparametric (SBM) monitoring scheme, the literature states that a data base, accurately representing the different operation conditions, must be available. Unfortunately, in our case there is no database that represents the system behavior, nor information about the nature of the available data, i.e. within the matrices X and Y, "normal"—as well as "faulty"—operation conditions can be found. To overcome this problem, an initial training subset within the available data was considered, and in base of the estimation error this database was extended.

A. Preprocessing

The first test to be done in order to justify the use of SBM techniques, is to verify if the available data presents different operation regions. In order to do so , the first 4 principal components of the variables in M, which capture the 87% of the data variation, are presented in Fig. 1 to identify the existence of clustered operation regions. From this figure well defined operation points are identified, this evidence suggests the use of a monitoring technique based on SBM, since it relies on finding the current operation condition of the system to estimate its output. It must be noted that this approach only reveals the existence of different operation conditions, but it does not provide any notion of these conditions' nature (normal or faulty).

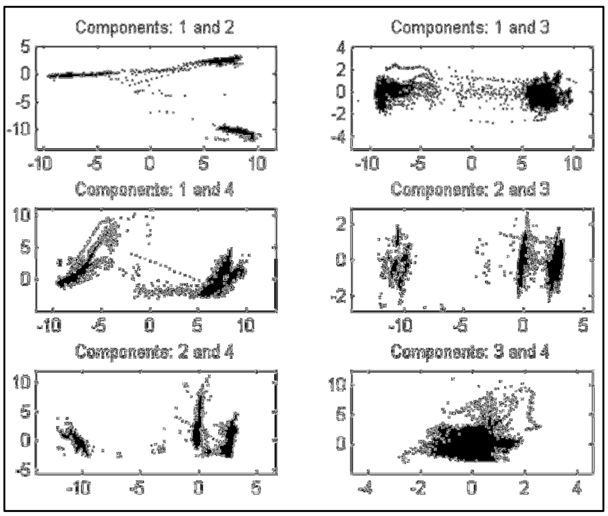

Fig. 1. Principal component analysis (PCA) of data from power generation plant. Clusters are the first indication of the existence of several operating points within the data set.

An important feature to discuss at this stage is the dynamic consideration of the system, and hence of the modeling structure. The system under study incorporates thermodynamical and mechanical subsystems, and more importantly, it is mainly used under a feedback control loop that ensures tracking of constant set points. These facts suggest the consideration of a static model such as the SBM presented in Section II.B, this assumption will indeed lead to estimation errors in case of set point changes, however, this time periods are not important when compared to the phenomena of interest (sustained faulty conditions).

Being stated that the data admits the use of SBM techniques, and assuming that the system dynamics can be neglected, a suitable similarity operator should be defined w.r.t to the statistical properties of the measurements. After a preliminary study, the similarity operator that best captured the data variability was the *saturated triangular operator* defined in (4).

$$A\Delta B = \begin{cases} d - \|A-B\| & \|A-B\| \leq d+\varepsilon \\ \varepsilon & \|A-B\| > d+\varepsilon \end{cases} \quad (4)$$

where $\varepsilon > 0$ is a small number that ensures $A\Delta B > 0$, and $d > 0$ is a threshold depending on the observations variance.

With these preliminary considerations, it is now possible to implement the SBM monitoring scheme, and according to its estimation errors, extend the database.

B. Database Extension

As stated above, the training matrices to be used as the initial database will be chosen within the available data, which were acquired using OSIsoft PI system. Data is comprised of 100 input-output pairs, evenly-spaced, within the first 4000 samples contained in the matrices X and Y, and it will be denoted by DB_0. In order to avoid biased—and high variance—estimates due to the different magnitudes in the measured data, the mean and variances of such variables have been estimated (by the sample mean and variance calculated using the training set) and used to normalize the data.

The Mean Squared Error (MSE) related to the normalized output and its SBM estimate is presented in Fig. 2. From this figure, it can be seen that using the specified database, the MSE related to the SBM estimation remains considerably low (provided that the signals to estimate are zero-mean, unit-variance) for the region that contain the training set (the first 4000 samples) and even for some regions that does not contain any training sample. This reveals that the training set constructed with samples taken in the first 4000 time instants is also representative of the system behavior in other periods (specially [10000 12000]) when using the SBM algorithm.

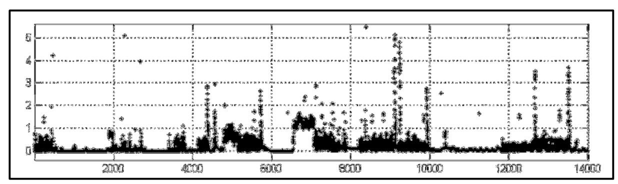

Fig. 2. Squared error associated to SBM model of power generation plant based on DB_0 (MSE:Error = 0.22341)

From Fig. 2 can also be noted that most of the high MSE estimates occur in reduced time periods, and hence they are of no interest in this analysis since they could be a consequence of neglecting the system dynamics, or merely sensor errors. However, there is a region that presents a sustained MSE, which could be produced by two possible scenarios, i) the system input is not represented in the training set, and hence the SBM algorithm is not capable of representing the output, or ii) the system input is recognized by the training set, but the system is actually behaving differently. To find out what scenario the system is, Fig. 3 shows the maximum—unnormalized—SBM weights series $\hat{w}$, which is an indicator of the successful input recognition by the current training database. This figure reveals that the area around time instant 7000, the system inputs are not found to be similar to any sample in the database (the first case mentioned above). Hence, to make sure that this is not a faulty condition, samples accurately representing this period should be incorporated into the database in order to derive a consistent modeling structure. Note that the figure also shows the maximum weights corresponding to the training set, which are all equal to 1, suggesting that the training samples are properly identified in the testing stage.

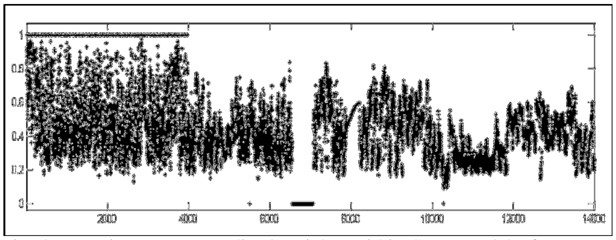

Fig. 3. Maximum unnormalized weights within SBM model of power generation plant as a representative measure of the identification of online measurements.

The extended database DB_1 contains the samples in DB_0 plus two samples in the mentioned area, specifically

$$DB_1 = DB_0 \cup \{(x_{6600}, y_{6600}), (x_{6800}, y_{6800})\}.$$

From Fig. 4 can be seen that the MSE associated to DB_1 does not present the sustained error region due to unknown system inputs. This reveals that the proposed algorithm is capable of identifying the presence of unknown operation points, and incorporating them in order to perform a successful modeling of the system.

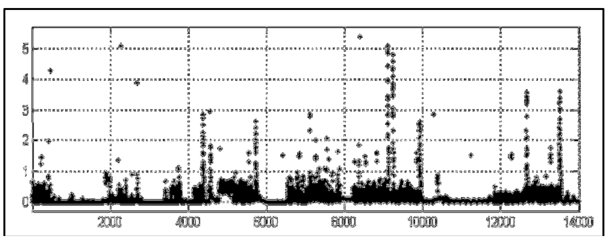

Fig. 4. Squared error associated to SBM model of power generation plant based on DB_1 (MSE-Error = 0.17032)

C. Relationship between measured variables

To verify that, in addition to minimizing the estimation MSE, the proposed algorithm also captures the relationship between the system variables, the initial database has been extended to minimize the MSE throughout the period [1 14000]. The new database DB comprises 288 samples (DB_0 plus 188 samples) manually chosen based on the MSE and the unnormalized weights just as explained in Section III.B; note that using a training set of 288 samples to estimate a sequence of 14000 observations means that about a 2% of the samples are being used to train the algorithm. The resulting MSE of this procedure (presented in Fig 5) presents lower values when compared to the MSE signals corresponding to DB_0 and DB_1.

Additionally, the first 4 error principal components, which represent the 92% of the error variation, are plotted in Fig. 6, where the Hotelling's test has been run for every pair of components to find the 95% confidence ellipse, using the software SCAN developed by CONTAC Engineers Ltda. This figure shows the portion of the estimates error that is out of the confidence ellipse, and also suggests a possible dependency between the error principal components; as a matter of fact, the cross-shaped plots reveal the presence of—at least—as many operation conditions as plot axes for each pair of components.

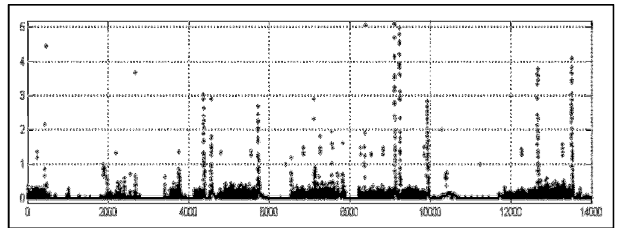

Fig. 5. Squared error associated to SBM model of power generation plant on testing data using DB (MSE-Error = 0.11641)

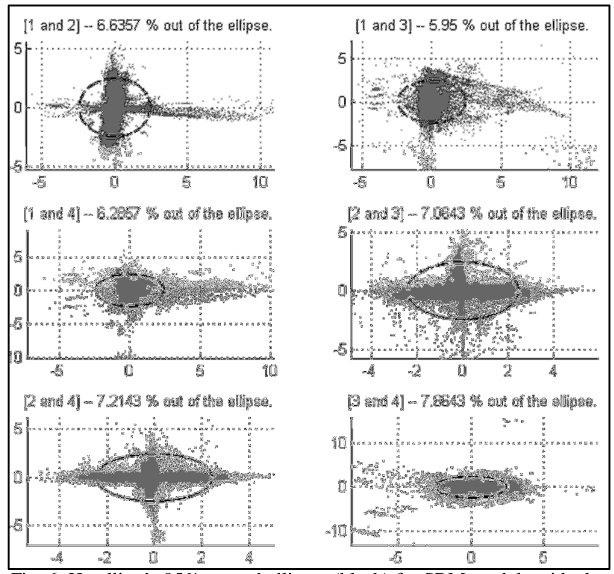

Fig. 6. Hotelling's 95% control ellipses (black) for SBM model residuals (grey) using the software SCAN (developed by CONTAC Engineers Ltda.). The corresponding components (square brackets) and the portion of the samples that rely out of the ellipse can be found above each plot.

As the overall estimation MSE (Fig. 5) and the Hotelling's test (Fig. 6) show that the system estimation is acceptable when appropriate samples are added to the training set as done in III.B, the output SBM estimates are presented as follows. Fig. 7 presents a comparison between three important normalized output signals and their SBM estimates, from top to bottom these variables correspond to a device temperature which is continuous and present fluctuations, an operation condition which is a Boolean, and a gas temperature which is continuous but it presents minimum variations around its values. Despite the different nature of these variables, the SBM algorithm delivers acceptable estimates, and it can be noted that the estimation delay is minimum. Note that for the estimate of Boolean states within the system, a saturator should be used.

IV. Conclusions

This paper presents and tests a scheme to detect anomalies in a power generation plant by comparing the online system outputs, and their estimates computed by a nonparametric modeling algorithm known as Similarity-based Model. The proposed scheme also allows the designer to include new samples in the training set in order to improve the estimation routine. An assessment of the estimation error, and the relationship between systems

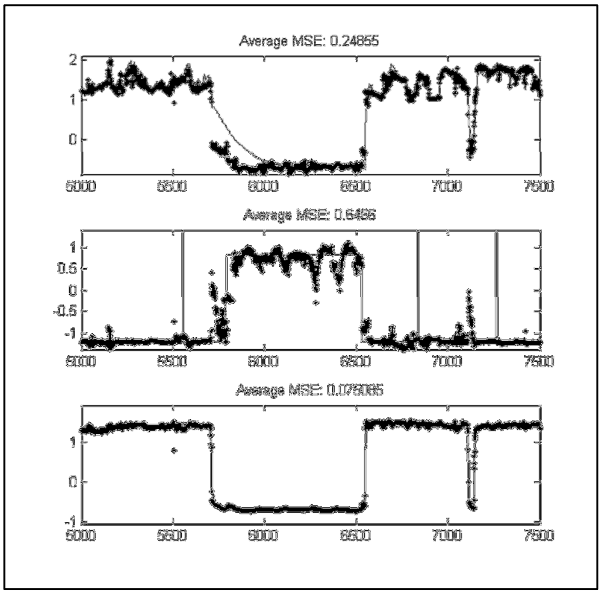

Fig. 7. Normalized output process (grey line) variables and their respective SBM representation (black dots) in a power generation plant. From top to bottom: device temperature, Boolean state, and gas temperature.

variables is also provided; this is done by means of Principal Component Analysis, and the Hotelling test. Once a representative training set is constructed, the proposed scheme estimate the system output with low squared errors and it also captures the relationships between the variables. Another key feature of this work is that the data set used to estimate the system output is considerable small when compared with the estimation horizon (228 samples to estimate 14000 observations).

REFERENCES

[1] M. Orchard and G. Vachtsevanos, "A particle filtering-based framework for real-time fault diagnosis and failure prognosis in a turbine engine," Mediterranean Conference on Control and Automation MED'07, Athens, Greece, July 2007.

[2] R. Isermann, "Supervision, fault detection and fault diagnosis methods – an introduction," *Journal of Control Engineering Practice*, 5(5):639–652, May 1997.

[3] R. Isermann and P. Balle, Trends in the Application of Model-Based Fault Detection and Diagnosis of Technical Processes, *Control Eng. Practice*, vol. 5, no. 5, 1997, pp. 709-719.

[4] B. Ayhan, M.-Y. Chow, M.-H. Song, Multiple discriminant analysis and neural network based monolith and partition fault detection schemes for broken rotor bar in induction motors, *IEEE T. Industrial Electronics*, Vol. 53, No. 4, pp 1298-1308, June, 2006.

[5] B. Ayhan, H. Trussell, M.-Y. Chow, M.-H. Song, On the use of a lower sampling rate for broken rotor bar detection with DTFT and AR-based spectrum methods, *IEEE T. Industrial Electronics*, Vol. 55, No. 3, pp 1421-1434, Mar. 2008.

[6] M. Khan, M. Rahman, Development and implementation of a novel fault diagnostic and protection technique for IPM motor drivers, *IEEE T. Industrial Electronics*, Vol. 56, No. 1, pp 85-92, Jan. 2009.

[7] A. Lebaroud, G. Cleac, Classification of induction machine faults by optimal time-frequency representations, *IEEE T. Industrial Electronics*, Vol. 55, No. 12, pp 4290-4298, Dec. 2008.

[8] D. Tolani, M. Yasar, S. Chin, A. Ray, "Anomaly Detection for Health Management of Aircraft Gas Turbine Engines," 2005 American Control Conference, pp. 459-464, June 2005, Portland, OR, USA.

[9] W. Zhou, T. Habetler, R. Harley, Bearing fault detection via stator current noise cancellation and statistical control, IEEE T. Industrial Electronics, Vol. 55, No. 12, pp 4260-4269, Dec. 2008.

[10] A. Ibrahim, M. El Badaoui, F. Guillet, F. Bonnardot, A new bearing fault detection method in induction machine based on instantaneous power factor, IEEE T. Industrial Electronics, Vol. 55, No. 12, pp 4252-4259, Dec. 2008

[11] N. Gebraeel and J. Pan, "Prognostic Degradation Models for Computing and Updating Residual Life Distributions in a Time-Varying Environment", IEEE T. Reliability, vol. 57, no.4, pp 539-550, 2008.

[12] S. Wegerich , "Condition Based Monitoring using Nonparametric Similarity Based Modeling". Technology Development, SmartSignal Corporation. 2008.

[13] L. Gong, and D. Schonfeld, "Space kernel Analysis*",* IEEE International Conference on Acoustics, Speech, and Signal Processing 2009 (ICASSP'09*),* Taipei, Taiwan, 2009.

[14] M. J. Pivoso and K.A. Kosanovich. Applications of multivariate statistical methods to process monitoring and controller design, *Int. J. of Control*, 59:743-765,1994.

[15] L. H. Chiang, E.L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems.* Springer-Verlag London Limited 2001.

[16] B. M. Wise, N. L. Ricker, D. J. Velkamp, and B. R. Kowalski. A theoretical basis for the use of principal component models for monitoring multivariate process. *Processes Control and Quality*, 1:41-51, 1990.

[17] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, 1983.

[18] G.D. Gonzalez, M. Orchard, J.L. Cerda, A. Casali and G. Vallebuona, "Local models for soft-sensors in a rougher flotation bank," *Minerals Engineering*, vol. 16, no.5, pp. 441-453, 2003.

[19] N. D. Tracy, J. C. Young, and R. L. Mason. Multivariate control charts for individual observations. *J. of Quality Control*, 24:88-95, 1992.