CHEMICAL ENGINEERING TRANSACTIONS
VOL. 33, 2013
Guest Editors: Enrico Zio, Piero Baraldi

ISBN 978-88-95608-24-2; **ISSN** 1974-9791

A publication of
AIDIC
The Italian Association
of Chemical Engineering
Online at: www.aidic.it/cet

DOI: 10.3303/CET1333051

Application of Data Driven Methods for Condition Monitoring Maintenance

Isabel Marton*[a] , Ana I. Sánchez[b],Sofia Carlos[a], Sebastián Martorell[a]

[a]Department of Chemical and Nuclear Engineering, Universitat Politècnica de València, Camino de Vera, s/n, 46022, València,Spain
[b]Department of Statistics and Operational Research, Universitat Politècnica de València, , Camino de Vera, s/n, 46022, València,Spain
ismarllu@upvnet.upv.es

Nowadays, there is an increasing demand for Condition Based Maintenance (CBM) activities as time-directed maintenance are observed to be inefficient in many situations. CBM is a maintenance strategy based on collecting information concerning the working condition of equipment, such as vibration intensity, temperature, pressure, etc., related to the system degradation or status in order to prevent its failure and to determine the optimal maintenance. Prognosis is an important part of CBM. Different methodologies can be used to perform prognosis and can be classified as: model-based or data-driven. Model-based methods use physical models of the process or statistical estimation methods based on state observers, to this approach belong Kalman filters, particle filters, etc. On the other hand, data-driven methods only makes use of the available monitoring data which to train a learning algorithm.
In this paper a data-driven approach is presented to detect abnormal behaviours in industrial equipment. The suggested approach combines two multivariate analysis techniques: principal component analysis (PCA) and partial least squares (PLS). With PCA the most important contributors to characterize the condition of the equipment are found. Next, PLS is used to predict the system state and detect abnormal behaviour. This behaviour can lead to perform maintenance tasks. Finally, an example of application to an asynchronous generator is presented.

1. Introduction

In recent decades, a change in the maintenance planning of industrial plants has been observed due to the development in information and communication technology. This planning is of great interest as the plant availability and economy depend on the maintenance activities developed. The total cost of maintenance, which includes human and material costs, needs to be reduced to be more competitive and guarantee the production. For this reason, in the last few decades, the application of techniques of condition monitoring in order to predict the status patterns or faults of the facilities (Kusiak and Verma, 2011) has increased. In addition, an early prediction of abnormal behaviour of the equipment helps maintenance managers as it indicates equipment degradation.
Condition based maintenance (CBM) techniques guide the decision making process to select the most appropriated maintenance plan. These techniques can be classified into diagnosis and prognosis techniques (Jardine et al., 2006). The first ones are based on detection, isolation and identification of faults and abnormal behaviour once they are detected. The second ones are focused on predicting faults or failures before they occur, so it is possible to avoid failures increasing the equipment availability, and production.
There are some typical approaches for fault prognostic based on the theory, methods and routes adopted in practical applications. Such approaches are generally fallen into two main categories, namely model-based approaches and data-based approaches.

Please cite this article as: Marton I., Sanchez A., Carlos S., Martorell S., 2013, Application of data driven methods for condition monitoring maintenance, Chemical Engineering Transactions, 33, 301-306 DOI: 10.3303/CET1333051

The model-based approach relies on the use of a mathematical models (set of algebraic or differential equations) to represent the system behavior, including the degradation phenomenon. The mathematical model comprises statistical and physical models. Statistical models are developed from the collected input/output data, which do not include not recorded conditions. And physical models are useful in accounting for all operating situations (Hao et al., 2009).

On the other hand, data-driven approaches use real data obtained from data acquisition system and track features revealing the components degradation and to forecast the global behaviour of a system.

Data-driven approaches can be divided into two categories: artificial intelligence (AI) techniques (neural networks, fuzzy systems, decision trees, etc.), and statistical techniques (multivariate statistical methods, linear and quadratic discriminators, partial least squares, etc.) (Dragomir et al., 2009; Lam et al., 2010).

In this work a data driven approach methodology which combines two techniques of multivariate analysis, principal component analysis (PCA) and partial least squares (PLS) to predict the state of the equipment is proposed and applied to detect the abnormal behaviour of an asynchronous generator. The paper is organized as follows. In section 2 PCA and PLS techniques are described. Section 3 presents the methodology proposed. Section 4 presents the case of application, and finally in section 5 the concluding remark are discussed.

2. Data driven methods. PCA and PLS.

Principal component analysis (PCA) is one of the most popular multivariate statistical techniques. PCA is a powerful tool which can be used in fault detection and isolation tasks. This technique is based on a linear transformation that produces new uncorrelated variables, named components, from the original correlated measured variables. This transformation implies a dimensionality reduction of the original data so, a few of these components are sufficient to adequately represent the hidden sources of variability in the process (Wold et al., 1987).

Mathematically, PCA is based on an orthogonal decomposition of the covariance matrix of the process variables along directions that explain the maximum variation of the data matrix, X. The data matrix, X, contains information of data obtained from the equipment historical data. The N rows in the matrix X are known as objects and the K columns are called variables which are measured for each object.

The projection of X down on A-dimensional subspace by means of the projection matrix, P', gives the object coordinates in the plane, T. The columns in T, t_a, are called score vectors and the rows in P', p_a, are called loading vectors. The p_a and t_a are orthogonal. The deviations between projections and the original values coordinates are the residuals, E. In matrix form PCA is expressed as:

$$X = TP' + E \tag{1}$$

The main purpose of PCA is to find factors that have a much lower dimension than the original data set which can properly describe the major trends in the original data set.

PCA can be extended when it is desirable to include all data available in the monitoring procedure and to use the historical data, X, to predict and to detect changes in the output variables Y. This extension is known as Partial Least Squared (PLS). The PLS model is used to obtain the relationship between the matrix data, X, and the variables, Y, reducing their dimension simultaneously.

PLS model is built on the properties of the NIPALS algorithm (Geladi and Kowalsky, 1986). As shown in Equation 1, the data matrix, X, can be represented by the score matrix, P. PLS model consists of the regression between the scores for the X and Y. As relation for block X is obtained as shown in the Equation 1, only the relationship for block Y would be considered, as given by:

$$Y = UC'' + F \tag{2}$$

Where U is the scores matrix of Y, C is the loading matrix of summarizes Y, and F is the residual matrix.

The scores matrix of Y, U, has an internal relationship with the score matrix of X, T, which can be represented as:

$$U = TB + R \tag{3}$$

where B is the matrix of regression coefficients and R is the residual matrix. Therefore the PLS model can be calculated using the following expression:

$$Y = TC' + G \tag{4}$$

being G the total residual matrix.

In some cases PCA and PLS are used together in Multivariate Statistical Process Control (SPC-Multivariate) (MacGregor and Kourti, 1995). The methodology is structured in several steps. In the first step an "in-control" model based on the historical data during the normal operation is established. The second step performs the projection of new observations onto the plan defined by loadings, the score and the residuals are obtained. Finally, the control chart is plotted for all principal components. Multivariate control charts using Hottelling´s T^2 can be plotted based on each principal component.

$$T_A^2 = \sum_{i=1}^{A} \frac{t_i^2}{s_{t_i}^2} \tag{5}$$

where, $s_{t_i}^2$ is the estimated variance of t_i and A is total principal components (PC).

Therefore T^2 based on the PCA provides a test for deviations in the variables X that are of greatest importance to the variance of Y.
The squared prediction error (SPEy) of the residuals is also used in the monitoring process, as T^2 monitoring only detects the variation in the variables Y of the PC´s explained by the common cause but no if a new type of special event , not presented in the reference data used to develop the PCA in-control model occurs. The expression of SPEy is the following one:

$$SPE_y = \sum_{i=1}^{N} \left(y_{new,i} - \hat{y}_{new,i} \right)^2 \tag{6}$$

Where $\hat{y}_{new,i}$ is computed from the reference PLS and PCA model and N is the number of observations.

Based on the SPE, the absolute distance to model is defined (DModY) as:

$$DModY = \sqrt{\frac{SPE_y}{K - A - 1}} \tag{7}$$

where K is the number of variables.

3. Methodology approach

The objective of this paper is focused on development of a methodology to detect abnormal behavior of the equipment. Figure 1 shows the framework methodology used in this paper.
The first stage of the methodology is the data collection. The data used in this paper are collected from the supervisory control and data acquisition (SCADA).

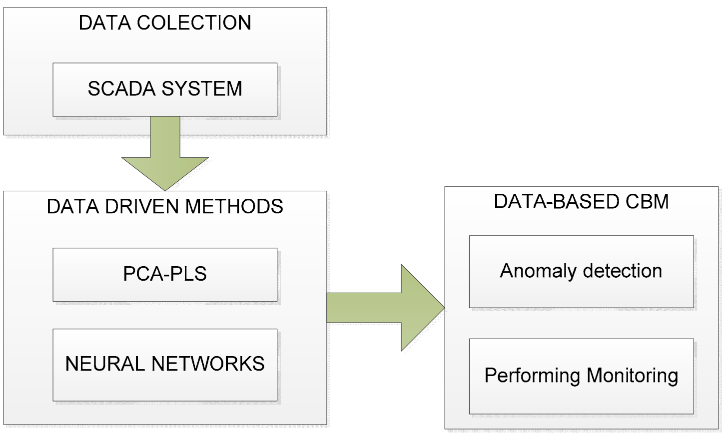

Figure 1: Framework methodology

Based on the equipment data collected, it is possible to obtain a model of the equipment normal behaviour using PCA and PLS.

Sometimes, the data collected from SCADA contain noise due to sensor errors and malfunctions, for this reason once the prediction model of each data driven is built, it is necessary to make a data filtering and remove the outliers and the influential data.
The predictive performance of the models is measured by further examination of the residual. This quantitative examination was initially conducted using the Mean Absolute Percentage Error (MAPE). MAPE is obtained from the following expression:

$$MAPE = \frac{\sum_{i=1}^{n} \left| (y_t - \hat{y}_t) / y_t \right|}{n} \cdot 100 \tag{8}$$

where, y_t is the actual value, $\hat{y}_t$ is the fitted value and n the number of observations.

Finally, the evolution of real and predicted data obtained from the models developed in step 2, will be revised in order to detect the abnormal behavior and to perform the condition monitoring of the equipment. The monitoring approach uses the predicted output to generate an alarm signal when the equipment behavior is abnormal.

4. Application case

The application case is focused on obtaining a model characterizing the normal behaviour of an asynchronous generator to detect deviations or malfunctions. In order to obtain this model, hourly measured variables from SCADA system are collected by sensors and saved on historical data files.
The data used to predict the abnormal state of an asynchronous generator has been collected from the historical information available. The historical data stored which allow generator behaviour monitoring are the following:

- Rotor speed (RS)
- Gearbox temperature (GT)
- Generator power (GP)
- Difference between bearings temperature of generator and ambient (DT_1)
- Difference between rings temperature of generator and nacelle (DT_2)

These data have been used to establish a relation between the measured operation variables and the normal function of generator in order to predict an abnormal behaviour. Data from historical information of period from 2008 to 2010 have been used. Table 1 shows the input and output variables used to build the three models considered. MAPE has been used to evaluate the goodness of prediction model fit.

Table 1: Selected normal models

Model	Input 1	Input 2	Input 3	Output	MAPE (%)
1	GP	RS		DT_1	14,84
2	GP	RS	GT	DT_1	14,78
3	GP	RS		DT_2	17,84

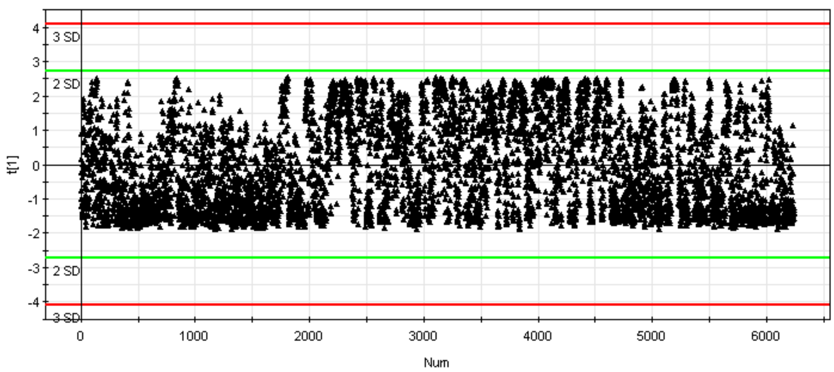

Figure 2: T2 Chart for monitoring of Model 1"in-control"

Usually, one model is better than another when its value of MAPE is lower but as shown in Table 1, Model 1 and 2 have a similar MAPE. Due to this fact we could both, model 1 or 2, and the prediction errors will be insignificant, so in this case the simplest model, model 1, has been used. Most of the data information is explained by the first principal component (PC1). Based on the PC1, Figure 2 shows the T^2 associated to the normal behaviour (in-control) of the equipment analyzed.

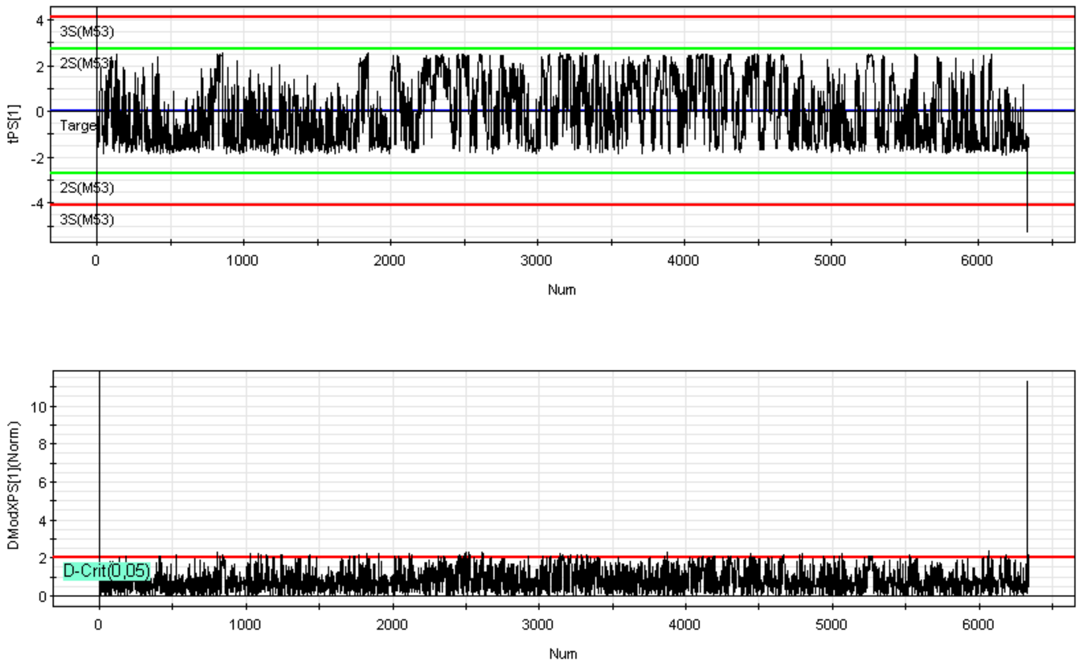

Figure 3: T^2 and DModX Chart for monitoring of Model 1"out-of- control"

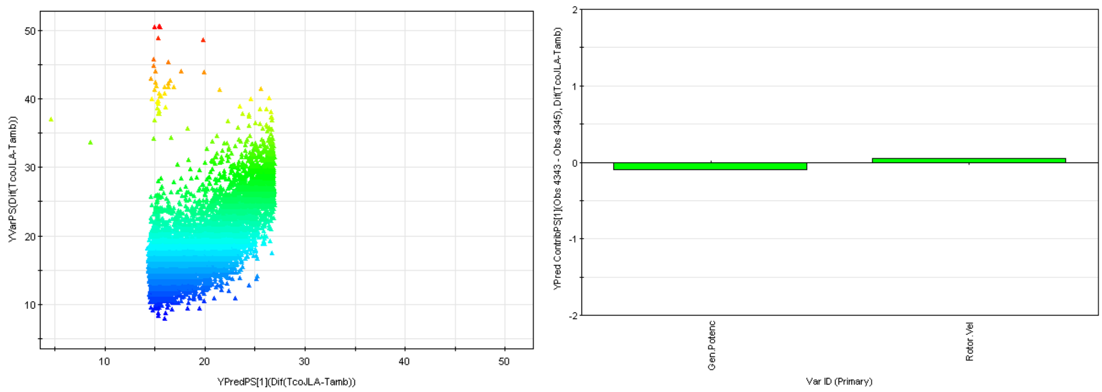

Figure 4.a) DT1 OBSERVED vs. DT1 Predicted, b) Variables Contribution plot

Figure 3 shows the results obtained from the monitoring using the model obtained to the PCA analysis,. Specifically, Figure 3 represents the contribution of the input variables to the T^2 and DModX when a new contribution is introduced. As it is observed in Figure 3 the abnormality can be clearly detected as T^2 and DModX remain beyond the control limit, what means normal behavior, until the 6350th sampling when an abnormal situation occurs. It is possible to detect and diagnose the origin of this behaviour. In this case, it is due to an increase in temperature trend, which could indicate degradation of the generator.

One possible way to diagnose this behaviour is shown in Figure 4. Figure 4 a) shows the representation of DT_1 observed versus predicted, coloured by the DT_1 observed, and b) shows the variables contribution of variables (GP and RS) for the points with higher DT1 observed (points highlighted in red). As shown in this figure, the observations with an abnormal behaviour presents a DT_1 observable higher than the DT_1 predicted having the variables contribution, GP and RS, a normal behaviour. The points displaced from the cloud and highlighted in red, in Figure 4 a), are those with a greater DT_1 also this points are the same that fall outside the control limits as shown in Figure 3.

5. Concluding remarks

A data driven approach, based on PCA and PLS, has been constructed and applied to identify and predict the status pattern of an asynchronous generator. A prediction model was built using operational historical data. A component performance monitoring model was developed to generate alarm signals based on the outputs predicted observations.

Using the methodology proposed in this work, an abnormal pattern of an equipment can be detected indicating a possible degradation of the equipment. In addition, these techniques can be used to support the condition-based maintenance in order to reduce maintenance costs and increase the availability of the generator.

Acknowledgments

Authors are grateful to the project CENIT-E Ocean Lider for the financial support of this work (CEN-20091039).

References

Dragomir O. E.,Gouriveau R., Dragomir F., Minca E., Zerhouni N., 2009, Review of Prognostic Problem in Condition-Based Maintenance, 2009, European Control Conference ECC’09, Budapest, Hungary

Geladi P.,Kowalsky B.R., 1986, Analytica Chimica Acta, 185, 1-17.

Hao L., Jinsong Y., Ping Z., Xingshan L., 2009, A Review on Fault Prognostics in Integrated Health Management, The Ninth International Conference on Electronic Measurement & Instruments, ICEMI’2009.

Jardine A.K.S, Lin D., Banjevic D., 2006, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing, 20, 1483-1510

Kusiak A., Verma A., 2011, Prediction of Status Patterns of wind turbines: A data mining approach, Journal of Solar Energy Engineering, 133, 110081-110090.

Lam H., Klemes J., Friedler F., Kravanja Z., Varbanov P., 2010, Software tools overview: process integration, modelling and optimisation for energy saving and pollution reduction, Chemical Engineering Transactions, Vol. 21, 487-491.

MacGregor, J.F., Kourti, T., 1995, Statistical process control of multivariate processes, Control Eng. Practice, Vol.3, No.3, 403-411.

Wold S., Esbensen K., Geladi P., 1987, Principal Components Analysis, Chemometrics and Intelligent Laboratory Systems, Vol. 2, 37-52.