Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia
Systems Engineering
Systems Engineering Procedia 1 (2011) 473–480

Detecting Changes in a Distillation Column by Using a Sequential Probability Ratio Test

Yahya CHETOUANI

University of Rouen,Chemical Engineering Department, Street Lavoisier, 76821, Mont Saint Aignan Cedex, France. Fax: (0033)235146130, Yahya.Chetouani@univ-rouen.fr

Abstract

In chemical plants, a reliable detection of anomalies is important for a safe operation. To this end, a fault detection (FD) method of abnormal operations applicable to a chemical process is presented in this paper. This method couples an Artificial Neural Network-Multi-Layer Perceptron (ANN-MLP) with a statistical module based on the sequential probability ratio test (SPRT) of Wald, for the analysis of the process residuals. To detect a change, this combination uses the mean and the standard deviation of the residual noise obtained from applying a NARX (Nonlinear Auto-Regressive with eXogenous input) model. The FD effectiveness is tested under real abnormal circumstances on a real plant as a distillation column. The experimental results obtained show the relevance of this method for the fast detection and the monitoring of this chemical process.

Selection and/or peer-review under responsibility of the Organising Committee of The International Conference of Risk and Engineering Management.
Keywords: Fault detection; Reliability; Safety; SPRT; ANNs; Distillation column

1. Introduction

In recent decades, an increasing demands for efficiency, product quality, reliability and process safety is observed. To meet these goals, operators have to equip the chemical plants with set-up real-time control systems. Unfortunately, a component fault, sensor fault or an abnormal operation can lead to an intolerable behaviour of the process and makes supervision delicate. In order to ensure an adequate safe operation, it is essential to set up procedures able of detecting these anomalies.

Traditionally, the most usually implemented FD methods have been based on model-based approaches. However, in modern process industry, there is a demand for data-based methods because of the complexity and the limited availability of the nonlinear models in chemical units. In order to meet the demands from industry concerning quality, efficiency and safety, numerous FD methods have been developed based on ANNs [1-6]. This neural approach is characterized with particular properties such as the ability to learn, generalization abilities, good approximation properties and simplicity of implementation [7-10].

Artificial Neural Networks (ANNs) have an adaptive behavior i.e. they are able to adjust and to modify their behavior according to nonlinear dynamics of processes. ANNs can be trained to learn new associations, complex modelling, functional dependencies and new patterns [11]. Owing to their inherent nature to model and learn complexities, ANNs have been successfully applied in medicine and biomedical studies [12]. Also they have found wide applications in various areas of chemical engineering [13, 14].

This paper introduces a real-time fault detection method, applicable to a chemical plant as a distillation column.

Selection and/or peer-review under responsibility of the Organising Committee of The International Conference of Risk and Engineering Management.
doi:10.1016/j.sepro.2011.08.069

This method consists in combining ANNs and the binary sequential probability ratio test of Wald for detecting changes. This strategy is based on a comparison of the real response with a forecast of this response. This forecast is provided by means of a neural model (NARX). The residual is analysed with the SPRT test. Experimental results show that this test is a powerful method and is very well suited for anomaly detection in this chemical unit.

This paper is organized as follows: Section 2 describes the neural training modelling, Section 3 describes the fault detection strategy, in Section 4 the experimental set-up is introduced and presents the experimental results with the proposed FD method, and Section 5 ends the paper.

2. Network training modelling

In many fields, there are several models representing systems that are dominated by nonlinear characteristics. Many physical processes which have nonlinear behavior can be well represented by a polynomial representation, Volterra or Wiener series, or other nonlinear techniques. It is showed that the AutoRegressive with eXternal input (ARX) model can be easily identified and outperform linear models [15]. Also, Nonlinear AutoRegressive Moving Average model with eXogenous inputs (NARMAX) model can provide a unified and simple representation for a wide class of discrete-time nonlinear stochastic systems. The NARMAX model is an input-output recursive model where the current output depends on lagged inputs, outputs and noise terms through a suitable nonlinear function. The simpler Nonlinear AutoRegressive with eXternal input (NARX) model is often preferred, although the absence of a disturbance model may result in bias problems i.e. a NARMAX model with the noise terms excluded. Structure determination of the NARX model is to choose sufficient and necessary terms for capturing the system dynamics. This model could represent a nonlinear model with a smaller number of parameters [13, 16]. In this study, the effectiveness of this technique for black-box nonlinear modelling is described.

The ANN propagates the error from the output layer (y_i) to the hidden layer to update the weight matrix (w_i). Each node produces an output signal, which is a function of the sum of its inputs:

$$y_i = \phi_{ANN}\left(\sum x_i\, w_i\right) \tag{1}$$

where $\phi_{ANN}(\cdot)$ is the activation function. In this study, the neural model is defined as follows [17, 18]:

$$y(t) = \phi_{ANN}(y(t-1),\ldots,y(t-n_y),\quad u_i(t-n_k),\ldots,u_i(t-n_k-n_u)) \qquad 1 \le i \le m \tag{2}$$

where $y(k)$ is the Auto-Regressive (AR) variable or system output; $u(k)$ is the eXogenous (X) variable or system input. n_y and n_u are the AR and X orders, respectively. m is the number of the used inputs. n_k is the time delay between u and y.

For smaller number of hidden layers and nodes, the performance of the ANN may not be satisfactory, while with too many hidden nodes, there is the risk of over-fitting the training data. In this case, the ANN generalization on the new data is poor. Different methods, both heuristic and systematic, can be chosen in order to select the number of hidden layers and the nodes [19]. There are a large number of selection methods of the ANN. Two most popular ways of significant structures are in general used. We start with a small number of nodes in the hidden layer and add new ones when a certain criterion is met (growing algorithms). On the other hand, we can start with a large number of nodes and delete some of them under certain conditions (pruning algorithms).

In this study, the hidden layer nodes have log-sigmoid transfer hyperbolic function as the activation function and the output have linear activation function. The ANN multi-layer feed forward network is trained to capture the underlying relationship between the input $u(k)$ and output $y(k)$ using the training data. The input data are presented to the network via the input layer. These data are propagated through the network to the output layer to obtain the network output. The network error (generally the mean square error function) is then determined by comparing the network output with the actual output. If the error is not smaller than a desired performance, the weights are adjusted and the training data are presented to the network again to determine a new network error. The aim of this step is to find the appropriate weights which minimize the cost function. This is usually done using an iterative procedure. One of the best known learning mechanisms for neural networks is the training algorithm of Levenberg-Marquardt [20], which is used along with back-propagation (BP). In this case the ANN is trained iteratively using the training dataset to minimize the performance function of mean square error (MSE) between the network outputs and the corresponding target values. The Levenberg-Marquardt Algorithm (LMA) shows the fastest convergence during the training process based on gradient descent methods because it performs as a compromise between the stability of the first-order optimization methods (steepest-descent method) and the fast convergence properties of the second-order optimization methods (Gauss-Newton method). In many cases, LMA is able to obtain lower mean square errors than any of the other algorithms [20]. For weights initialization, the Nguyen-Widrow initialization method [21] is best suited for the use with the sigmoid/linear network which is often used for function approximation. After training, the networks thus developed are tested with the test data set to assess the generalization capability of each developed network. The best ANN model developed is trained off-line and then used on-line for detecting faults in the separation unit. All computations have been made on Matlab® 7.0.4.

2. Basic theory of the SPRT method

The FD method used in this paper is the sequential probability ratio test (SPRT) introduced by Wald [22]. It is a powerful method of testing an alternative hypothesis (abnormal mode) against a null hypothesis (normal mode). The SPRT is then able to determine whether the hypothesis has been accepted or rejected, or if further information is required to determine the true answer. This makes the method relevant for anomaly detection. According to Wald [22] , the SPRT method is optimal in the sense that it requires a minimum number of samples needed to identify each mode. This is extremely important when the change occurs abruptly. The SPRT offers an advantage over some of the other methods by allowing formal hypothesis testing. In fact, this test method incorporates selection of type I and II error rates and a threshold of an unacceptable odds ratio for an outcome [23].

The SPRT test was used to supervise systems [24-27] used the SPRT as a real-time FD method, applicable to sewer networks, for the follow-up of rainy events. They used the innovation generated by the Kalman filter for building this statistical test. Sheu [28] used it for a real-time detection and characterization of freeway incidents. He showed that the proposed technology is capable of detecting freeway incidents in real time as well as characterizing incidents in terms of time-varying lane-changing fractions and queue lengths in blocked lanes, the lanes blocked due to incidents, and duration of incident.

In this paper the SPRT method is used to detect a change. For this reason, the false alarm probability, the alarm failure probability and the average time to alarm, are important parameters for FD reliable method. The SPRT algorithm tests and selects which of two hypotheses; a binary decision of one mode (normal) against an alternative mode (degradation).

The residual is a fault-indicating parameter. It represents the difference between the measurement and the prediction by the ANN model. Thus, the residual is an evaluation of the capacity of the ANN to predict the observed behaviour of the process. If this behaviour is normal and the measurement is free from fault, the residual will be low. In this case, the residual is assumed as a white noise and has a Gaussian distribution with a zero mean value but other value when a fault is present.

Taking into account these observations, the problem of fault detection in the process is formalized as:

Hypothesis H_0*:* the residual is symptomatic of a normal operating condition of the process. It is a random variable of Gaussian law with common variance σ_0^2 and a zero mean.

Hypothesis H_1*:* the innovation is symptomatic of an abnormal operating condition of the system. It is a random variable of Gaussian law with common variance σ_1^2 and a mean $\mu \neq 0$.

This test consists in computing the likelihood ratio l_k:

$$l_k = P_{\theta_1}(\gamma_{1.....k} / H_1) / P_{\theta_o}(\gamma_{1.....k} / H_0) \qquad (3)$$

where $\gamma(k) = y(k) - \hat{y}(k)$ is the sequence of the residual between time-step 1 and K and p the probability law conditioned upon the hypothesis H_0 and H_1.

Under assumption of independence the likelihood ratio is:

$$L_k = \prod_{i=1}^{k} \left(P_{\theta_1}(\gamma_i) / P_{\theta_o}(\gamma_i) \right) = \prod_{i=1}^{k} l_k \qquad (4)$$

Let us suppose that the probability follows a normal law:

$$p(\gamma;\theta) = \exp\left(-0.5(\gamma-\theta)^2/\sigma^2\right) / \left(2\pi\sigma^2\right)^{1/2} \qquad (5)$$

This likelihood ratio is compared to two thresholds α and β, in order to make the following decisions:

$$\alpha < L_k = \prod_{i=1}^{k} P_1(\gamma_i)/P_o(\gamma_i) = \prod_{i=1}^{k}(\exp\left(-0.5(\gamma_i - \bar{\gamma}_1)^2/\sigma_1^{\,2}\right)/\left(2\pi\sigma_1^{\,2}\right)^{1/2})/(\exp\left(-0.5(\gamma_i - \bar{\gamma}_0)^2/\sigma_0^{\,2}\right)/\left(2\pi\sigma_0^{\,2}\right)^{1/2}) < \beta \qquad (6)$$

The decision rule is as follows:

If $L_k > \alpha$ then the hypothesis H_0 is accepted; $L_k < \beta$ then H_1 is accepted. Otherwise no decision is taken with this sample. Thus, the sample numbers of the residual used to calculate the likelihood ratio is increased until knowing the state of the system.

The thresholds (α and β) are determined according to the probability of false alarm P_{fa} and missed alarm P_{nd}:

$\alpha = P_{nd}/(1-P_{fa})$ and $\beta = (1-P_{nd})/P_{fa}$

The strategy adopted here is to attempt to supervise both the variance and the mean of the residuals. En fact, a fault can induce simultaneously an increase in noise or a change of the mean without changing the nature of the distribution.

The Wald test performances are directly connected to the probabilities false alarm P_{fa} and P_{nd}. The behaviour is compared in real-time to the position of two line $\log\left(P_{nd}/(1-P_{fa})\right)$ and $\log\left((1-P_{nd})/P_{fa}\right)$

4. Experimental results

4.1. Experimental set-up

The proposed FD scheme is applied to a distillation unit. The feed tank contains a mixture to be separated (toluene-methylcyclohexane) with a mass composition at 23 % in methylcyclohexane. The operation in continuous mode involves charging the still with the mixture to be separated, bringing the column to equilibrium under total reflux. The product is introduced through the optimal feed tray so that the light components are volatilized, while the heavy part goes down again with the reflux in the column reboiler. The quality of the collected top product of the column depends on the reflux flow rate. The reflux ratio is varied through the magnetic valve by changing the relative quantities of material returning to the column and flowing to product storage. Feed preheating system is constituted by three elements of 250 W each one. In addition it has a low liquid level switch in order to avoid the running if the level is excessively low. The reciprocating feed pump is constituted by a membrane allowing firstly the suction of the mixture and the discharge towards the tank with a flow capacity F = 4.32 L.h^{-1}. The column has also a reboiler of 2 liters hold-up capacity, an immersion heater of a power Q_b = 3.3 kW and of a level liquid switch sensor which allows the automatic stop of heating if the level is insufficient. The column can be used in atmospheric (P_{atm}) or in vacuum conditions (P_r). The stirring of the mixture in the reboiler is ensured by the boiling mixture. The internal packing is made of Multiknit stainless 316L which enhances the mass transfer between the vapor and liquid phases. In order to approach the adiabatic conditions, a heat-insulating made of glass wool is laid around the column. A condenser is placed at the column overhead in order to condense the entire vapor coming out from the column. The cooling (Q_c) used in exchangers is water. The heat-transfer area of the total overhead condenser is 0.08 m^2. Moreover the reflux timer (R_t) allows to control the reflux ratio (R_r). It is monitored by the overhead product temperature (T_d). When the required distillate temperature (T_d) is attained, the reflux timer opens. In the opposite case, it remains closed. Distillation supervision control system allows to modify the parameters and to follow their evolution such as the pressure drop (ΔP), the flow or the temperatures at different points of the distillation column. This control system, therefore, must hold product compositions as near the set points as possible. The thermocouples are coupled to a calibrated amplification circuit (4-20 mA, 0-150°C) whose signals are inputted to the computer on-line, which permits the bottom and top temperatures to be obtained. The unit has twelve sensors which measure continuously the temperature throughout the column.

4.2. Reliability of modelling of the overhead product temperature (T_d)

The number of nodes in the input and the output layers depend on the number of input and output variables, respectively. In this study, the adopted strategy is chosen as follows; the initial model has a low number of parameters and hidden nodes were gradually added during learning until the optimal result is achieved in the test subset. In the present work, the number of hidden nodes was modified from 1 to 12. Also, one hidden layer was used. The complete training process of networks took approximately 80,000 epochs using the Levenberg-Marquardt algorithm. In this case, the model composed by the set of inputs (ΔP, R_t, Q_b, Q_f, Q_f, T_f, P_r and F) and the output (T_d) is given as follows:

$$T_d(t) = \phi(T_d(t-1), T_f(t-3), R_t(t-4), \Delta P(t-8), \ Q_f(t-8) \tag{7}$$

The difficult trade-off between model accuracy and complexity can be clarified by using model parsimony indices from linear estimation theory [18] such as Akaike's Information Criterion (AIC), Final Prediction Error (FPE) and Bayesian Information Criterion (BIC). A strict application of the indices would select a number N_h=5 because it exhibits the lowest of three indices for all the model structures compared. In conclusion, the developed network architecture used consisted of 5 nodes in the input layer and 5 nodes for the hidden layer. This reduced neural model is considered as a reliable one for describing the dynamic behaviour of the studied distillation column. In conclusion, the identified model (Fig.1) is reduced from a (9-9-1) neural structure to (5-5-1).

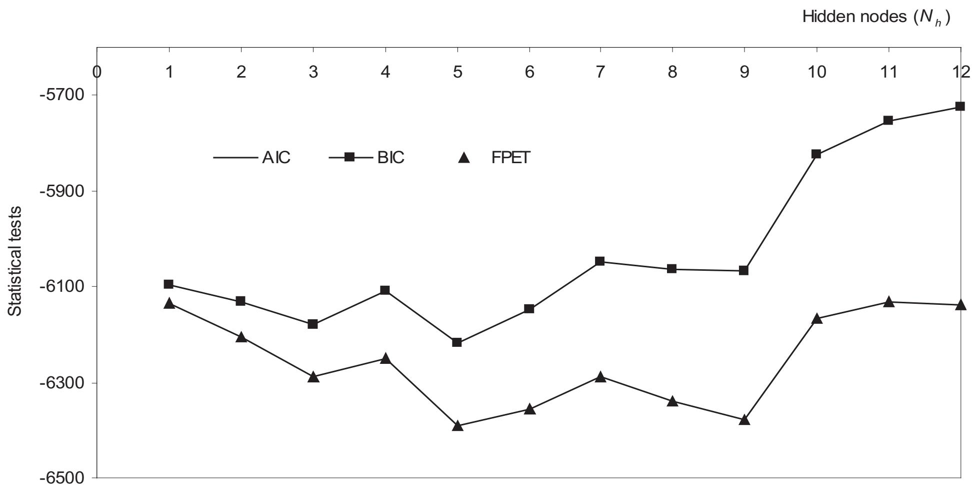

Fig. 1. Evolution of the statistical criteria

4.3. Application of the developed fault detection

Once the identified ANN is trained and tested, it is ready for detecting faults. The developed neural model is used both for the prediction and the FD procedure using the analysis of residuals. The proposed method is based on the SPRT as described in Section 3. This test should be processed to detect any real fault condition, rejecting any false alarms caused by noise. In this section, the proposed FD is applied to an experimental fault to verify its workability and effectiveness under real measurement conditions.

Statistics concerning the incidents which occurred in the distillation column showed that the most frequent faults are due to an inadequate heating power, feed pump, feed preheating, reflux or pressure used in the column. Consequently, the choice was carried out on this kind of faults. To illustrate the adopted approach for the fault detection, it was decided to attempt among these faults to detect a sudden closing of the reflux timer (R_t=0%). This fault is frequent in the distillation operation and introduces a large deviation in comparison with the normal behaviour. It is important to notice that this fault occurs at 8173 s causes a large decrease of the overheat temperature (T_d).

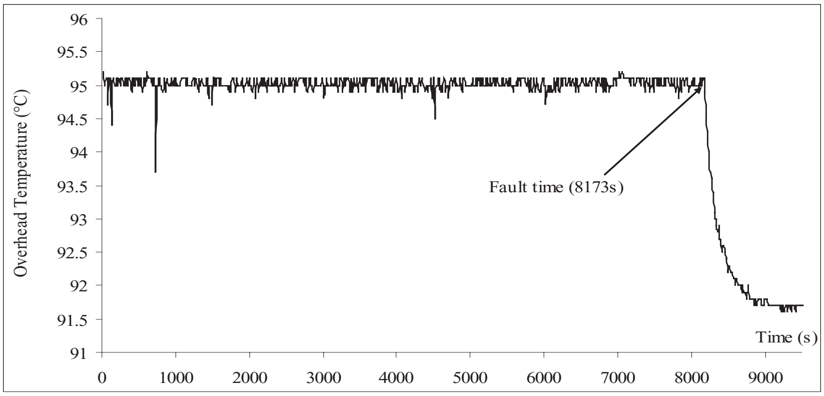

Fig. 2. Evolution of the residual of the overheat temperature

Fig. 2. shows the evolution of the residual in time. It is done by using a NARX which makes a prediction of the behaviour process. This prediction is compared with the actual behaviour. It is declared normal when the residual noise is white. If not, the plant operation is in abnormal mode. Fig. 2 shows also a residual decrease at 8173 s (fault time). This decrease should be detected by the SPRT because it exceeds a defined statistical threshold. The Wald test objective is to determine if a the observation is representative of the normal operation. These confidence levels (false and missed alarm rates) allows to require the shortest sample length for the decision. For the FD, The type I error is important from a viewpoint of economic maintenance but the type error II is more important, because it is linked to the distillation column safety. In fact, a chemical process requires the quick detection of abnormal values in the process. In this evaluation, α and β error levels of 0.05 were used to define acceptable type I and II error rates.

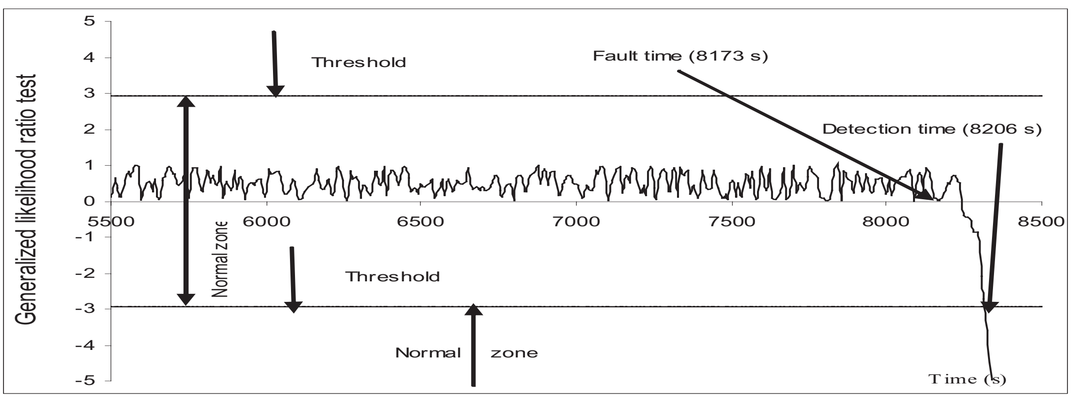

Fig. 3. Fault detection test based on SPRT

Fig. 3 shows two operating zones: a fault region and a confidence one. It is important to notice that the fault which occurs at 8173 s, is detected at 8206 s i.e. with a delay of 33 s which corresponds to a difference ($\Delta T \approx 0.5$ °C) between the temperature of the desired overhead temperature (T_d =95 °C) and the fault one (T_d =94.5 °C). This delay in detection defined as the difference between the time of the fault occurrence and the time of its detection, depends essentially on the evolution of the process dynamics.

5. Conclusion

Faults that change the system dynamics by causing surges of drifts of components, abnormal measurements, sudden shifts in the measurement channel, and other difficulties such as the decrease of instrument accuracy, an increase of background noise, reduction in actuator effectiveness etc., effect the characteristics of the SPRT test. This study shows that the combination of the ANN and the SPRT solves efficiently the FD problem.

The experimental results show that the one-layer perceptron network provides promising assignments to normal and faulty states of the investigated reference process. The dedicated example indicates the strength of the proposed approach for reliability of models, prediction of the future data and fault detection. It makes this combination very attractive for solving the modelling issues and fault detection in real time operation of complex plants as a distillation column. SPRT is used to advise the operator of an abnormal behavior of the process by setting-on the suitable alarm. Such alarm could be used in order to initiate the back-up procedure, to stop the operating plant or even to begin the diagnosis algorithm of the fault physical origin.

References

1. R. Isermann, Model-based fault-detection and diagnosis-status and applications, Annual Reviews in Control. 29 (2005) 71-85.

2. K.P. Ferentinos, Biological engineering applications of feedforward neural networks designed and parameterized by genetic algorithms. Neural Networks. 18 (2005) 934-950.

3. J. Korbicz, J.M. Koscielny, Z. Kowalczuk, W. Cholewa, (2004) Fault diagnosis: Models. Artificial intelligence, applications (1st ed.). Berlin: Springer.

4. T. Bouthiba, Fault location in ehv transmission lines using artificial neural networks, International Journal of Applied Mathematics and Computer Science. 14 (2004) 69-78.

5. M.M. Gupta, J. Liang, N. Homma, (2003) Static and Dynamic Neural Networks. New Jersey: Wiley.

6. O. Nelles, (2001) Non-linear Systems Identification. From Classical Approaches to Neural Networks and Fuzzy Models. Berlin: Springer,.

7. S.C Tan, C.P. Lim, M.V.C. Rao, A hybrid neural network model for rule generation and its application to process fault detection and diagnosis. Engineering Applications of Artificial Intelligence. 20 (2007) 203-213.

8. V. Uraikul, W.C. Chan, P. Tontiwachwuthikul. Artificial intelligence for monitoring and supervisory control of process systems. Engineering Applications of Artificial Intelligence. 20 (2007) 115-131.

9. J. Vieiraa, F.M. Diasb, A. Motac, Artificial neural networks and neuro-fuzzy systems for modelling and controlling real systems: a comparative study. Engineering Applications of Artificial Intelligence. 17 (2004) 265-273.

10. Chetouani Y., Using neural networks and statistical tests for detecting changes in the process dynamics. International Journal of Modelling, Identification and Control. 3 (2008) 113-123.

11. R. Fuller, (2000) Introduction to Neuro-Fuzzy Systems. New York: Springer.

12. J. Zurada, W. Karwowski, W.S. Marras, A neural network-based system for classification of industrial jobs with respect to the risk of low back disorders. Applied Ergonomics. 28 (1997) 49-58.

13. R. Sharma, K. Singh, D. Singhal, R. Ghosh, Neural network applications for detecting process faults in packed towers. Chemical Engineering and Processing. 43 (2004) 841-847.

14. D.M. Himmelblau, Applications of artificial neural networks in chemical engineering. Korean J. Chem. Eng., 17 (2000) 373-392.

15. T. Söderström, P. Stoica, (1989) System Identification. Englewood Cliffs, NJ: Prentice-Hall.

16. Y. Chetouani, Experimental study and nonlinear modelling by artificial neural networks of a distillation column. International Journal of Reliability and Safety. 4 (2010) 265-284.

17. A. Zaknich, (2003) Neural networks for intelligent signal processing. Singapore: World Scientific.

18. L. Ljung, (1999) System Identification, Theory for the User. NJ: Prentice Hall, Englewood Cliffs.

19. S. Haykin, (1999) Neural networks. Prentice Hall.

20. T.C. Chen, D.J. Han, F.T.K. Au, L.G. Tham, Acceleration of Levenberg-Marquardt training of neural networks with variable decay rate. In: Neural networks. Proceedings of the international joint conference. 3 (2003) 1873-1878.

21. D. Nguyen, B. Widrow, Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. International Joint Conference of neural networks. 3 (1990) 21-26.

22. A. Wald, (1947) Sequential Analysis. New York: Wiley.

23. D. Spiegelhalter, O. Grigg, R. Kinsman, Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. Int J Qual Health Care. 15 (2003) 7-13.

24. C. Bonivento, A. Tonielli, A detection estimation multifilter approach with nuclear application. In: The Ninth IFAC World Congress, Budapest. (1984) 1771-1776.

25. P. Coirault, J.D. Gabano, J.D. Pincon, J.C. Trigeassou, Monitoring a D.C. motor using parameter estimation and binary sequential tests of hypothesis. In: International Conference on Fault Diagnosis. 1993, Toulouse, France.

26. A. Racz, Detection of small leakages by a combination of dedicated Kalman filters and an extended version of the binary sequential probability ratio test. Nucl. Technol. 104 (1993) 128-146.

27. E. Piatyszek, P. Voignier, D. Graillot, Fault detection on a sewer network by a combination of a Kalman filter and a binary sequential probability ratio test. Journal of Hydrology. 3 (2000) 258-268.

28. J.B. Sheu, A sequential detection approach to real-time freeway incident detection and characterization. European Journal of Operational Research. 2 (2004) 471-485.