Asian Journal of Information Technology 5 (10): 1166-1171, 2006

Fault Diagnosis of Technical Processes Based on the Multi-Model Approach

Messaoud Ramdani and Noureddine Doghmane
Department of Electronics, Faculty of Engineering, University of Annaba,
BP. 12, 23000, Annaba, Algeria

Abstract: The supervision of many technical systems is often a challenging task due mainly to various nonlinearities. In this study, a multi-model approach for fault detection and diagnosis is proposed as an effective way since it allows to derive good process models valid over a wide range of operation and subsequently to detect changes of the current process behaviour. The diagnosis task is accomplished by decomposing the complex process into several sub-processes in order to generate a set of structured residuals. The validity of the approach is illustrated on the well known academic three tanks benchmark and different faults can be detected and isolated continuously, over several operating regimes.

Key words: Parity equations, nonlinear dynamic processes, multi-modelling, fuzzy approximate reasoning

INTRODUCTION

Due to the increasing complexity of modern technical processes and the growing demands for quality, cost efficiency, availability, reliability and safety, there is a need for effective fault tolerant control and process supervision techniques. Fault tolerance can be achieved either by passive or by active strategies. The main principle behind the passive approach is to ensure that the controlled system becomes insensitive with respect to faults. However, the active study provides fault accommodation, i.e. the reconfiguration of the control policy when a fault has occurred. Thus, fault diagnosis has become an important issue and during the last three decades a lot of work has been done in this area, resulting in different techniques with various acceptances in practice.

Most of the model-based FDI technologies have been developed for the linear systems, but the monitoring and the diagnosis of nonlinear processes remains a challenge. In general, the nonlinear systems are firstly linearised at an operating point and then robust techniques are applied to generate residuals, which are robust against limited parameter variations. The strategy only works well when the linearization does not cause a large mismatch between linear and nonlinear models and when the system operates near the specified operating point. Therefore such techniques have limited robustness when considering gross plant changes and nonlinearity.

In the lack of first-principle models, empirical models like neural networks can be used for the purposes of process supervision. The main problems with these approaches are the difficulty in analysing, in a rigorous mathematical way, their robustness/sensitivity and the scalability; i.e. a network trained for a specific plant may be inappropriate for other plant. To overcome the problem of precision and accuracy in FDD, various approaches based on fuzzy logic have been also suggested. However, the fuzzy logic approach is not only required on its own, but as a framework for combining different paradigms. More specifically, quantitative model-based and soft-computing are combined to exploit the benefit of each. Another powerful approach for residual generation is based on observers. The common way is to obtain a set of residuals by comparing the actual measurements with their estimates obtained with the help of observers. Unfortunately, the design of nonlinear observers is not a straightforward task, even if the nonlinear process is completely known.

Overview of fault detection and diagnosis: Different approaches for fault detection and diagnosis using mathematical models have been initiated and developed from the early seventies to now, see[1-4]. They can be splinted in two categories. The first one is based on state estimation and includes detection filter, parity space approaches as well as observer-based methods[5,6]. The second category includes parameter estimation techniques[7].

Corresponding Author: Messaoud Ramdani, Department of Electronics, Faculty of Engineering, University of Annaba, BP. 12, 23000, Annaba, Algeria

Generally, automatic fault diagnosis can be viewed as a sequential process involving two stages:

- symptom extraction
- fault diagnosis

Symptom extraction is mainly required for data reduction. Many methods have been developed during the past decade and the choice of a specific model is somewhat dependent on the nature of the process. The symptoms can be analytic or heuristic. While analytic symptom generation is based on measurements, heuristic symptom extraction requires a human operator observing the process. In the fault diagnosis stage, the task consists of the detection of the type of fault with as many details as possible such as the fault size, location and time of detection.

In model-based fault detection and diagnosis, the most important task is the generation of residual signals which are independent of the disturbances. The most common way uses observers. The basic idea behind the observer or filter-based methods is to estimate the states and the outputs of the system from a subset of the measurements by using either Luenberger observer(s) in a deterministic setting or Kalman filter(s) in a stochastic setting. Subsequently, the weighted output estimation error is used as residual signals.

MATERIALS AND METHODS

The major motivation for the multiple modelling methodologies is that locally there are less relevant phenomena and interactions are simpler. Under this study, the underling nonlinear mapping is inferred by a local approximation using only nearly states.

The basic philosophy behind this modelling strategy is to partition the input domain into multiple subsets. Such local representations include RBF nets and fuzzy systems[8,9]. The locality property can be used to make models more interpretable and computationally efficient.

A large number of nonlinear, dynamic processes with m inputs u and one output y can be described in the discrete domain by means of Eq. 1

$$\begin{aligned} y(k) &= f(x(k)) \\ x(k) &= [u_1(k-1),\cdots,u_1(k-n_1), \\ &\quad u_m(k-1),\cdots,u_m(k-n_m), \\ &\quad y(k-1),\cdots,y(k-n_y)\ ] \end{aligned} \tag{1}$$

The nonlinear function f (•) can be approximated with the Local Model Networks

$$\hat{y}(x,\theta) = \sum_{i=1}^{c} \hat{y}_i(x,\theta_i)\rho_i(\tilde{x}) \tag{2}$$

where, x is the observed input vector and $\theta = [\theta_1^T,\ldots,\theta_c^T]$ is the parameter vector and $\tilde{x}$ defines the operating point of the system, usually given by a function $\tilde{x}$ =H(x). This is a vector which often can be defined on a lower dimensional subspace of the input space to form the gating or weighting functions for the local models (which are defined on the full input space). The basis functions used are defined,

$$\rho_i(\tilde{x}) = \frac{\phi\left(d(\tilde{x},c_i,\sigma_i)\right)}{\sum_{j=1}^{c}\phi\left(d(\tilde{x},c_j,\sigma_j)\right)} \tag{3}$$

where $\phi(.)$ is the underlying normalized basis function, e.g. a Gaussian $\phi(d) = \exp(-d^2/2)$ is a weighted Euclidean distance metric which measures the distance of the current operating point $\tilde{x}$ from the basis function centre c_i the normalized basis function $\rho_i(.)$ now sum to unity. The local models used are linear:

$$\hat{y}_i(x,\theta_i) = \left[1x^T\right]\cdot\theta_i \tag{4}$$

If the centres and standard deviations of the validity functions are known, the estimation of the local linear model parameters θ_i is a linear optimization problem. The parameters can therefore be evaluated by employing linear least squares optimization algorithms. In the following subsection, we consider the learning of parameters of the local models for a given model structure, i.e. we are estimating θ for an a priori given set of c, σ.

Parameter estimation: For locally linear models, the coefficients can be estimated using the least squares method on the prediction targets Y. Once an initial structure is defined (i.e. c_i's and σ_i's) and local models as in (4), so the learning problem is a straightforward application of linear regression techniques to find parameters θ which best fit the data. The available data samples are collected in the regression data matrix X and the output vector

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix},\quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \tag{5}$$

where, N is the number of training data. Staking the data into matrices, we get the following regression model:

$$Y = \Phi\theta^* + \varepsilon \tag{6}$$

where Φ is the design matrix, the rows of which are defined by

$$\phi_k = \left[\rho_1(\tilde{x}_k)\left[1x_k^T\right],\cdots,\rho_c(\tilde{x}_k)\left[1x_k^T\right]\right] \tag{7}$$

So that the design matrix Φ, vector of output measurement Y and errors ϵ are

$$\Phi = \left(\phi_1^T,\cdots,\phi_N^T\right)^T,\; Y = \left(y_1,\cdots,y_N\right)^T$$
$$\varepsilon = \left(\varepsilon_1,\cdots,\varepsilon_N\right)^T$$

The standard least squares criterion for this estimation problem is

$$J(\theta) = \frac{1}{N}\left(Y - \Phi\theta\right)^T\left(Y - \Phi\theta\right) \tag{8}$$

and the Moore-Penrose pseudo inverse of Φ, Φ^+ is used to estimate the weights:

$$\hat{\theta}_{LS} = \Phi^+Y = \left(\Phi^T\Phi\right)^{-1}\Phi^TY \tag{9}$$

The computation of the pseudo inverse uses the Singular Value Decomposition (SVD) to decompose any N x p matrix Φ, such that $\Phi = USV^T$ and the pseudo inverse of Φ is: $\Phi^+ = US^+V^T$

Then, the solution of the regression problem (6) can be calculated:

$$\hat{\theta}_{LS} = VS^{-1}U^TY \tag{10}$$

In this way, the learning is global because it is based on assumption that all the parameters θ would be learned in a single regression operation and the local models cooperate to solve the regression task. Unfortunately, this may not always be computationally feasible if a large number of training samples or local models are needed for a particular problem. With a global learning, the parameters of the local models cannot be interpreted independently of neighbouring local models, which mean that they cannot be seen as local approximation of the underlying system. An alternative to global learning which is less prone to these disadvantages is to locally estimate the parameters of each of the local models as defined in Eq. 3 independently. This is achieved using a set of local estimation criteria for the ith local model

$$J_i(\theta_i) = \frac{1}{N}\left(Y - \Phi_i\theta_i\right)^T Q_i\left(Y - \Phi_i\theta_i\right) \tag{11}$$

where i =1, . . . ,c. Q_i is an N×N diagonal weighting matrix defined as:

$$Q_i = \mathrm{diag}\left(\rho_i\left(\tilde{x}_1\right),\cdots,\rho_i\left(\tilde{x}_N\right)\right) \tag{12}$$

Now, the criteria J_i is minimized by the locally Weighted Least Squares (WLS) estimate of the local model parameter vector θi. In matrix terms, now we have

$$\begin{aligned}\hat{\theta}_{WLS} &= \left(\hat{\theta}_{WLS,1},\cdots,\hat{\theta}_{WLS,c}\right)^T\\ \hat{\theta}_{WLS,i} &= \left(\Phi_i^TQ_i\Phi_i\right)^{-1}Q_i^TQ_iY,\quad i = 1,\cdots,c\end{aligned} \tag{13}$$

where Φ_i is an Nx(n+1) submatrix of Φ corresponding to the ith local model.

Diagnosis by fuzzy approximate reasoning: The fuzzy approximate reasoning scheme is an effective way to take into account the vagueness and the uncertainty which are inherently present in real world applications. In the context of diagnosis, the domain of possible events Ω = (f_1, f_2,..., f_M, ff) where f_i denotes a particular faulty state, while ff = f_{M+1} stands for the fault-free state. The symptoms $\{S_1,S_2, . . . ,S_r\}$ provide evidence of faults. Every symptom S_i is typically described by some fuzzy sets like Low, Height or Negative, Zero, Positive as shown in Fig. 1. The observation of some symptoms is regarded as a source of partial information that matches some rules of a fuzzy rule based system which depends on the elements of the incidence matrix $\Lambda = \lambda_{ij}$. An entry $\lambda_{ij} \neq 0$ means that the jth fault causes the ith analytical symptom to become different from zero, i.e. Height. The approximate reasoning model consists of K rules of the form:

$$\begin{aligned}&R^i : \text{IF } S_1 \text{ is } A_1^i \text{ and} \cdots \text{and } S_r \text{ is } A_r^i\\ &\quad \text{Then } f_1 \text{ is } B_1^i \text{ and} \cdots \text{and } f_{M+1} \text{is } B_{M+1}^i\end{aligned}$$

The firing strength of the ith rule is defined by the product of the membership degrees of the corresponding fuzzy sets:

$$\mu^i(S) = \prod_{j=1}^{r}\mu_{A_j^i(S_j)} \tag{14}$$

where $\mu_{Aj(Sj)}^{i}$ is the membership function of the fuzzy set $A_j^i(S_j)$. The overall output is computed by the weighted average[7]. In this way, the

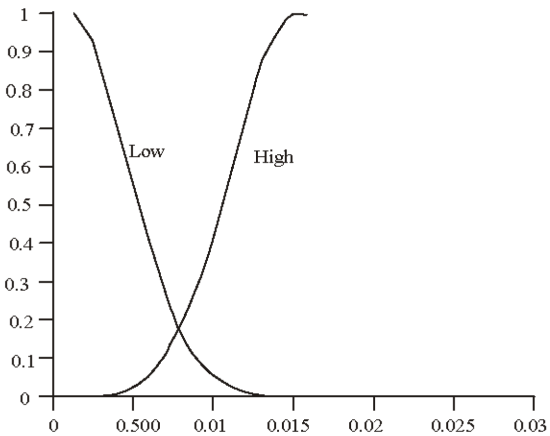

Fig. 1: Analytic symptom description by linguistic variables {Low, High} and their membership functon $\mu_{AiJ}(S^j)$, i=1,2.

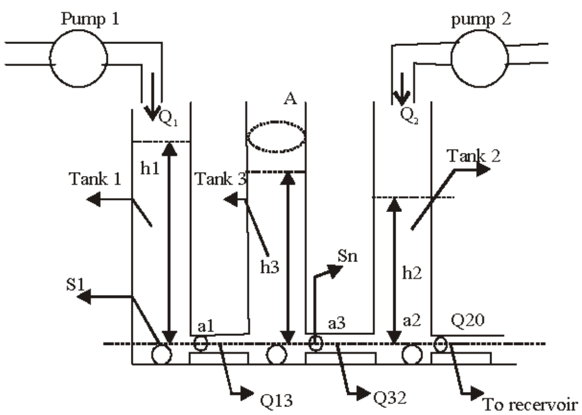

Fig. 2: The three-tank benchmark system

final result of the fuzzy inference system is a set of membership value to the different faulty states.

RESULT AND DICUSSION

Benchmark description: The academic three-tank benchmark process is used in the simulations. The plant consists of three cylinders T1, T3 and T2 with the equivalent cross section A (Fig. 1). These are connected serially with each other by cylindrical pipes with the cross section Sn. Located at T2 is the single so-called ”nominal outflow valve”. It has also a circular cross section Sn. The out flowing liquid (usually distilled water) is collected in a reservoir, which supplies the pumps 1 and 2. The pump flow rate Q1 and Q2 denote the input signals, which are controllable. The required level measurements are carried out by piezo-resistive differential pressure sensors and the reference pressure is the atmospheric pressure.

Let us define the following variables and the parameters: a_i, outflow coefficients; h_i, liquid levels (m);

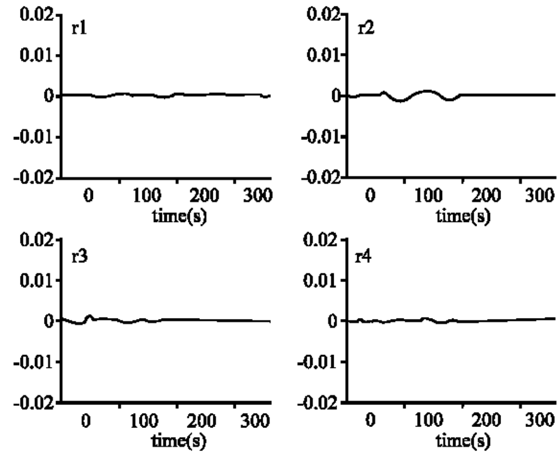

Fig. 3: Fault free

Q_{ij}, flow rates (m3/s); Q_1 and Q_2, supplying flow rates (m3/s); A, section of cylinder (m2); Sn, section of connection pipe (m2); where i = 1,2,3 and (i, j) = {(1, 3); (3,2); (2, 0)}. By using the balance equations for the three cylinders, the model is setup as follows:

$$\begin{aligned} A\frac{dh_1}{dt} &= -Q_{13} + Q_1 \\ A\frac{dh_2}{dt} &= Q_{13} - Q_{32} \\ A\frac{dh_3}{dt} &= Q_{32} - Q_{20} + Q_2 \end{aligned} \tag{15}$$

where the follows are given by generalized Torricelli-rule,

$$\begin{aligned} Q_{13} &= a_1 S_n \operatorname{sign}(h_1 - h_2)\sqrt{2g|h_1 - h_3|} \\ Q_{32} &= a_3 S_n \operatorname{sign}(h_3 - h_2)\sqrt{2g|h_3 - h_2|} \\ Q_{20} &= a_2 S_n \sqrt{2gh_2} \end{aligned} \tag{16}$$

The state vector is $x = [h_1\ h_2\ h_3]^T$ and the input vector is $u = [u_1\ u_2]^T$. The actual parameters are

$$\begin{aligned} &A = 0.0154m^2, S_n = 5\times10^{-5}m^2, Q_{1max} \\ &= Q_{2max} = 100ml/s, h_{max} = 62\pm1cm, g = 9.81m/s^2, \\ &a_1 = 0.450, a_2 = 0.611, a_3 = 0.462 \end{aligned}$$

The mechanistic model can be used to simulate different faults such as

- Faults in level sensors;
- Faults in pump output;
- Blockages in connecting pipes between tanks;
- Leaks in each tank.

Let us take the leakage in tank 1 as a fault caused by a hole of radius r, then according to the generalized Torricelli-rule:

$$Q^{1}_{leak} = a_1 \pi r^2 \sqrt{2gh_1} \tag{17}$$

and the dynamics of tank 1 with fault becomes:

$$A\frac{dh_1}{dt} = -Q_{13} + Q_1 - Q^{1}_{leak} \tag{18}$$

Model structure: The task of fault detection scheme is to detect changes of the current process behaviour compared to the nominal one during fault-free operation. One possibility for the generation of symptoms is the comparison of measured output signals with estimated signals.

A set of structured residuals can be designed, each independent of different inputs. Therefore, some inputs do not have impact on specific residuals and the decoupled residuals remain small, whereas the other residuals are deflected. The pattern of deflected and undeflected symptoms are are appropriate deatures for the assignment of faults.

The selection of some suitable mathematical relationships is not an easy task and often some expert knowledge is necessary. Obviously, the expert knowledge is also critical to determine the structure of the models, i.e. the number of fuzzy sets and the shape of the membership functions. In the study under consideration, three basic relationships are considered:

$$\Delta h_1 = F_1(h_{13}, Q_1), \text{ where } h_{13} = H_1 - H_3 \tag{19}$$

$$\Delta h_2 = F_2(h_{32}, Q_2, H_2), \text{ where } h_{32} = H_3 - H_2 \tag{20}$$

$$\Delta h_3 = F_3(h_{13}, h_{32}) \tag{21}$$

To obtain an additional residual, another relation has been proposed

$$(\Delta h_1 + \Delta h_3) = F_4(h_{32}, Q_1) \tag{22}$$

In the identification of these relationships, the membership functions are assumed to be Gaussian with a number of five for every variable.

In a secod step, the different sub-models have to be identified by means of suitable data. To obtain good identification results the choice of the excitation signals is of geat importance. Good results can be achieved by using the so-called Amplitude-Modulated Pseudo Random Binary Signal (AMRBS) which is an extension of the standard PRBS. Instead of applying only two amplitudes-which is sufficient for linear processes-the amplitudes are distributed over the whole range of values in order to ensure that all amplitudes of the excitation signal appear with the same probability

Diagnostic reasoning: The effects of all faults are summarized in the incidence matrix of Table 1. Each fault

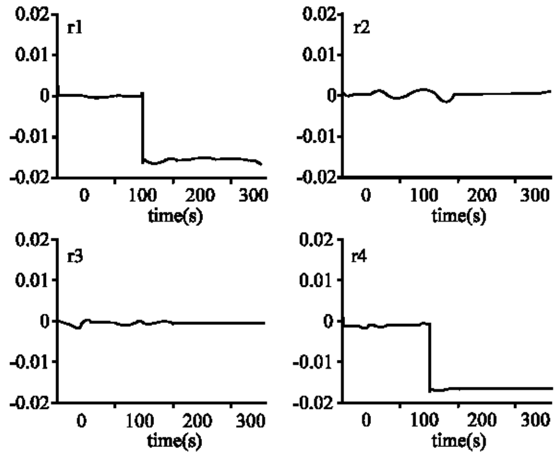

Fig. 4: Fault in Q1

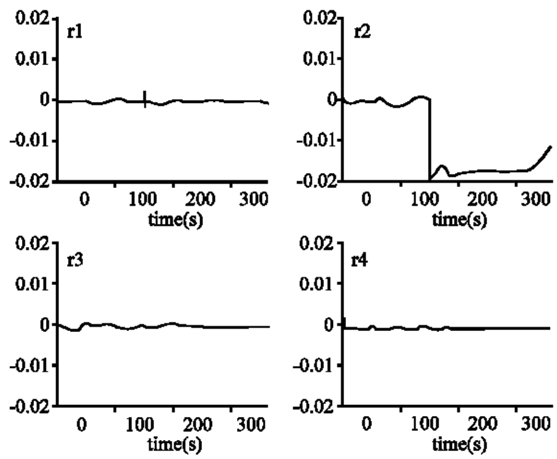

Fig. 5: Leak in tank 2

Table 1: Incidence matrix

	Faults										
r_i	f_{h1}	f_{h2}	f_{h3}	f_{Q1}	f_{Q2}	f_{Q13}	f_{Q32}	f_{Qout}	f_{L1}	f_{L2}	f_{L3}
r_1	1	0	1	1	0	1	0	0	1	0	0
r_2	0	1	1	0	1	0	1	1	0	1	0
r_3	1	1	1	0	0	1	1	0	0	0	1
r_4	1	1	1	1	0	0	1	0	1	0	1

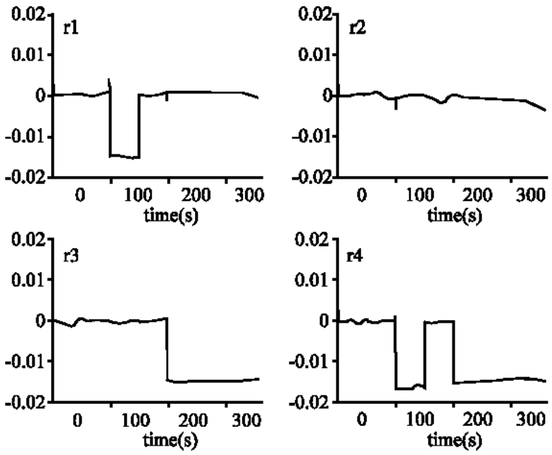

Fig. 6: Fault in pump Q_1 followed by a leak in tank 3

leads to significant deviation in at least one residual. Therefore all faults can be detected. However, using the proposed residual set, only seven different faulty states can be isolated:

- $f_1 = f_{Q_1} \vee f_{L_1}$, • $f_2 = f_{Q_{13}}$
- $f_3 = f_{Q_2} \vee f_{Qout} \vee f_{L_2}$, • $f_4 = f_{L_3}$
- $f_5 = f_{h_1}$, • $f_6 = f_{h_2} \vee f_{Q_{32}}$
- $f_7 = f_{h_3}$

The behaviours of the four residuals r_1,..., r_4 (analytical symptoms) are shown in Fig. 4-6. Fig. 3 corresponds to the fault free situation but the rest of Fig., a fault is provoked at time t = 150s.

Fig 4 shows the values taken by the residuals when a stuck in pump Q1 is provoked. As can be seen in Table 1 shows the residuals r1 and r4 have to change their values, but r2 and r3 must stay near zero. Fig. 5 corresponds to a leak in tank 2 where the only residual r2 has changed. Fig. 6 shows the effect of a fault in pump Q1 between t = 100s and t = 150s followed by a leak in tank 3 after t = 200s. The fuzzy rule-based system consists of eight rules; one for each faulty state and one for the fault free state.

CONCLUSION

This study describes the fault detection and diagnosis of technical plants in a systematic way. The analytic redundancy concept is used to detect changes of the current process behaviour compared to the nominal one during fault-free operation. The symptoms are extracted by the comparison of measured output signals with estimated signals. In order to cope with uncertainty, the residual evaluation step is based on fuzzy approximate reasoning scheme to produce a ranked list of possible fault candidates.

REFERENCES

1. Chen, J. and R.J. Patton, 1999. Robust model-based fault diagnosis for dynamic systems, Kluwer, Boston.
2. Gertler, J., 1998. Fault detection and diagnosis in engineering systems, Marcel Dekker, New York.
3. Simani, S., C. Fantuzzi and R.J. Patton, 2003. Model-based fault diagnosis in dynamic systems using identification techniques, Springer, London.
4. Frank, P.M., 1990. Fault diagnosis in dynamic systems using analytical and knowledge based redundancy, Automatica, 26: 459-474.
5. Willsky, A.S., 1976. A survey of design methods for failure detection systems, Automatica, 12: 601-611.
6. Gertler, J., 1990. A new structural framework for parity Equation based failure detection and isolation, Automatica, 26: 381-388.
7. Isermann, R., 1997. Supervision, fault-detection and fault-diagnosis methods. An introduction, Control Eng. Practice, 5: 639-652.
8. Johansen, T.A. and B.A. Foss, 1993. Constructing Narmax models using ARMAX models, Intl. J. Control, 58: 1125-1153.
9. Takagi, T. and M. Sugeno, 1985. Fuzzy identification of systems and its applications to modeling and control, IEEE Trans. Syst., Man, Cybern., 15: 116-132.
10. Tsoukalas, L.H. and R.E. Uhrig, 1997. Fuzzy and Neural Approaches in Engineering. New York: Wiley.