PERGAMON

Engineering Applications of Artificial Intelligence 13 (2000) 345–356

Engineering Applications of
ARTIFICIAL
INTELLIGENCE

www.elsevier.com/locate/engappai

Neural network based fault diagnosis using unmeasurable inputs

S.H. Yang[a,*], B.H. Chen[a], X.Z. Wang[b]

[a]Department of Computer Science, Loughborough University, Loughborough, LE11 3TU, UK
[b]Department of Chemical Engineering, University of Leeds, Leeds, LS2 9JT, UK

Received 1 December 1998; accepted 1 January 2000

Abstract

Much of the earlier work presented in the area of on-line fault diagnosis focuses on knowledge based and qualitatively reasoning principles and attempts to present possible root causes and consequences in terms of various measured data. However, there are many unmeasurable operating variables in chemical processes that define the state of the system. Such variables essentially characterise the efficiency and really need to be known in order to diagnose possible malfunction and provide a basis for deciding on appropriate action to be taken by operators. This paper is concerned with developing a soft sensor to assist in on-line fault diagnosis by providing information on the critical variable that is not directly accessible. The features of dynamic trends of the process are extracted using a wavelet transform and a qualitative interpretation, and then are used as inputs in the neural network based fault diagnosis model. The procedure is illustrated by reference to a refinery fluid catalytic cracking reactor.

Keywords: Fault diagnosis; Soft sensor; Neural network; Wavelet transform; Dynamic trend; Fluid catalytic cracking (FCC)

1. Introduction

With the increasing integration and complexity of chemical processes on-line fault diagnosis has been widely recognised as a powerful support tool for operators. This is particularly important where conditions give rise to higher probabilities of equipment failure and operator error. Any know-how to help to deal with process deviations that could result quality and safety problems is clearly desirable. A number of researchers have proposed fault diagnosis architectures for industrial processes and robotic systems. Much of the previous work on this topic has been based on knowledge-based systems, deep mathematical models and/or statistical models (Wennersten et al., 1996; Vemuri and Polycarpou, 1997a). In the last two decades many contributions have been made using neural networks (Maki and Loparo, 1997; Meghlaoui et al., 1998; Vemuri and Polycarpou, 1997b; Hoskins and Himmelblau, 1988; Venkatasubramanian and Chan, 1989; Venkatasubramanian et al., 1990; Hoskins et al., 1991; Wang et al. 1996). However, these approaches are static in nature because the neural networks are trained using only steady-state data. If the steady-state operating conditions are changed, the network must be retrained in order to work properly. Some researchers trained the network using dynamic data, a number of sets of time series data (Ohga and Seki, 1993), and types of behaviour of dynamic trend such as increasing, decreasing and steady (Maki and Loparo, 1997).

There are two features in chemical processes that should be addressed in any fault diagnostic systems. One feature is that there are many unmeasured operating variables in chemical processes that can play a key role in defining the state of the process and are critical in the diagnosis of possible malfunctions. Vemuri and Polycarpou (1997a, 1997b) described a fault diagnostic algorithm for a class of non-linear dynamic systems with modelling uncertainties when not all states of the system are measurable. The main idea behind their

* Corresponding author.
E-mail address: s.h.yang@lboro.ac.uk (S.H. Yang).

0952-1976/00/$ - see front matter
PII: S0952-1976(00)00005-1

Nomenclature

A_{rcs}	cross sectional area of the slide valve (m^2)
Cp_a	heat capacity of air (kJ/Nm^3 °C)
Cp_s	heat capacity of catalyst (kJ/kg °C)
CSVS	catalyst slide valve stuck
Dp_{rcs}	differential pressure of regenerated catalyst slide valve (MPa)
DX_{max}	maximum deviation from the steady state
E	expectation operator
$E\mathbf{X}$	expectation of vector of key operating variables
F_a	airflow rate to the regenerator (Nm^3/h)
FCW	feedstock containing water
F_w	steam flow rate to the regenerator (kg/h)
g_n	coefficient of the high pass filter
G_{rc}	catalyst circulation rate (kg/h)
G_{rc2}	catalyst circulation rate calculated from Eq. (2) (kg/h)
G_{rc3}	catalyst circulation rate calculated from Eq. (3) (kg/h)
h_n	coefficient of the low pass filter
H_{rg}	catalyst hold-up in the regenerators (kg)
Hw_a	steam enthalpy at inlet temperature to regenerators (kJ/kg)
Hw_{fg}	steam enthalpy at flue gas temperature from the regenerators (kJ/kg)
N	number of extrema points in a window length
O_{fg}	oxygen concentration in flue gas of the regenerator (%)
O_{fg2}	oxygen concentration in flue gas of second regenerator (%)
P_{ra}	pressure of the reactor (MPa)
Q_{loss}	heat loss from the regenerator (kJ)
T_a	air temperature (°C)
TCSRF	temperature control system of the reactor failure
T_f	temperature of feedstock (°C)
T_{fg}	flue gas temperature in the regenerator (°C)
T_{ra}	outlet temperature of the reactor (°C)
T_{rg}	temperature of the regenerator (°C)
x	measure of the degree of the specified behaviour
X_0	starting point in a window length
X_{avg}	average value of the staring, extrema, and end points in a window length
X_i	extrema point in a window length ($i = 1, \ldots, N$)
X_{max}	maximum value of the staring, extrema, and end points in a window length
X_{max}^{end}	last maximum point of extrema in a window length
X_{max}^{first}	first maximum point of extrema in a window length
X_{min}^{end}	last minimum point of extrema in a window length
X_{min}^{first}	first minimum point of extrema in a window length
X_{min}	minimum value of the staring, extrema, and end points in a window length
X_{ss}	steady-state value
X_{N+1}	end point in a window length
Y	vector of key operating variables
Y_{k+1}	vector of key operating variables at instant $k + 1$
ΔH_{cb}	heat of coke combustion in the regenerator (kJ/kg coke)
ρ_{rc}	weight density of catalyst flowing through regenerated catalyst slide valve (kg/h)
υ_{oc}	Nm^3 amount of oxygen spent by burning 1 kg coke in the regenerator
ϵ	a small positive number
δ	a small positive number
ζ	flow rate coefficient of regenerated catalyst slide valve
ζ_{next}	flow rate coefficient of regenerated catalyst slide valve at next instant
$\zeta_{current}$	flow rate coefficient of regenerated catalyst slide valve at current instant
η	filter coefficient less than unity

Superscript

T	transposition of matrix

approach is to monitor the plant for any off-nominal system behaviour due to faults utilizing a non-linear online approximator with adjustable parameters. The online approximator provides an estimator of the fault and the uncertainties. Another feature of chemical processes is that the fault usually occurs during transient periods of operation and fault symptoms are embedded in dynamic trends of process variables. There are different ways to deal with dynamic trends (Maki and Loparo, 1997; Basseville, 1998; Yang et al., 1995; Yang et al., 1997; Chen and Jong, 1993) and a comprehensive summary can be found in Wang et al. (1999). A two-stage neural network is proposed as the basic structure of the fault detection and diagnostic system by Maki and Loparo (1997). The first stage of the network detects the dynamic trend of each measurement, and the second stage of the network detects and diagnoses the faults. Observation histories

are categorized into three types of behaviour: increasing, decreasing and steady, the first stage network is trained to give this type of trend information including the extent of change. In our previous works (Yang et al., 1995; Yang et al., 1997; Wang et al., 1997) the dynamic response is divided into three stages, each stage corresponds to particular feature of the process behaviour. Stage 1 is associated with the order of the system, Stage 2 with the maximum rate of change of the system when maximum control input is applied and Stage 3 relates to the setting stage and is an indication of the stability of the system.

The major motivation of this work is to develop neural networks that can use unmeasurable variables and deal with dynamic trends for the purpose of fault diagnosis. It is the nature of most of chemical processes that important operating variables associated with product quality and process performance cannot be measured directly or easily, despite the expenditure on the special measurement techniques (Yang et al., 1998). Many important variables have to be determined by off-line laboratory analysis; even on-line product quality measurements often suffer from excessive time delays and in many cases can only be done intermittently. Soft sensors are designed to exploit information from model-based procedures, which estimate system parameters to describe the state of the system. This makes it possible to compensate for relatively infrequent sampling and time delays in measurement by essentially providing real-time process information which can be used as input to fault diagnostic systems. The concern here is with developing techniques for soft sensors, which can be incorporated into fault diagnostic procedures. Graphically represented dynamic trends provide a direct way for the process operator to make decisions on the operational status of a process. However, in order to make efficient use of trends in a fault diagnostic system, a proper interpretation method is required. In this study, the wavelet transform technique is employed for identifying and representing dynamic trends through extracting from dynamic trends.

The paper is organised as follows. First, a Fluid Catalytic Cracking (FCC) process is described as a target process for this study. Second, the structure of the neural-network-based fault diagnostic system is developed, in which a soft sensor and a dynamic trend interpreter have been involved. Third, a strategy for developing suitable soft sensors is then described. Following this, a qualitative interpretation of dynamic trends and a wavelet transform are introduced to capture and identify the features of the dynamic trends. Then, the neural network for fault diagnosis is established for the FCC process. Finally, training, simulation and utilisation of the proposed method are accomplished.

2. The most serious faults in FCC reactor

The Fluid Catalytic Cracking (FCC) process is a dominant feature of most refinery operations, and is characterised by being physically and chemically complex and operating at high temperatures and pressures. It is also inherently hazardous because of the rise of potential fires and explosion arising from any malfunction. A typical FCC process as shown in Fig. 1 is selected as a target plant in this work. It includes a riser tube reactor and a regenerator and is designed to convert heavy oils into light hydrocarbons and coke. The combined feed undergoes on endothermic reaction in the riser promoted by the regenerated catalyst, which is recycled from the regenerator. The coke produced from the catalytic reaction is deposited on the spent catalyst, which is circulated to the regenerator where the coke is burned and then the catalyst is returned to the reactor. The heat generated from the combustion is used to support the cracking reactions. The outlet temperature of the riser reactor is one of the most important process variables and is controlled by manipulating the catalyst circulation rate through the slide valve. It is difficult to measure the catalyst circulation rate directly because of the high temperature and the tendency of the catalyst to ignite. The most often disturbance to the reactor operation is the change of the characteristic of feedstock, for example, feedstock implicitly contains water. Prediction of the working state of the temperature control system, the slide valve, and the disturbance occurrence from the feedstock will be highly beneficial for operators. The most serious and frequently observable faults in the operation of the FCC riser reactor are selected for this study as listed in Table 1. Three faults that affect the controller, actuator, and process are considered.

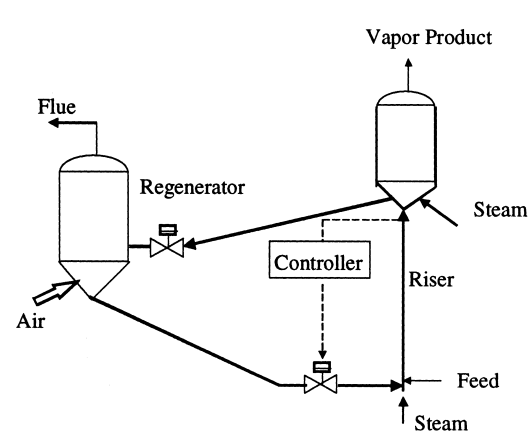

Fig. 1. The FCC reactor–regenerator system.

Table 1
List of faults studies

Fault name	Fault description	Faults class
TCSRF	Temperature control system of the riser failutre	Malfunction of controller
CSVS	Catalyst slide valve stuck	Malfunction of actuator
FCW	Feedstock containing water	Malfunction of process

3. Structure of the neural network based fault diagnostic system

The capability of neural network based methods for fault diagnosis has been established in previous work. The particular attention of this study is to address (a) using unmeasurable inputs in the neural networks, (b) efficiently using dynamic trends of process variables as inputs of the neural networks. The neural network based fault diagnostic system developed in this study has a basic structure as shown in Fig. 2.

1. *Soft sensor*: In order to employ *unmeasurable* variables in the neural network a soft sensor for the catalyst circulation rate is developed and briefly described here but detailed discussion will be given in the next section. A soft sensor can be seen as a simple local process model or an inferential measurement, which generates on-line estimations of unmeasured variables using computational models together with other on-line measurements as well as off-line laboratory tests (Ansari and Tade, 1996; Enrique and Luyben, 1992).
2. *Feature extraction in a moving time window*: A moving time window is an indispensable technique to track dynamic data. As shown in Fig. 3, the window moves forward at each time increment. The right side of each window corresponds to the current time; the time span of the window is the product of the number of samples in the window and the time increment. The window length is adjustable for each application. For this study, 100 data points are included in a single window length. There will be an input explosion if all data points in the window length are directly fed into the following neural network. Suppose 10 dynamic trends are required to act as inputs for the neural network and each dynamic trend is composed of 100 data points, then 1000 data points will be fed into the neural network for any prediction. An alternative is to feed the features of dynamic trends into the neural network instead of all raw data points. A wavelet transform and a qualitative interpretation are adopted here to extract the features of dynamic trends. Data points in a window length are categorized into six types of behaviour: increasing, decreasing, steady, unstable, mean value high, and mean value low, and the wavelet transform and the qualitative interpretation identify this type of trend information. More detail in feature extraction will be given in Section 5.
3. *Neural network for fault diagnosis*: Neural networks have generated considerable interest in the field of fault diagnosis in the past two decades because their ability of being able to approximate any continuous non-linear functions. The fundamental element of neural networks is a neuron, which has multiple inputs and a single output. Each input is multiplied by a weight, the inputs are summed and this quantity is operated on by the transfer function of the neuron to generate the output. A sigmoid function

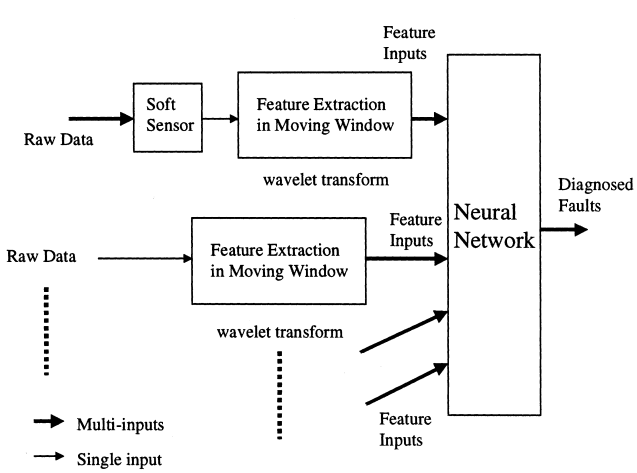

Fig. 2. Schematic diagram of fault diagnostic system.

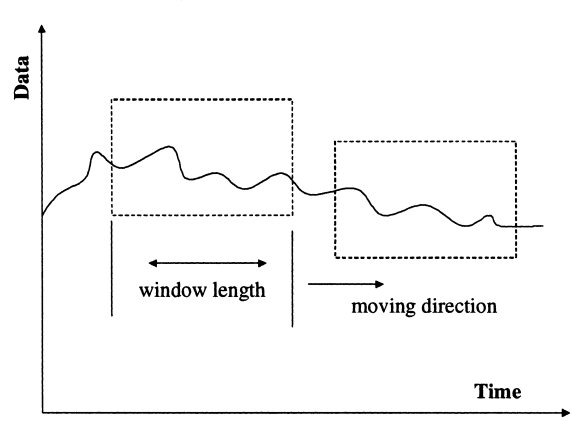

Fig. 3. Moving time window.

is usually selected as a transfer function for the neuron.

The fault diagnosis is a mapping process between a fault symptom set M and a fault source set D. The relationship between M and D can be embedded in a neural network if treating M as inputs and D as outputs in the network. In this study, D is composed of three faults listed in Table 1; M is composed of features of dynamic trends of 6 process variables. The structure with a hidden layer is selected. The back propagation algorithm is used to train the network. More detail will be discussed in Section 6.

4. Soft sensor of catalyst circulation rate

This section looks at a strategy of developing a soft sensor of the catalyst circulation rate in the FCC process. The novel feature of the approach is the development of a self-learning procedure. The structure of the self-learning procedure is shown in Fig. 4. The self-learning function is activated only if a steady state operation condition has been recognised, otherwise an online estimation is used directly.

A steady state energy balance around the regenerator is usually used in calculating the average catalyst circulation rate (Pierce, 1983). It assumes that the heat generated from burning coke be counterbalanced by the difference in sensible heat between the exit and inlet streams. The exit streams include the flue gas and the catalyst returning to the reactor as well as the heat taken out from the regenerator. The inlet streams include the combustion air and the spent catalyst from the reactor together with direct addition of steam to the regenerator. The energy balance over the regenerator can be written as following:

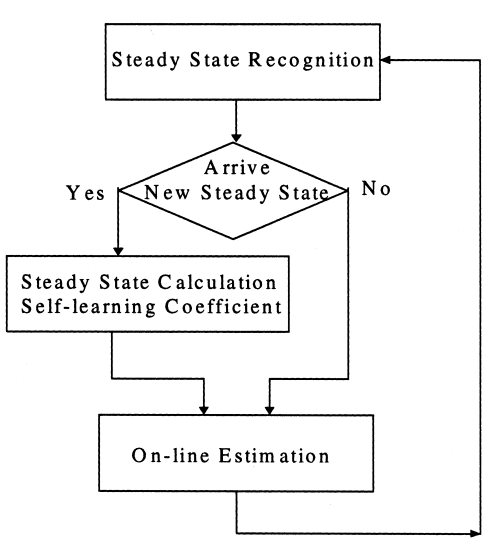

Fig. 4. Strategy of catalyst circulation rate soft sensor.

Heat generated from burning coke

$$= \text{Heat of exit streams} - \text{Heat of inlet streams}$$

Heat generated from burning coke

$$= F_a(0.21 - O_{fg})\Delta H_{cb}/\upsilon_{oc}$$

Heat of exit streams

$$= T_{fg}F_aCp_a + T_{rg}G_{rc}Cp_s + F_wHw_{fg} + Q_{loss}$$

Heat of inlet streams

$$= T_aF_aCp_a + T_{ra}G_{rc}Cp_s + F_wHw_a \tag{1}$$

Where ΔH_{cb} is the heat of combustion of the coke in the regenerator, υ_{oc} represents the volume Nm^3 of oxygen required to burn 1 kg coke. Q_{loss} is the heat lost to environment by radiation. Q_{loss} is primarily a function of the regenerator geometry and may be estimated from experience. More detail can be found in our previous work (Gu and Yang, 1995; Yang et al., 1998).

Therefore, the catalyst circulation rate in a steady state, G_{rc}, is given as following:

$$G_{rc} = \Big[F_aCp_a(T_a - T_{fg}) - F_w(Hw_{fg} - Hw_a) - Q_{loss} + F_a(0.21 - O_{fg})\Delta H_{cb}/\upsilon_{oc}\Big]/(T_{rg} - T_{ra})Cp_s \tag{2}$$

On the other hand, the catalyst circulation rate can be calculated in terms of the opening of the catalyst slide valve as following:

$$G_{rc} = \zeta \cdot A_{rcs} \cdot \sqrt{Dp_{rcs} \cdot \rho_{rc}} \tag{3}$$

where A_{rcs} represents the cross-sectional area of the slide valve. It is a segmented non-linear function of the opening without any uncertainty.

Unfortunately, because of the complexity of the catalyst flow and measurement noise associated with differential pressure Dp_{rcs} and density ρ_{rc}, of catalyst flowing through the catalyst side valve, the estimation using Eq. (3) may be incorrect after the operation condition changes. The approach here is to lump all uncertainties in Eq. (3) into the flow rate coefficient of the slide valve, ζ, which is dealt with as an adaptive parameter. The soft sensor for the catalyst circulation rate is based on the on-line estimation procedure given by Eq. (3). Once arriving a new steady state, the adaptive parameter will be updated according to the steady-state estimation of Eq. (2), following the sequence shown in Fig. 4. In practice, the on-line

learning for the flow rate coefficient of the slide valve can be based on any adaptive recursive algorithm, such as Recursive Least Squares. In the simplest case, the adaptive parameter ζ can be updated according to a first-order filter as following:

If $Dp_{\mathrm{rcs}} \cdot \rho_{\mathrm{rc}} \neq 0$

$$\zeta_{\mathrm{next}} = \zeta_{\mathrm{current}}\eta + \left(|G_{\mathrm{rc}_2} - G_{\mathrm{rc}_3}|/A_{\mathrm{rcs}}\sqrt{Dp_{\mathrm{rcs}} \cdot \rho_{\mathrm{rc}}}\right) \cdot (1-\eta) \tag{4}$$

Otherwise

$$\zeta_{\mathrm{next}} = \zeta_{\mathrm{current}},$$

where η is a filter constant less than unity.

The FCC reactor/regenerator section is usually regarded as being at a steady state if the outlet temperature of the riser T_{ra}, the regenerator dense phase temperature T_{rg}, and the catalyst hold-up in the regenerator H_{rg} are remained stable, i.e. Eqs. (5) and (6) are satisfied. Therefore, the following two equations can be used to identify the steady state in the sequence shown in Fig. 4.

$$|Y_{k+1} - EY| \leq \varepsilon, \tag{5}$$

$$E(Y_{k+1} - EY) \rightarrow 0, \tag{6}$$

here, $Y = (T_{\mathrm{ra}} \quad T_{\mathrm{rg}} \quad H_{\mathrm{rg}})^{\mathrm{T}}$, and ε is a small positive number.

5. Feature extraction from dynamic trends

5.1. Features of dynamic trend

The dynamic trends in process variables provide important clues in fault detection and diagnosis. An appropriate trend interpretation method is required in order to make an efficient use of dynamic trends in the neural network based fault diagnostic system. In this approach the dynamic trends are categorized into six types of behaviour: increasing, decreasing, steady, unstable, mean value high, and mean value low. Six values are used to interpret the trends, namely:

variable_name(increasing, x)
variable_name(decreasing, x)
variable_name(steady, x)
variable_name(unstable, x)
variable_name(mean_value_high, x)
variable_name(mean_value_low, x)

x is a measure of the degree of the specified behaviour. In the first four variable pairs only one x has a value of 1, the rest of them are 0. For example, when x in variable_name(increasing, x) is 1, then x in variable_name(decreasing, x), variable_name(steady, x), and variable_name(unstable, x) being 0. Differently, x has a value from 0 to 1 in the last two variable pairs. When x in variable_name(mean_value_high, x) is greater than 0, then this variable will result in variable_name(mean_value_low, x) being 0. With this set of values, a dynamic trend can be represented by six variables in the network for fault diagnosis.

5.2. Wavelet transform

In this work the measure of the degree of the specified behaviour, x, is calculated based on the extrema including maxima and minima of process measurements, where wavelet transform is adopted to identify such extrema. Further details can be seen in our recent work (Chen et al., 1999; Wang et al., 1999; Wang, 1999).

For a continuous function $f(t)$, extrema can be identified based on function derivation. On the other hand, extrema are often measured with Lipschitz exponents (Mallat and Hwang, 1992) in discrete signals. In practice, however, the relationship does not provide a simple and direct way of detecting extrema of a signal because evaluating the Lipschitz exponent of a signal is a time consuming procedure. In fact, wavelet transform of a signal can be linked to Lipschitz exponent so that it can be used to detect extrema of discrete signals. Particularly, the wavelet analysis can capture the time location of extrema of signals when wavelet function and filter are selected properly (Mallat and Hwang, 1992; Berman and Baras, 1993).

Here, non-subsampled multi-resolution analysis filter bank developed by Cvetkovic and Vetterli (1995) is used to analyse signals. A finite impulse response (FIR) is recommended for use in Cvetkovic and Vetterli's approach. Here, the cubic spine wavelet given by Mallat and Hwang (1992) and Mallat and Zhong (1992) is employed as basis function. The coefficients of the filters h_n and g_n are listed in Table 2, where h_n is the low pass and g_n is the high pass filter. Using this method, extrema of wavelet analysis of a signal exactly correspond to the maxima or minima of the signal with time location as illustrated in Fig. 5. Thus, these

Table 2
Filter coefficient for cubic spine wavelet and scale function

n	h_n	g_n
0	0.3750	−0.59621
1	0.2500	0.59621
2	0.0625	0.10872
3	0.0000	0.01643
4	0.0000	0.00008

points can be identified based on extrema of wavelet analysis of the signal.

The method of calculating x is illustrated as follows. Suppose there exist N extrema points in the window length. The start point and the end point are denoted by X_0 and X_{N+1}, respectively. N extrema points are denoted by $X_1, X_2, \ldots, X_N$. The first maximum point of $X_1, X_2, \ldots, X_N$ is denoted by $X_{\max}^{\text{first}}$, the last maximum point by $X_{\max}^{\text{end}}$ if available. Similarly, the first minimum point of $X_1, X_2, \ldots, X_n$ is denoted by $X_{\min}^{\text{first}}$, the last minimum point by $X_{\min}^{\text{end}}$ if available. The steady-state value of the variable is X_{ss} that is identified according to Eqs. (5) and (6) in Section 4. The average value of the start, extrema and end points in the window length is denoted as X_{avg}, X_i is the minimum point in all extrema points, X_j is the maximum point in all extrema points. Therefore, x can be calculated as the following:

$$X_{\text{avg}} = \frac{\sum X_k}{N+2} \tag{7}$$

$$X_{\min} = X_i = \min(X_0, \ldots, X_{N+1}) \tag{8}$$

$$X_{\max} = X_j = \max(X_0, \ldots, X_{N+1}) \tag{9}$$

Case 1: if $X_{\max} - X_{\min} \leq \delta$ then
$x = 1$ in variable_name(steady, x),
$x = 0$ in variable_name(increasing, x), variable_name(decreasing, x), variable_name(unstable, x);

Case 2: if $N \leq 3$ or $|X_{\max}^{\text{first}} - X_{\max}^{\text{end}}| \leq \delta$ or $|X_{\max}^{\text{first}} - X_{\min}^{\text{end}}| \leq \delta$ then
If $i < j$ then
$x = 1$ in variable_name(increasing, x),
$x = 0$ in variable_name(decreasing, x), variable_name(steady, x), variable_name(unstable, x);
Else if $i > j$ then
$x = 1$ in variable_name(decreasing, x),
$x = 0$ in variable_name(increasing, x), variable_name(steady, x), variable_name(unstable, x);

Case 3: if $X_{\max}^{\text{first}} < X_{\max}^{\text{end}}$ and $X_{\min}^{\text{first}} < X_{\min}^{\text{end}}$ then
$x = 1$ in variable_name(increasing, x),
$x = 0$ in variable_name(decreasing, x), variable_name(steady, x), variable_name(unstable, x),

Case 4: if $X_{\max}^{\text{first}} < X_{\max}^{\text{end}}$ and $X_{\min}^{\text{first}} > X_{\min}^{\text{end}}$ then
$x = 1$ in variable_name(unstable, x),
$x = 0$ in variable_name(increasing, x), variable_name(decreasing, x), variable_name(steady, x),

Case 5: if $X_{\max}^{\text{first}} > X_{\max}^{\text{end}}$ and $X_{\min}^{\text{first}} > X_{\min}^{\text{end}}$ then
$x = 1$ in variable_name(decreasing, x),
$x = 0$ in variable_name(increasing, x), variable_name(steady, x), variable_name(unstable, x),

Case 6: if $X_{\max}^{\text{first}} > X_{\max}^{\text{end}}$ and $X_{\min}^{\text{first}} < X_{\min}^{\text{end}}$ then
If $X_{\max}^{\text{first}} - X_{\max}^{\text{end}} > X_{\min}^{\text{first}} - X_{\min}^{\text{end}}$ then
$x = 1$ in variable_name(decreasing, x),
$x = 0$ in variable_name(increasing, x), variable_name(steady, x), variable_name(unstable, x);
else
$x = 1$ in variable_name(increasing, x),
$x = 0$ in variable_name(decreasing, x), variable_name(steady, x), variable_name(unstable, x);

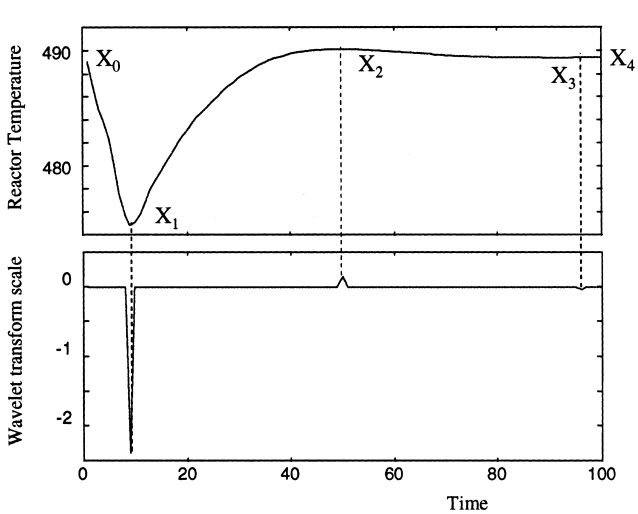

Fig. 5. Detect extrema of dynamic trend by wavelet transform

Here, δ is a small positive number that can be justified for various process variables. Case 1 means the dynamic trend is stable only if the difference between the maximum and the minimum of the extrema points is less than δ. In Case 2, the number of the extrema points is less than 4, or the value of the first maximum point is close to the value of the last maximum point in the window length, or the value of the first minimum point is close to the value of the last minimum point in the window length. In this case, the direction of the trend is simply derived by comparing the values of the start point and the end point. In Cases 3–6, the first and the last maximum points and minimum points

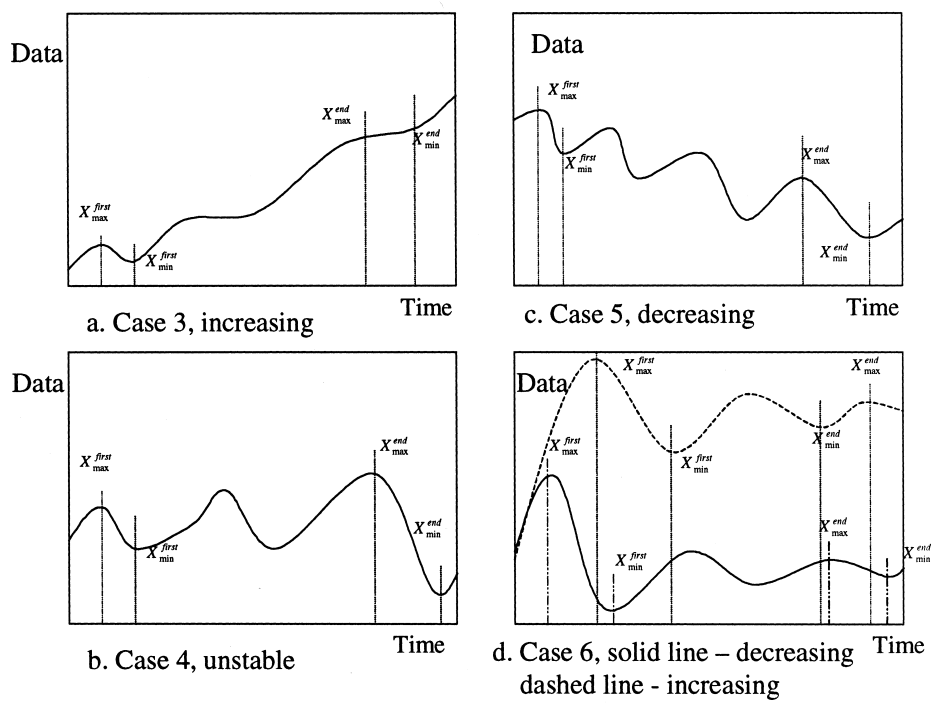

Fig. 6. Examples of dynamic trends.

in the dynamic trend are used to achieve the values of x. Fig. 6 shows the examples for Cases 3–6.

In the last two variable pairs x has a value from 0 to 1. Giving a maximum deviation from the steady state $DX_{\max}$, then x value can be calculated as the following:

If $X_{\text{avg}} \geq X_{\text{ss}}$ then
$x = \min(1, |X_{\text{avg}} - X_{\text{ss}}|/DX_{\max})$ in variable_name(mean_value_high, x),
$x = 0$ in variable_name(mean_value_high, x),

else
$x = 0$ in variable_name(mean_value_low, x),
$x = 0$ in variable_name(mean_value_high, x),
$x = 0$, $x = \min(1, |X_{\text{avg}} - X_{\text{ss}}|/DX_{\max})$ in variable_name(mean_value_low, x).

6. Neural network for fault diagnosis

Fig. 7 depicts the basic structure of the neural network for fault diagnosis for the FCC process developed in this work. A feedforward type network is used. Features of dynamic trends of six process variables are used in the input layer. The number of units in the input layer is 36. Three faults are used in the output layer. The number of units in the output layer is 3. One hidden layer is selected here. The number of hidden units is adjustable and after some trial and error during the learning phase it is chosen at 20. Process variables in the input layer are listed in Table 3.

The dynamic trend of each process variable is expressed using six feature variable pairs proposed in Section 5. The value of $G_{\text{rc}}$is generated by the soft sensor of the catalyst circulation rate and is then used as one of input dynamic trends as shown in the dashed line box in Fig. 7. It is obvious that if there is not soft sensor information, the network shown in Fig. 7. will need to be introduced other more process variables and it will make the model complex and inefficient, even impractical. From this point of view the network with a soft sensor could be considered as a kind of hybrid model that integrates an inference model for a soft sensor with a neural network model.

Table 3
List of input variables of the neutral network

Input name	Description
G_{rc}	Catalyst circulation rate
T_{ra}	Temperature in the reactor
T_{rg}	Temperature in the regenerator
P_{ra}	Pressure in the reactor
T_{f}	Temperature of the feed of the reactor
F_{f}	Feed flow rate of the reactor

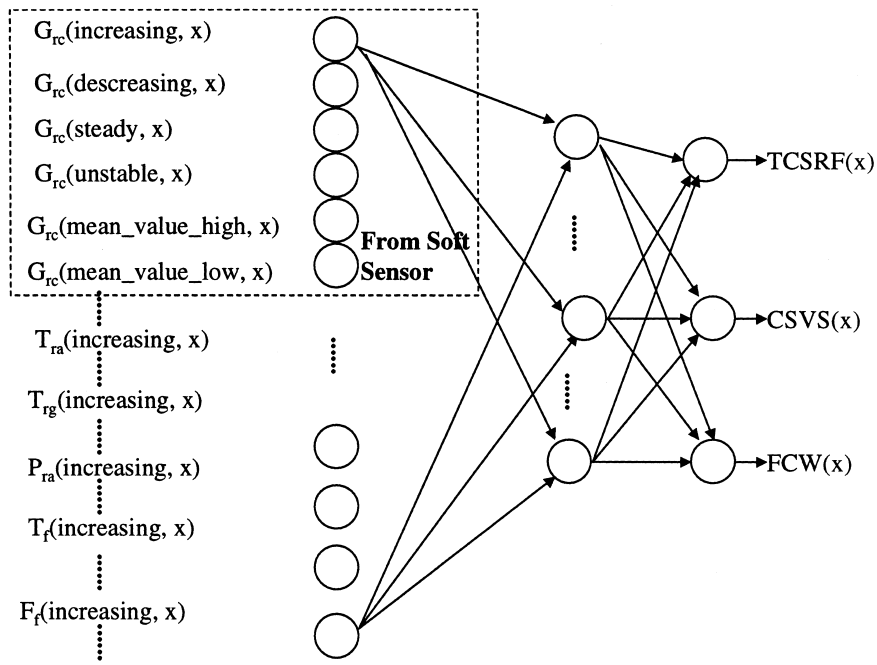

Fig. 7. The neural network structure.

7. Simulation results

7.1. Training and testing data

Training and testing data are generated from a dynamic training simulator for the FCC process. The structure of the simulator is shown in Fig. 8. It comprises a network of three personnel computers. The process dynamic simulation software is run on the instruction station that can be viewed as the emulation of the real plant dynamics. The operator station enables the complete simulation of an industrial control system including a special keyboard and its operational functions for process engineers. Operators can be trained on the operator station of start-up, shut down, normal operation, and fault detection and diagnosis as well as emergency treatment. The field operations such as manipulation of a hand valve or a pump are also simulated on the operator station. Faults can be deliberately initiated on both the operator station and the instruction station, and the corresponding process responses can be observed on the operator station and be received by the fault diagnostic system.

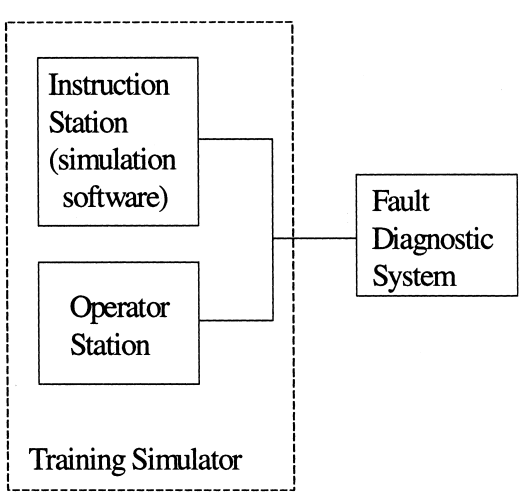

Fig. 8. The structure of the training simulator based fault diagnostic system

The dynamic simulator used for developing fault diagnostic system has distinctive features compared with traditional differential equation based dynamic simulators. The latter are not usually able to simulate the process behaviour under abnormal conditions. In addition, many differential equation based simulators are designed only for open loop conditions. Also, it is important that the simulator used for developing fault diagnostic system should be able to handle the detailed plant structure, including all utility and purge streams. The simulator is able to carry out the continuous simulation of start-up, normal operation and shutdown procedures. More importantly, it can simulate not only malfunctions caused by equipment failure, such as catalyst side valve stuck (CSVS), but interruptions in electricity and steam supply as well. It also allows for changes in properties of the feedstock such as density, feedstock-containing water (FCW).

Training and testing data are generated by the simu-

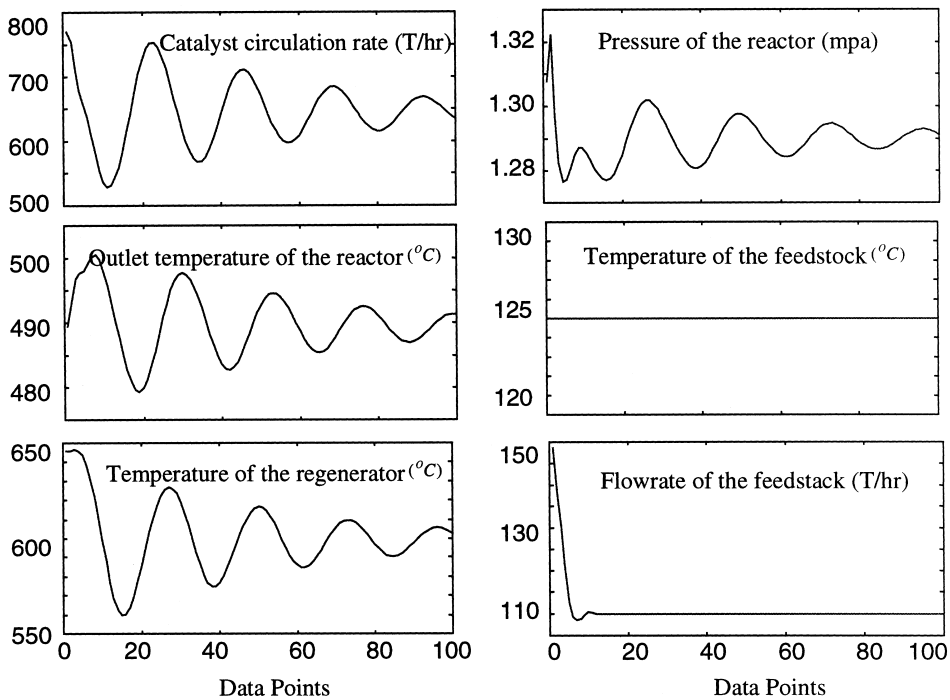

Fig. 9. Dynamic response of the FCC process under step change of 50 T/h in the set point of flow rate of the feedstock.

lator and fed into the fault diagnostic system. In order to generate patterns for normal operations various changes have been introduced into the simulator. Fig. 9 illustrates the dynamic responses of the FCC process under the step change of 50 T/h in the setpoint of the flow rate controller of the feedstock. After the step change in the setpoint of the flow rate controller of the feedstock, the catalyst circulation rate, the pressure of the reactor, and the temperature of the regenerator are going down. The catalyst circulation rate is estimated by the soft sensor developed in Section 4. The temperature of the feedstock is independent on the feedstock flow rate and keeps unchanged. The feedstock flow rate is decreased to 100 T/h. The outlet temperature of the reactor is significantly disturbed by the change of the feedstock flowrate, but still converges to the original value due to the action of the temperature controller. Training patterns for this normal operation is achieved by implementing the method proposed in Section 5 and is shown in Table 4. Similarly, training patterns for the other normal operations and three faults listed in Table 1 can be gained and used for training the network.

7.2. Training method

Even though there exist many different training methods, only the backpropagation (BP) algorithm is selected to train the above network. Comparisons with other training methods will not be discussed here since selecting a suitable training method is not the motivation of this work. The connection weights in the network are adjusted so that the average squared error between the network output and the desired output for a set of given reference inputs is minimised. Learning continues iteratively until the sum of the squared error is below a certain goal. The neural network toolbox in

Table 4
Training patterns for the step change in the setpoint of the flow rate controller of the feedstock

Dynamic trend	Increasing	Decreasing	Steady	Unstable	mean_value_high	mean_value_low	TCSRF	CSVS	FCW
G_{rc}	0	1	0	0	0	0.4	0	0	0
T_{ra}	0	0	1	0	0	0			
T_{rg}	0	1	0	0	0	0.5			
P_{ra}	0	1	0	0	0	0.3			
T_f	0	0	1	0	0	0			
F_f	0	1	0	0	0	0.6			

Table 5
Prediction of fault patterns

Specified faults in dynamic simulator	Recall or prediction of probability for TCSRF occurrence	Recall or prediction of probability for CSVS occurrence	Recall or prediction of probability for FCW occurrence
TCSRF (recall)	**0.9998**	0.0001	0.0004
CSVS (recall)	0.0005	**0.9991**	0.0002
FCW (recall)	0.0001	0.0002	**0.9976**
TCSRF + CSVS (prediction)	**0.9018**	**0.9210**	0.0950
TCSRF + FCW (prediction)	**0.9120**	0.1002	**0.9289**
CSVS + FCW (prediction)	0.0250	**0.9029**	**0.9489**
TCSRF + CSVS + FCW (prediction)	**0.8421**	**0.8238**	**0.8940**

MATLAB is used to carry out the network training in this work (Demuth Beale, 1993).

7.3. Simulation results

The proposed fault diagnostic system is expected to not only recall pretrained single faults correctly, but also to predict multiple faults without any further training. It is important to train the network by presenting the target pattern of a single fault and the patterns for the normal operations. The reason is that if the network is trained only using the target pattern of a single fault, then only a single neuron in the output layer is fired for any input (Maki and Loparo, 1997), therefore, the trained network will be hard to predict multiple faults.

The error goal of the backpropagation algorithm is set at 1×10^{-4} for all training experiments. Recall results for single fault and predicted results for double and triple faults are illustrated in Table 5. The first column displays the faults set on the dynamic simulator. The rest columns correspond to the predicted possibilities of these faults by the fault diagnostic system. The faults identified have been highlighted in Table 5 which are the same as those set on the dynamic simulator. It proves that the fault diagnostic system gives good recall results for single faults and good predicted results for double and triple faults.

8. Conclusions

The design of a neural network based fault diagnostic system for an industrial catalytic cracking riser reactor has been described. A distinguished feature of the system is that it uses an unmeasurable process variable as input to the network. Development of the system essentially involves three steps: soft sensor design for the unmeasurable variable, extraction of features from dynamic trends using qualitative interpretation and wavelet transform and construction and training of a neural network.

Unmeasured operating variables exist in most chemical processes and may play a key role in operation and control. Introducing soft sensors into fault diagnostic system is a promising way to make the fault diagnostic system efficient. The dynamic trends of process variables provide important clues in fault detection and diagnosis. The appropriate trend interpretation and feature extraction method makes it possible to invoke the dynamic trend information in the fault diagnostic system. A qualitative interpretation approach is combined with wavelet transform and used for this purpose. Neural network is used in this study because it has proven capability in mapping a fault symptom to a fault source set. The training data is generated by setting a number of malfunctions on a dynamic training simulator. It can be also collected from a real plant if possible. It is obvious that using the dynamic training simulator for generating the training data is more practical and economic compared with collecting data from a real plant. Even though the neural network is trained using single fault patterns the results have shown that it can also predict multiple faults effectively.

References

Ansari, R.M., Tade, M.O., 1996. On-line inferential models in nonlinear multivariable control applications. Developments in Chemical Engineering and Mineral Processing 4 (1/2), 71–83.

Basseville, M., 1998. On-board component fault detection and isolation using the statistical local approach. Automatica 34 (11), 1391–1415.

Berman, Z., Baras, J.S., 1993. Properties of the multiscale-maxima and zero-crossings representations. IEEE Trans. Signal Processin 41, 3216–3231.

Chen, B.H., Wang, X.Z., Yang, S.H., McGreavy, C., 1999. Application of wavelets and neural networks to diagnostic system

development, 1, feature extraction. Computer & Chemical Engineering 23, 899–906.

Chen, Chieh-Li, Jong, Ming-Jyi, 1993. Fuzzy predictive control for the time-delay system. In: Proc. 2nd IEEE Int. Conf. Fuzzy Systems, 236–240.

Cvetkovic, Z., Vetterli, M., 1995. Discrete-time wavelet extrema representation: design and consistent reconstruction. IEEE Trans. Signal Processing 43, 681–693.

Demuth, H., Beal, M., 1993. Neural Network Tool Box User's Guide. The Math Works, Inc, MA, USA.

Enrique, Q.M., Luyben, W.L., 1992. Inferential model-based control of multicomponent batch distillation. Chemical Engineering Science 47 (4), 887–898.

Gu, Z.W., Yang, S.H., 1995. Modelling of Industrial Processes. Zhejiang University Press, Zhejiang, People's Republic of China, pp. 38–54.

Hoskins, J.C., Himmelblau, D.M., 1988. Artificial neural-network models of knowledge representation in chemical engineering. Computers Chem. Eng 12 (9/10), 881–890.

Hoskins, J.C., Kaliyur, K.M., Himmelblau, D.M., 1991. Fault diagnosis in complex chemical plants using artificial neural networks. AIChE J 37 (1), 137–141.

Maki, Y., Loparo, K., 1997. A neural-network approach to fault detection and diagnosis in industrial processes. IEEE Transaction on Control Systems Technology 5 (6), 529–541.

Mallat, S., Hwang, W.H., 1992. Singularity detection and processing with wavelet. IEEE Trans. on Information Theory 38 (2), 617–643.

Mallat, S., Zhong, S., 1992. Characterization of signals from multiscale edges. IEEE Trans. Pattern Anal. Machine Intell 14, 710–732.

Meghlaoui, A., Thibault, J., Bui, R.T., Tikasz, L., Santerre, R., 1998. Neural networks for the identification of the aluminium electrolysis process. Computers Chem. Eng 22 (10), 1419–1428.

Ohga, Y., Seki, H., 1993. Abnormal event identification in nuclear power plants using a neural network and knowledge processing. Nuclear Techno 101, 159–167.

Pierce, E., 1983. Calculating FCC Heat Balance. Hydrocarbon Processing. 39–42.

Vemuri, A.T., Polycarpou, M.M., 1997a. Robust non-linear fault diagnosis in input-output systems. International Journal of Control 68 (2), 343–360.

Vemuri, A.T., Polycarpou, M.M., 1997b. Neural-network-based robust fault diagnosis in robotic systems. IEEE Transaction on Neural Networks 8 (6), 1410–1420.

Venkatasubramanian, V., Vaidyanathan, R., Yamamoto, Y., 1990. Process fault detection and diagnosis using neural networks. Part I: steady State processes. Computers Chem. Eng 14 (7), 699–712.

Venkatasubramanian, V., Chan, K., 1989. A neural network methodology for process fault diagnosis. AIChE J 35 (12), 1993–2002.

Wang, X.Z., 1999. Data Mining and Knowledge Discovery for Process Monitoring and Control. Springer, London.

Wang, X.Z., Chen, B.H., Yang, S.H., McGreavy, C., 1996. Neural nets. fuzzy sets and digraphs in safety and operability studies of refinery reaction processes, Chemical Engineering Science 51 (10), 2169–2178.

Wang, X.Z., Chen, B.H., Yang, S.H., McGreavy, C., 1999. Application of wavelets and neural networks to diagnostic system development. Part 2: An integrated framework and its application. Comput. Chem. Engng 23, 945–954.

Wang, X.Z., Lu, M.L., McGreavy, C., 1997. Learning dynamic fault models based on a fuzzy set covering method. Computer and Chemical Engineering 21 (6), 621–630.

Wennersten, R., Narfeldt, R., Granfors, A., Sjokvist, S., 1996. Process modelling in fault diagnosis. Computers & Chemical Engineering 20 (S), s665–s670.

Yang, S.H., Wang, X.Z., McGreavy, C., 1995. Machine learning fault model of industrial R-FCC processes. Pre-Prints of the IFAC Detection and Supervision in the Chemical Process Industries, Newcastle, June, 174–179.

Yang, S.H., Chen, B.H., Wang, X.Z., McGreavy, C., 1997. Soft sensor based fault diagnosis of industrial fluid catalytic craking reactor. IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, Hull, UK, August, pp. 240–245.

Yang, S.H., Wang, X.Z., McGeravy, C., Chen, Q.H., 1998. Soft sensor based predictive control of industrial fluid catalytic cracking processes. Chemical Engineering Research and Design, Trans. IChemE 76, 499–508.

Shuanghua Yang is a lecturer in the Computer Science Department at Loughborough University. He received his BSc and MSc in Control Engineering from Petroleum University in China in 1983, and 1986, and PhD, in Control Engineering from Zhejiang University in China in 1991.

Binghui Chen was a research associate at Loughborough University but is now currently working in the Department of Biochemical Engineering at University College of London. He received his BSc in Chemical Engineering from Huaquiao University in China in 1984, MSc in Chemical Engineering from Zhejiang University in China in 1991, and PhD in Chemical Engineering from Leeds University in 1999.

XueZhong Wang is a lecturer in the Chemical Engineering Department at Leeds University. He received his BSc, MSc and PhD in Chemical Engineering from Petroleum University in China in 1984, 1987, and 1990, respectively.