

Data Driven Fault Diagnosis and Fault Tolerant Control: Some Advances and Possible New Directions

WANG Hong^{1,2}CHAI Tian-You²DING Jin-Liang²BROWN Martin¹

Abstract This paper presents a selected survey covering the advances of fault diagnosis and fault tolerant control using data driven techniques. A brief summary of the general developments in fault detection and diagnosis for industrial processes is given, which is then followed by discussions on the widely used data driven and knowledge-based techniques. A successful application example is also given, which deals with faults caused by the misplacement of control loop set points and several areas of potential future directions are included in the paper.

Key words Fault detection, fault diagnosis, fault tolerant control, data driven techniques

Fault detection and diagnosis (FDD) and fault tolerant control (FTC) have been the subject of considerable interest in the control research community^[1–39]. This is in response to the ever increasing requirements on the reliable operation of control systems, which are, in most cases, subject to a number of faults either in the internal closed loops or from environmental factors. Once system faults have occurred, they can cause unrecoverable losses and result in unacceptable environmental pollution, etc. Occasionally, the occurrence of a minor fault has resulted in disastrous effects. For example, it has been observed that faults have caused a 3% ~ 8% reduction of the oil production in the United States, leading to \$20 billion losses in the country's economy per year. Also, in 1997, the faults in a chemical plant in Beijing caused heavy direct losses. Therefore, effective FDD is of vital importance to the safe operation of industrial plants. Indeed, FDD and FTC have now become an integral part of industrial process control.

In general, system faults can be grouped into several categories, namely, actuator faults, sensor faults, system faults and also abnormal operating faults caused by either the misplacement of control loop set points or unexpected variations in the raw materials to be processed. The purpose of FDD is to use available signals to detect, identify, and isolate possible sensor faults, actuator faults, and system faults. Conversely, FTC calculates the required actions (either controller modification or reconfiguration) so that the system can still continue to operate safely even under faulty conditions^[2–3,40]. In terms of condition monitoring or FDD, the existing methods can also be grouped into the following two categories:

- 1) Model based FDD;
- 2) Data driven FDD including knowledge based FDD.

In the early days (1980's onwards), model based FDD constituted the main stream of research, and a number of techniques were developed. Depending on whether the system model can be represented as either a state space model or an input-output model, FDD can be classified into two groups: observer based FDD^[1] and system identification based FDD^[4]. Also, to combine the best features of these two approaches, there is another group of FDD methods called adaptive observer based fault diagnosis, which uses parameter tuning principles from model reference adap-

tive control to directly estimate fault parameters online so that effective fault diagnosis can be achieved^[5–6]. However, model based FDD uses mathematical system models to estimate the system state and parameters, and in general these methods can only be applied to low dimensional systems.

Alternatively, data driven based FDD can deal with high dimensional data, and data dimension reduction techniques are generally used to highlight important information in the data volume^[7–14]. However, there are many important challenges in its use in FDD, when the system is time-varying and highly nonlinear. Knowledge based FDD uses a composition of a knowledge base (such as process input and output variables, abnormal process models, fault characteristics, operational constraints and assessment criteria) and a set of the qualitative models of the system to perform the required FDD^[15–21]. Typical techniques are cause-effect analysis based FDD, fault feature tree analysis based FDD, rule based and case based reasoning for FDD^[3].

After a fault has been detected and diagnosed, FTC can subsequently be used to guarantee the safe system operation and to prepare for an economical plant repair shutdown. FTC approaches can be generally classified as either passive or active FTC. The former uses the results of FDD to adjust some parameters of the controller in a similar manner as the well known adaptive controllers, such as self-tuning control, whilst the latter uses the FDD results to reconfigure the controller. A detailed survey on FTC was published in 2005^[2].

1 Challenges as a result of growing system complexity

With the ever increasing complexity of industrial systems, distributed control systems (DCS) have been widely used to realize whole plant monitoring and control (see Fig. 1). Typical examples are the processes seen in steel making, car manufacturing, material processing, paper-making, chemical plant and mineral processes, etc, where these production processes are controlled via a multi-layer computer network. For the system infrastructure shown in Fig. 1, the lower levels of computers (microprocessors and PLCs) are used to directly control the individual process units on the production line, whilst the higher level computers are used to manage the overall system operation and to produce required control loop set points based on production planning and scheduling results. DCS can therefore provide a platform for the global management and optimization of the whole production line to achieve an optimal operation in terms of improved product quality, high

Received January 8, 2009; in revised form March 12, 2009
Supported by National Basic Research Program of China (973 Program) (2009CB320600), National Natural Science Foundation of China (60828007, 60534010, 60821063), the Leverhulme Trust (F/00 120/BC) in the United Kingdom, and the 111 Project (B08015)

1. Control Systems Centre, University of Manchester, Manchester M60 1QD, UK 2. Key Laboratory of Integrated Automation for Process Industry, Ministry of Education, Northeastern University Shenyang 110004, P. R. China

DOI: 10.1016/S1874-1029(08)60093-2

production efficiency, minimized effluent discharge, and reduced energy costs, etc.

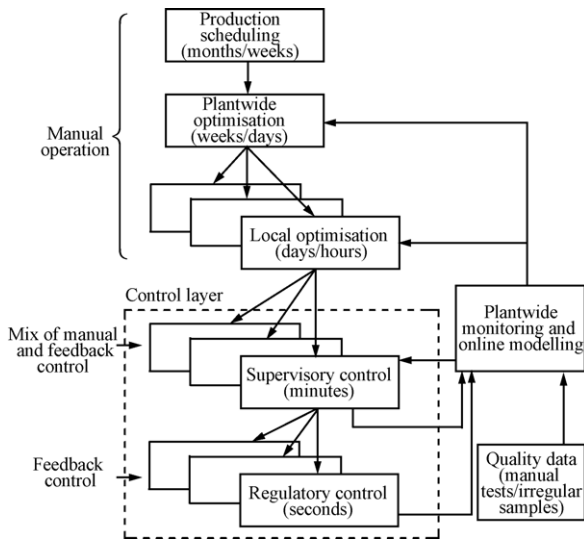


Fig. 1 DCS based plant-wide performance optimization

The advantage of using DCS systems is that a large number of process and product quality data is available for the reliable and optimal operation of the whole production line. As complex industrial processes cannot be easily represented using accurate or first principles mathematical models, data driven based FDD and FTC must be considered. This means that FDD and FTC of complex industrial systems must effectively use process data from the production line to identify faults and then use FTC techniques to realize safe operation of the whole plant until an economic shutdown can be achieved. It can therefore be expected that data driven based FDD and FTC techniques will play an increasingly important role in improving the reliable operation of complex industrial systems and data driven based FDD and FTC has already been applied in many processes^[20–31]. In this paper, the following sections will give a brief survey on these techniques.

With ever increasing automation in complex industrial systems, there exists a greater chance of unstable system operation which presents new challenges for FDD. Specifically, as there are many uncertainties in the process and the complexity of the process includes nonlinearities, time-varying, and strong coupling effects amongst the variables, etc, it is generally difficult to establish a plant-wide system model and hence perform model based FDD. Moreover, in large scale industrial systems, network based sensors measure process variables and operational variables such as control loop set points. This data provides the required information for on-site operators to improve product quality, production safety, and process efficiency. However, such large data volumes present difficulties in building process models whilst they are very suited for data dimension reduction and the related FDD.

To perform effective FDD and FTC, in line with the DCS controlled system structure, Leung proposed to divide the production line into three layers: lower, middle, and high layers^[15]. In the lower layers, tasks such as data acquisition, loop control, and signal processing is performed. In the middle layer, data analysis, process monitoring, and FDD takes place, whilst in the high layer planning and scheduling are carried out. As such, a fault can be defined

as one or more parameter deviations from its, or their, nominal values. This constitutes a strict request on the middle and high level operations of the concerned production line.

At present, FDD units embedded in DCS and SCADA systems make use of single variable monitoring techniques. The FDD methods can be represented as classification based FDD, knowledge based system (KBS), contribution diagram, qualitative intelligent analysis such as fault tree analysis, as well as rule based, knowledge based FDD^[15] and case based reasoning^[3]. Following the FDD process, FTC can be realized using system redundancy in terms of control structure configuration^[2].

Condition monitoring, in fact, performs the same task as FDD, albeit some papers use condition monitoring and others use FDD. In this paper, we will not deliberately draw a clear line between them as their purposes are similar — finding out faults in the system. Condition monitoring uses mean and variance statistics of important process variables, or their magnitude and frequencies as the basis. This includes multivariable statistics and signal processing based techniques etc. In terms of the data, nominal operation data form the basis of data driven FDD.

2 Data driven FDD – from signal based FDD to multi-variable statistics and knowledge based FDD

Data driven FDD has gone through three main phases in its development. These three phases are referred to as signal based FDD, multi-variable statistics based FDD, and knowledge based FDD. The common feature of these methods is that they all use raw system data and process knowledge to carry out the required FDD.

2.1 Signal based FDD

The first group of data driven FDD methods is signal based. Signal based FDD methods use signal processing methods consisting of correlation functions, signal model identification, signal parity checks, and spectral analysis using fast Fourier transformation and wavelet transformation. This is similar to the signal detection and trend detection for important variables using the available data. The key idea is that unexpected changes in the magnitude, phase shift and/or frequencies of the important signals can be regarded as the faults in the system. For example, in [7] wavelet transformations have been used for FDD in steel mills. In addition, statistical process control (SPC) has been applied to detect abnormal distribution changes of quality data in many production lines so that real-time alarms are produced when the data lies outside the upper and lower distribution limits. In SPC, Shewhart and Cusum charts have been widely applied to check whether important variables can be declared as normal or not. However, since SPC entirely relies on the data of the process quality, they cannot be used to detect the abnormal statistical distribution of the quality data. For example, if the quality data is not Gaussian, SPC does not produce a reliable test for FDD.

2.2 Multi-variable statistics based FDD

The key concept in principal component analysis (PCA) is to reduce a high dimensional data volume into a lower dimensional space, where the low dimensional data contains most of the useful information/variance contained in the original data set. The projection axes are referred to as principal components. As such, PCA has been widely used in industrial process control as a stan-

dard technique for data analysis and process abnormality identification^[8–9,13,22–23]. In terms of fault detection, a set of PCA components should be determined for the healthy data set and then fault detection can be performed by checking whether or not the new incoming data lies in the space spanned by the healthy principal components. PCA divides the whole observable space into a principal component subspace and a residual subspace, and then performs the FDD using Q-test and Hotelling T^2 test. In this context, the statistics (SPE) used for FDD is given by:

$$SPE = \|\tilde{\mathbf{x}}\|^2 \leq \delta_\alpha^2 \quad (1)$$

$$T^2 = \mathbf{x}^T P \Lambda^{-1} P^T \mathbf{x} \leq \chi_\beta^2 \quad (2)$$

where \mathbf{x} is the new data row, $\tilde{\mathbf{x}}$ is the residual and P is the matrix of healthy principle components and Λ is the (diagonal) matrix of eigenvalues. The PCA based FDD process is therefore given by:

Step 1. Data pre-processing in terms of scaling and filtering, etc.

Step 2. Establish the PCA $\{P, \Lambda\}$ of healthy data set and obtain the thresholds of the statistics SPE and Hotelling T^2 .

Step 3. For a new exemplar \mathbf{x} , check whether the two statistics are over their thresholds in (1) and (2). If the answer is “yes”, then it can be concluded that the system has a fault. Otherwise, the system is healthy.

Step 4. When a fault is detected, fault diagnosis can then be performed.

It can therefore be seen that PCA has the advantage of being able to detect process faults in large multivariate data sets. However, since PCA does not use direct input-output relationships, it is more suitable for fault detection rather than diagnosis. Moreover, PCA assumes that the data is (multivariate) Gaussian distributed, which limits its application in complex industrial systems that exhibit a time-varying, non-Gaussian and nonlinear characteristics. Therefore, PCA is not suited for dynamic systems with high nonlinearities and in general a high false alarm rate can occur.

To overcome this difficulty, modifications to standard PCA based FDD have been made and dynamic PCA techniques (such as moving PCAs and nonlinear PCAs using neural networks) have been developed and also applied to some industrial processes. For example, Ku proposed a modified PCA^[24], where the PCA was applied to a time-delayed data set and the linear relationships amongst important variables were retrieved. The key idea was to embed the time delays into the data base so as to compensate for the time-variations during the PCA phase. A recursive PCA was proposed in [10], where the data matrix was updated constantly so as to cope with the time-varying nature of the system dynamics. There are also modified PCA algorithms for batch processes such as multi-block PCA techniques. In terms of nonlinear systems, PCA can be extended using neural network and kernel based techniques^[25]. For example, in the kernel based PCAs, a nonlinear mapping is used to project the input data into a high dimensional feature space, then standard linear PCA can be applied directly to this high dimensional space data to produce the required fault diagnosis results. Moreover, to cope with the non-Gaussian nature of the data volume, minimum entropy based PCA has also been developed where the principal components are obtained so that the residual of the recovery error contains the minimum uncertainty^[41].

Alternatively, independent component analysis (ICA) divides the observed data into a linear combination of independent components^[11]. This allows the use of statistics charts to perform the required on-line monitoring and fault diagnosis. An interesting observation is that ICA does not require the data to be drawn from a Gaussian distribution, and it is therefore more applicable to industrial processes, where typically the data set does not obey a Gaussian distribution. A dynamic independent component analysis (DICA) technique was proposed in [12], where the ICA has been used to isolate the non-Gaussian part of the observed data set^[11] and the time delayed variables are used to extend the data matrices so that mutually independent variables can be easily identified. However, this DICA technique is not well suited for dealing with slowly time-varying faults, which are caused by equipment wear and tear, and false alarms occur when healthy process models are used to diagnose the faults.

For process data where input-output relationships can be identified, partial least squares (PLS) can be used as part of online monitoring techniques. For example, PLS uses the input and output matrices to locate the optimal vector-orientation based on predictability from the input to output space. Indeed, PLS is a modeling technique for input and output data and as such, the PLS based fault diagnosis works using a contribution of input variables to predict the quality data in production lines. There are also a number of industrial processes which are regarded as batch processes, where the data volume enters the DCS in a batch-by-batch way, and traditional continuous-time based techniques cannot be directly used to perform the required FDD and also FTC. In this case, there are three approaches, namely, the multi-way PCA, ICA, and PLS that can be used^[19,42–43].

2.3 Knowledge based FDD as one of active aspects for data driven FDD

Using continuously accumulated process knowledge, knowledge based FDD can be applied to perform online monitoring for industrial processes^[15]. Since the process knowledge is obtained from operational system data, knowledge based FDD can also be regarded as data driven FDD. Knowledge based FDD includes cause-effect analysis, expert systems and classification, where the cause-effect analysis uses the model, such as symbolic graphs or fault trees, of the fault to obtain the required fault diagnosis. Expert systems, however, attempt to mimic the reasoning phase of human experts to carry out FDD tasks and the classification phase uses the relationship between the data and the faults to perform the required fault diagnosis. Moreover, to make best use of the gained knowledge, combinations between fuzzy logic and neural networks have been used for knowledge based FDD^[32,44–46]. For example, one can use fuzzy logic to analyze the residual signals for fault diagnosis^[46], and the fault diagnosis threshold can also be made adaptive in response to process nominal changes^[32]. Moreover, neural network based FDD^[44–45] uses the available mapping between the process variables and the faults to identify the system faults. Neural networks have also been used to carry out residual signal analysis and classify the residual signals into healthy and unhealthy categories.

As discussed above, multi-variable statistic based FDD uses the data to obtain the information on the system faults and abnormal operation. However, fault diagnosis is generally difficult to perform as the method itself does not make use of available knowledge of the system. As such,

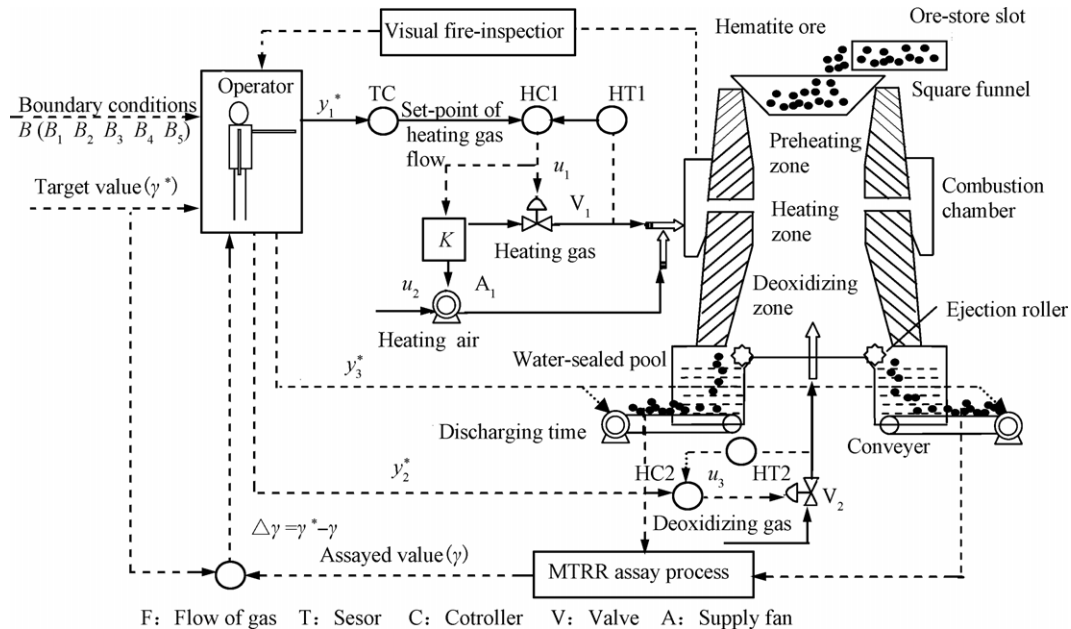


Fig. 2 Technical process and control of the shaft furnace

FDD should also involve the use of the process knowledge. This constitutes a group of methods referred to as combined data driven and knowledge based FDD^[15–17, 33]. An example is the well-known KBS technique, which uses historical operational data and understanding and the rules of the process to perform the required FDD. This technique is therefore suited to industrial processes where dynamic models are difficult to establish. However, since the accuracy and the sensitivity to the faults are not high, KBS is only suitable for simple variable detection. As such, by combining multi-variable statistics with KBS, this can improve the accuracy of the fault diagnosis process. The basic idea is to use multi-variable statistics to realize the process monitoring and use KBS to explain the statistic data produced by the multi-variable statistics. This procedure is therefore given by the following steps:

Step 1. Use multi-variable statistics techniques to invoke the knowledge base.

Step 2. Apply the contribution diagram to identify the fault.

Step 3. Use knowledge from KBS, process behavior and process operational knowledge to estimate the faults and then produce the required actions.

However, this method is largely based on expert rules and qualitative models of the system, where the statistics of the raw data is transferred into simple information for the fault diagnosis. Further development is therefore necessary so as to enlarge the dimension of the knowledge base and the raw data^[15–20].

2.4 Applications of multi-statistics based FDD

The application of data driven FDD has been successfully achieved in a number of different industries such as chemical plants, drug production, and steel mills, etc. In chemical engineering applications, data driven FDD has been applied to crystallization^[21], fermentation processes, continuous stirred tank reactor^[22, 24, 26] and waste water treatment plant^[27]. Successful applications have also been made to the benchmark Eastman processes^[9, 12, 28]. Moreover, data driven FDD has also been successfully ap-

plied to semi-conductor production^[23], steel industry, drug production^[29–31] and roughing mill^[14, 34], etc.

3 Abnormality caused by improper control loop set points: an example system

As previously described, abnormal operational conditions can also be regarded as faults in industrial processes. One of the issues related to process abnormality is the determination of control loop set points supplied by the top level to the bottom level of the DCS systems as shown in Fig. 1, where inappropriate control loop set points can also lead to faults. This is particularly true for systems in which human operators are used to determine the control loop set points. As such, it is important to develop novel FDD and FTC techniques so that the faults caused by the improper selection by on-site operators of control loop set points can be diagnosed and prevented by the relevant FTC techniques. In this section, a successful application of such a technique, namely the FTC of a shaft furnace roasting system^[3], will be described.

3.1 Shaft furnace roasting process and its characteristics

In the shaft furnace roasting system, the purpose is to effectively transfer the weak-magnetic low-grade ore into strong-magnetic ore^[3]. A shaft furnace structure is shown in Fig. 2, where the process consists of the ore feeding, ore preheating, deoxidizing, cooling and discharge phases. In the ore feeding phase, the raw hematite ore is dropped into the furnace through an ore-store slot and a square funnel at the top of the shaft furnace. Once the ore are fed into the furnace, they will go through the preheating phase, where the ore is preheated to reach a temperature of between 100 °C and 150 °C through their contacts with the ascending hot gas. Ore heating is an important phase in the operation of the shaft furnace where two combustion chambers heat the ores to the required temperature range of 700 °C ~ 850 °C. This is achieved using the heat

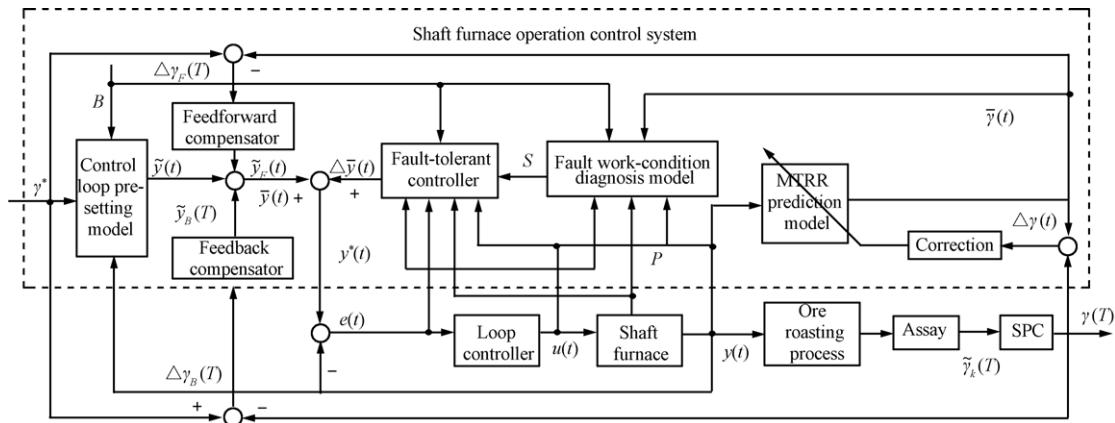


Fig. 3 Operation control strategy for the roasting process of shaft furnace

produced by the inflammation of air-mixed heating gas in these combustion chambers, which take the heating gas as the input. The control of the heating gas flow is achieved by a temperature control system whose set point y_1^* in Fig. 2 should be correctly selected. The deoxidizing phase is used to produce a chemical reaction such that the hot low-magnetic ore from the heating zone are deoxidized to high magnetic ore. This chemical reaction is partly defined by the flow rate of the deoxidizing gas, which is controlled via a local control loop around the valve denoted by V_2 . The set point y_2^* to this local loop is another important variable that needs careful selection. Cooling and discharge are the final stages, where the ore are cooled by pouring them into a water-sealed pool via two ore-ejection rollers. Then, the ore are discharged out of the furnace by two belt-conveyer machines that run synchronously with their corresponding ejection rollers. This two belt-conveyer system is driven by two motors and the start-stop operations of these two motors define the discharging time. It is therefore important to obtain the correct set point (i.e., y_3^* as denoted in Fig. 2) for the control systems of these two motors.

This is a semi-batch process where ore feeding and discharge are realized periodically. In this process, magnetic tube recovery rate (MTRR) has been used as an index for both the product quality and process efficiency. Subject to temperature variations in the combustion chamber, the flow rate of the gas fuel and the ore discharge time, the system exhibits large nonlinearities and strong coupling between each loop. Specifically, the feeding and discharge phases are performed in a synchronized way. For this process, the key issue is that the ore should be kept in the deoxidizing phase for a particular time, so that the metal recovery percentage can be maximized. In the reaction zone, the weak magnetism ores (i.e., Fe_2O_3) are turned into strong magnetism ores (Fe_3O_4). The chemical reactions during deoxidization depend on the reaction temperature, the density of the reactants and the reaction time. In the roasting process of shaft furnace, the reaction temperature is determined by the temperature of the combustion chamber and the ore discharging time. The density of the reactants is determined by the flow rate of the deoxidization gas, whilst the reaction time is decided by the ore discharging time. As a result, the MTRR is primarily influenced by the control quality on the combustion chamber's temperature, the flow rate of the deoxidization gas and the ore discharge time. To achieve the desired control performance of these local control loops and to ensure that the system operates far away from possible fault condition's, the set points to these local con-

trol loops need to be carefully selected during the process operation.

3.2 Faults caused by operator's misplacement of control loop set points

Since the system models for this process vary with the operating conditions and are generally difficult to establish, the set points of each control loop are determined using the operational experience of operators. Moreover, when the source ore is subjected to unpredictable variations in their size, grades, and composition, the set points of each control loop cannot be accurately determined. This leads to inaccurate control of the performance indicators and often faults develop such as fire-emitting, ore-melting, and under-deoxidization. These faults are classified as abnormal system operation caused by the three set points.

Under such faulty conditions, operators modify the set points for the combustion temperature, the gas fuel flow rate and the ore discharge time using operational experience together with a visual inspection of the surface status of the furnace. This can cause the actual operating points to gradually drift away from healthy operational conditions. Moreover, since the operators cannot always determine the operation condition correctly and modify the set points in time, the control performance will deteriorate and even lead to the system failure. Therefore, the roasting process control in the furnace operation is a key factor that not only affects the product quality, efficiency and various consumptions, but is also heavily related to the operational stability and safety. It is therefore imperative that set points are correctly selected so that the whole system operates far away from the fault conditions.

3.3 An FTC solution

To solve this problem, a hybrid intelligent control method for the optimal process operation has been developed^[3] with the purpose of controlling all the technical indices to their desired ranges by on-line adjustment of the control loop set points in response to the variations in working-conditions. The proposed method is composed of a pre-setting model for control loops, a feed forward and a feedback compensator, an MTRR prediction model, a fault diagnosis unit and a fault-tolerant controller as shown in Fig. 3.

As illustrated in Fig. 3, the fault working condition diagnosis model identifies faults such as fire-emitting, ore-melting, and under-deoxidization caused by incorrect set points. This is followed by a fault tolerant controller which provides the necessary adjustments to the three set points

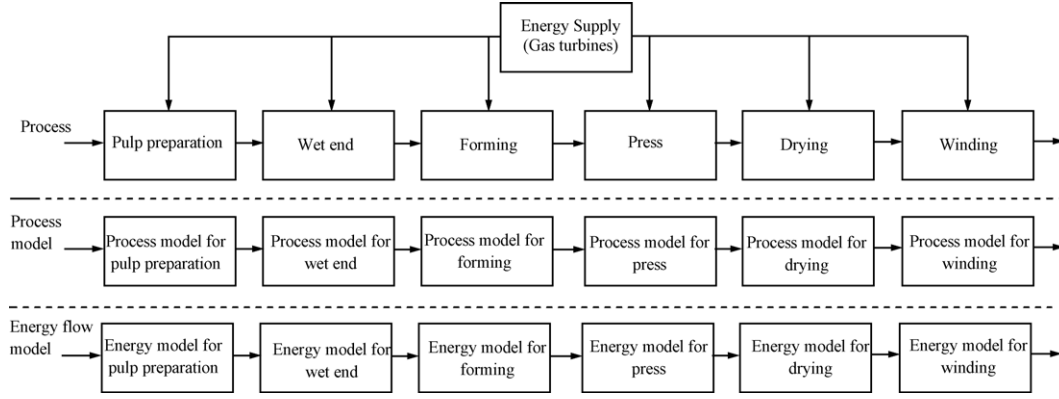


Fig. 4 Data driven multi-scale models for FDD and FTC

so that the required FTC can be achieved to keep the operating conditions away from the faults. Case-based and rule-based reasoning has been used in these two blocks to realize the required functions.

The proposed FDD and FTC methods have been successfully applied to the roasting process of twenty-two shaft furnaces in one of the largest mineral processing factories in China, where the MTRR has been controlled within its targeted range with an increased rate of 2%. Moreover, the operational efficiency of all the equipment has been improved by 2.98% and the metal recovery percentage has been increased by 2.01% with a total of 0.57% increase on the concentrated grade.

4 Possible future directions

In this section, a number of research issues that should be addressed in the near future will be discussed. Specifically, issues related to data driven modeling, probability density function based FDD and FTC, collaborative systems FDD and FTC, are described.

4.1 Multi-scale plant-wide modeling for FDD and FTC

As previously described, data based FDD and FTC face various challenges to make them practically useful as an integrated part of complex industrial processes. Primarily, data driven models should be considered because they have the advantage of exploiting the large volume of process data that exists in DCS systems (see Fig. 1). Once these data driven models are established, they can be used for a number of purposes — plant-wide optimization, plant-wide performance monitoring, and of course FDD and FTC.

Complex industrial processes are examples of sequentially arranged multi-agent systems where each production unit is connected in series so as to process the required intermediate products. Each product unit can be regarded as an agent and each agent would have a number of control systems so as to make it autonomous. Each agent has its own performance that needs to be considered in the plant-wide optimization and control. The agent's performance can be characterized in terms of its control performance, energy efficiency and the local environment impact. Therefore, an important difference from traditional modeling approaches, where only control function models are used, is that other criteria such as agent's energy model should also be used for data driven FDD and FTC. For example, in paper-making processes^[35], the production line consists of a number of sections that are connected in series as shown in Fig. 4.

In this figure, the top layer represents the process structure where there is a number of production units connected sequentially (from the pulp preparation, wet end, forming, press, drying to the final paper web winding phase). This arrangement turns the original raw materials (either wood chips or deinked pulps) into the finished paper. Along the production line, various control systems are used and the typical DCS (see Fig. 1) has been employed for many years. Traditionally, one would establish control models for each production units along the production line so as to optimize the effect of process control. However, with the ever increasing focus on energy reduction, the energy consumption of each unit needs to be considered in order to assess the plant-wide energy performance. This means that one should establish along the production line, energy models, and energy supply — usage distribution models. Therefore, data based modeling should aim to establish multi-scale plant-wide models for the systems. Here, the multi-scale models mean that these models work together and have different focuses such as process dynamic models and energy consumption models, etc.

Potential future research on multi-scale modeling could focus on investigating where process data should be best used for modelling of sequentially linked multi-agent systems. For example, along a general material processing production line, each unit (agent) accepts the material to be processed from the previous unit. Its purpose is to process the material so that it will have the required properties after passing through the unit. One can therefore represent the vector of material properties as \mathbf{X}_n and \mathbf{X}_{n+1} , prior to entering and after leaving the n -th unit, respectively. Then, the following structure represents the functioning of the n -th unit.

$$\mathbf{X}_n \oplus \mathbf{U}_n[t_n, t_{n+1}] \Rightarrow \mathbf{X}_{n+1} \quad (3)$$

where $\mathbf{U}_n[t_n, t_{n+1}]$ are the control variables applied in the n -th unit between time t_n and t_{n+1} so as to transform the material properties from \mathbf{X}_n into \mathbf{X}_{n+1} , and $\Delta t_n = t_{n+1} - t_n$ is the time duration for the material to pass through the n -th unit. Although different plants will have different models for each of the sequentially arranged units, the above structure can be generally applied as a signal pathway representation of material flow. In this context, the relationship between $\{\mathbf{X}_n, \mathbf{U}_n\}$ and \mathbf{X}_{n+1} can be generally represented as dynamic and possibly stochastic equations which are to be determined.

In (3), \mathbf{X}_n has components represented either by time-dependent variables (such as strength of the materials, etc) or by variable distributions (such as the particle size dis-

tributions represented by a probability density function (PDF) in polymerization processes). Moreover, the components in \mathbf{X}_n can be classified into two groups, namely measurable and un-measurable components. Since $\mathbf{U}_n[t_n, t_{n+1}]$ can be either the control inputs to each unit or the set point values from the DCS top layer, the obtained models will be multi-layer in structure. In addition, the communication between each unit and across each layer should also be considered in the modelling phase. The model therefore has a multi-layer structure in terms of control loops in each unit and the material properties transformations between \mathbf{X}_n and \mathbf{X}_{n+1} . As all the units are connected in series, the time delay of each unit with respect to the final products will be taken into account. Specifically, the reference time can be selected as the time when the products are produced. Due to the variations in the processing duration when the materials pass through each unit, these time delays will be time-varying in nature. It is expected that a general multi-scale model structure which takes into account of process dynamics will be developed for each agent as well as the whole system.

Using the models in (3), one of the future areas of research is to develop global PDF models that link relevant control variables in each production unit with the PDF of the final quality variables. For this purpose, a back propagation modelling procedure will be developed that links various unit model together via the time delay and process inherent multi-layer connections so as to form the following global and conditional PDF model for production qualities^[36].

$$\begin{aligned} \gamma(\mathbf{y}, \mathbf{U}_k) &= \gamma(\mathbf{y}, \mathbf{U}_N[t_{N-1}, t_N], \mathbf{U}_{N-1}[t_{N-2}, t_{N-1}], \dots, \\ &\quad \mathbf{U}_n[t_{n-1}, t_n], \dots, \mathbf{U}_1[t_0, t_1]) \\ \mathbf{U}_k &= (\mathbf{U}_N[t_{N-1}, t_N], \mathbf{U}_{N-1}[t_{N-2}, t_{N-1}], \dots, \\ &\quad \mathbf{U}_n[t_{n-1}, t_n], \dots, \mathbf{U}_1[t_0, t_1]) \end{aligned} \quad (4)$$

where \mathbf{U}_k denotes all the controllable variables of the system. When the sectional control variables are high dimensional, PCA algorithms can be used for data reduction so as to reduce the dimensionality of \mathbf{U}_k . Therefore, the final control vector that appears in the quality variables' PDFs will be either in terms of the original variables or in a projected subspace. Probability theory can also be applied to each sectional model to develop the quality variables' PDFs, where the recently developed recursive PDFs model^[36–37, 47] can also be used.

Indeed, research into such data driven multi-scale models would therefore constitute a potential starting point for data driven FDD and FTC that are discussed in the next subsection.

4.2 PDF based FDD and FTC using multi-scale plant-wide models

In line with the multi-scale models built for industrial processes discussed in Subsection 4.1, FDD and FTC can also be developed and traditional FDD and FTC, either model-based or data-based, can be used to deal with the faults for individual agents. However, a possible new research direction would be to investigate the quality data distribution of each agent or the final production quality. It is well-known that the product quality monitoring has been used for many years, however, the existing techniques are generally based on monitoring the means and the variances of important product quality variables in the data stream. This assumes that the process quality data is Gaussian distributed, however, for complex industrial processes, the quality data distribution is generally non-Gaussian. Of-

ten, it is only during healthy operation that the quality data distribution is approximately Gaussian distributed.

This phenomenon provides a potential insight towards the realization of data based FDD and FTC for complex industrial plant. For example, when the quality data PDF is not Gaussian, then it may indicate that there is either a fault in the process or a variation in the raw material. As such, the purpose of fault diagnosis is to firstly classify whether the non-Gaussian PDF is caused by the material variations or by a process fault. Once it has been confirmed that the non-Gaussian PDF is caused by the process fault, further fault diagnosis analysis is performed to locate the fault and estimate its size. In this context, observer based fault diagnosis can be used, where PDF residuals^[38–39, 48] are constructed for the sequential PDF models. Adaptive tuning rule-based fault diagnosis methods can be developed so as to guarantee the performance of the fault diagnosis. Once the fault diagnosis results are obtained, they will be combined with the relevant FTC control algorithm so that the whole plant-wide control strategy will be fault tolerant^[49]. Moreover, FTC design can also be made to minimize the entropy of the quality data so that a consistent product quality with minimum uncertainty (randomness) can be achieved^[50–52].

5 Conclusions and other challenging issues

Although multi-variable statistics based FDD has been applied to many industrial processes, there are still a number of challenging issues that need to be addressed. For example, multi-variable statistics based FDD is much more effective for fault detection rather than for fault diagnosis. Another issue that needs to be addressed is how effective combinations can be made between the data based FDD and the knowledge based FDD^[17–18]. Both methods make use of a large data volume, yet a combination of the human knowledge with process data needs to be made autonomous so as to apply them effectively to industrial processes. Practical issues are also important in applying data based FDD. This is because of the nonlinear, time-varying, batch and large scale nature of the systems that are widely seen in industrial processes. These characteristics make the available FDD algorithms complicated and novel techniques for data driven FDD need to be further explored.

Another potential area of research is the development of effective FDD and FTC for collaborative control systems. Consider an industrial process that is structured as a sequentially arranged multi-agent system. When an agent has been diagnosed to have a fault, it is important to explore how the subsequent healthy agents can be re-organized so that the end product is minimally affected. For example, in the paper making process illustrated in Fig. 4, when there is a fault in the pulp preparation unit so that the water content of the wet fiber entering the forming section is too high, then an FTC action should be made in the forming and drying sections so that the moisture content in the finished paper can still meet the product quality specification. Of course, such FTC should also be able to minimize the energy consumptions and other process specifications when energy distribution models are also used.

Fault prediction is also a potential research area^[53] in line with data driven based FDD and FTC. When a fault can be predicted, then an earlier FTC action can be made to prevent its occurrence. Indeed, fault prediction has been studied before in 1993 by the first author^[54] of this pa-

per, where the dynamics of the faults were represented by a first order state space model, and the original system input-output model was used as an extended output equation with respect to the fault dynamic model. Under this structure, the well-known Kalman filter can be applied to predict the faults for the required length of the time horizon.

To conclude, FDD and FTC for dynamic systems have been an area of intensive study and many successful process data based algorithms and their applications have been reported in the literature. Therefore, this paper does not aim to be a complete review, rather it is a collection of topics that the authors believe are important for dealing with data driven FDD and FTC for complex industrial processes.

References

- Chen J, Patton R J. *Robust Model-Based Fault Diagnosis for Dynamic System*. Boston: Kluwer Academic Publishers, 1999
- Jiang Jin. Fault tolerant control systems — an introductory overview. *Acta Automatica Sinica*, 2005, **31**(1): 161–174
- Chai Tian-You, Ding Jin-Liang, Wang Hong, Su Chun-Yi. Hybrid intelligent optimal control method for operation of complex industrial processes. *Acta Automatica Sinica*, 2008, **34**(5): 505–515 (in Chinese)
- Isermann R, Balle P. Trends in the application of model based fault detection and diagnosis of technical processes. In: *Proceedings of the 13th IFAC World Congress*. San Francisco, USA: Elsevier, 1996. 1–12
- Wang H, Daley S. Actuator fault diagnosis: an adaptive observer based approach. *IEEE Transactions on Automatic Control*, 1996, **41**(7): 1073–1078
- Wang H, Huang Z J, Daley S. On the use of adaptive updating rules for actuators and sensor fault diagnosis. *Automatica*, 1997, **32**(2): 217–225
- Lieftucht D, Druger U, Irwin G W. Improved reliability in diagnosing faults using multivariate statistics. *Computers and Chemical Engineering*, 2006, **30**(5): 901–912
- Martin E, Morris J, Lane S. Monitoring process manufacturing performance. *IEEE Control Systems Magazine*, 2002, **22**(5): 26–39
- Kano M, Hasebe S, Hashimoto I, Ohno H. A new multivariate statistical process monitoring method using principal component analysis. *Computers and Chemical Engineering*, 2001, **25**(7-8): 1103–1113
- Li W H, Yue H H, Valle-Cervantes S, Qin S J. Recursive PCA for adaptive process monitoring. *Journal of Process Control*, 2000, **10**(5): 471–486
- Lee J M, Yoo C, Lee I B. Statistical process monitoring with independent component analysis. *Journal of Process Control*, 2004, **14**(5): 467–485
- Lee J M, Yoo C, Lee I B. Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chemical Engineering Science*, 2004, **59**(14): 2995–3006
- Dunia R, Qin S J, Edgar T F, McAvoy T J. Identification of faulty sensors using principal component analysis. *American Institute of Chemical Engineers Journal*, 1996, **42**(10): 2797–2812
- Kano M, Nakagawa Y. Data-based process monitoring, process control, and quality improvement: recent developments and applications in steel industry. *Computers and Chemical Engineering*, 2008, **32**(1-2): 12–24
- Leung D, Romagnoli J. An integration mechanism for multivariate knowledge-based fault diagnosis. *Journal of Process Control*, 2002, **12**(1): 15–26
- Tatara E, Cinar A. An intelligent system for multivariate statistical process monitoring and diagnosis. *ISA Transactions*, 2002, **41**(2): 255–270
- Undey C, Tatara E, Williams B A, Birol G, Cinar A. A hybrid supervisory knowledge-based system for monitoring penicillin fermentation. In: *Proceedings of the American Control Conference*. Chicago, USA: IEEE, 2000. 3944–3948
- Glasse J, Montague G, Mohan P. Issues in the development of an industrial bioprocess advisory system. *Trends in Biotechnology*, 2000, **18**(4): 136–141
- Chiang L H, Leardi R, Pell R J, Seasholtz M B. Industrial experiences with multivariate statistical analysis of batch process data. *Chemometrics and Intelligent Laboratory Systems*, 2006, **81**(2): 109–119
- Albert S, Kinley R D. Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision. *Trends in Biotechnology*, 2001, **19**(2): 53–62
- Simoglou A, Georgieva P, Martin E B, Morris A J, de Azevedo S F. On-line monitoring of a sugar crystallization process. *Computers and Chemical Engineering*, 2005, **29**(6): 1411–1422
- Seongkyu Y, MacGregor J F. Principal-component analysis of multiscale data for process monitoring and fault diagnosis. *American Institute of Chemical Engineers Journal*, 2004, **50**(11): 2891–2903
- Qin S J, Cherry G, Good R, Wang J, Harrison C A. Semiconductor manufacturing process control and monitoring: a fab-wide framework. *Journal of Process Control*, 2006, **16**(3): 179–191
- Ku W F, Storer R H, Georgakis C. Disturbance detection and isolation by dynamic principal components analysis. *Chemometrics and Intelligent Laboratory Systems*, 1995, **30**(1): 179–196
- Choi S W, Lee C, Lee J M, Park J H, Lee I B. Fault detection and identification of nonlinear processes based on kernel PCA. *Chemometrics and Intelligent Laboratory Systems*, 2005, **75**(1): 55–67
- Deng X G, Tian X M. Multivariate statistical process monitoring using multi-scale kernel principal component analysis. In: *Proceedings of the 6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*. Beijing, China: Elsevier, 2006. 108–113
- Choi S W, Lee I B. Multiblock PLS-based localized process diagnosis. *Journal of Process Control*, 2005, **15**(3): 295–306
- Kano M, Nagao K, Hasebe S, Hashimoto I, Ohno H, Strauss R. Comparison of multivariate statistical process monitoring methods with applications to the Eastman challenge problem. *Computers and Chemical Engineering*, 2000, **24**(2-7): 175–181
- Zhao C H, Wang F L, Mao Z Z, Lu N Y, Jia M X. Adaptive monitoring based on independent component analysis for multiphase batch processes with limited modeling data. *Industrial and Engineering Chemistry Research*, 2008, **47**(9): 3104–3113
- Lee J M, Yoo C, Lee I B. On-line batch process monitoring using a consecutively updated multiway principal component analysis model. *Computers and Chemical Engineering*, 2003, **27**(12): 1903–1912
- Zhao C H, Wang F L, Gao F R, Lu N Y, Jia M X. Adaptive monitoring method for batch processes based on phase dissimilarity up-dating with limited modeling data. *Industrial and Engineering Chemistry Research*, 2007, **46**(14): 4943–4953
- Lopes J A, Menezes J C. Multivariate monitoring of fermentation processes with non-linear modelling methods. *Analytica Chimica Acta*, 2004, **515**(1): 101–108
- Norvilas A, Negiz A, DeCicco J, Cinar A. Intelligent process monitoring by interfacing knowledge-based systems and multivariate statistical monitoring. *Journal of Process Control*, 2000, **10**(4): 341–350
- Lesecq S, Gentil S, Taleb S. Fault detection based on wavelets transform: application to a roughing mill. In: *Proceedings of the 6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*. Beijing, China: Elsevier, 2006. 1115–1120
- Smook G A. *Handbook for Pulp and Paper Technologists*. Canada: Angus Wilde Publications, 1992
- Wang H. *Bounded Dynamic Stochastic Systems: Modelling and Control*. London: Springer-Verlag, 2000

- 37 Wang H. Control of conditional output probability density functions for general nonlinear and non-Gaussian dynamic stochastic systems. *IEEE Proceedings — Control Theory and Applications*, 2003, **150**(1): 55–60
- 38 Guo L, Wang H. Minimum entropy filtering for multivariate stochastic systems with non-Gaussian noises. *IEEE Transactions on Automatic Control*, 2006, **51**(4): 695–700
- 39 Guo L, Wang H. Fault detection and diagnosis for general stochastic systems using B-spline expansions and nonlinear filters. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2005, **52**(8): 1644–1652
- 40 Diao Y X, Passino K M. Stable fault-tolerant adaptive fuzzy/neural control for a turbine engine. *IEEE Transactions on Control Systems Technology*, 2001, **9**(3): 494–509
- 41 Guo Z H, Yue H, Wang H. A modified PCA based on the minimum error entropy. In: *Proceedings of the American Control Conference*. Boston, USA: IEEE, 2004. 3800–3801
- 42 Ferreira A P, Lopes J A, Menezes J C. Study of the application of multiway multivariate techniques to model data from an industrial fermentation process. *Analytica Chimica Acta*, 2007, **595**(1-2): 120–127
- 43 Yoo C K, Lee J M, Vanrolleghem P A, Lee I B. On-line monitoring of batch processes using multiway independent component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2004, **71**(2): 151–163
- 44 Patan K, Witczak M, Korbicz J. Towards robustness in neural network based fault diagnosis. *International Journal of Applied Mathematics and Computer Sciences*, 2008, **18**(4): 443–454
- 45 Fuessel D, Isermann R. Hierarchical motor diagnosis utilising structural knowledge and a self-learning neuro-fuzzy scheme. *IEEE Transactions on Industrial Electronics*, 2000, **47**(5): 1070–1077
- 46 Koppen-Seliger B, Frank P M. Fuzzy logic and neural networks in fault detection. *Fusion of Neural Networks, Fuzzy Sets, and Genetic Algorithms*. New York: CRC Press, 1999. 169–209
- 47 Guo L, Wang H, Wang A P. Optimal probability density function control for NARMAX stochastic systems. *Automatica*, 2008, **44**(7): 1904–1911
- 48 Guo L, Zhang Y M, Wang H, Fang J C. Observer-based optimal fault detection and diagnosis using conditional probability distributions. *IEEE Transactions on Signal Processing*, 2006, **54**(10): 3712–3719
- 49 Zhang Y M, Guo L, Yu H S, Zhao K Y. Fault tolerant control based on stochastic distributions via MLP neural networks. *Neurocomputing*, 2007, **70**(4-6): 867–874
- 50 Wang H. Minimum entropy control for non-Gaussian dynamic stochastic systems. *IEEE Transactions on Automatic Control*, 2002, **47**(2): 398–403
- 51 Yue H, Wang H. Minimum entropy control of closed loop tracking errors for dynamic stochastic systems. *IEEE Transactions on Automatic Control*, 2003, **48**(1): 118–122
- 52 Yue H, Zhou J L, Wang H. Minimum entropy control of B-spline PDF systems with mean constraint. *Automatica*, 2006, **42**(6): 989–994
- 53 Xu Z G, Ji J D, Zhou D H. Real-time reliability prediction for a dynamic system based on the hidden degradation process. *IEEE Transactions on Reliability*, 2008, **57**(2): 230–242
- 54 Wang H, Daley S. An approach to fault detection using nonlinear modelling and estimation. In: *Proceedings of the IFAC World Congress*. Sydney, Australia: Elsevier, 1993. 291–294



WANG Hong Received his bachelor degree from Huainan University of Mining Engineering, Huainan, China, in 1982, and his master and Ph.D. degrees from Huazhong University of Science and Technology, Wuhan China, in 1984 and 1987, respectively.

From 1988 to September 1992, he was a research fellow at Salford, Brunel, and Southampton Universities, UK. He then joined the University of Manchester in 2002, and has been a professor in process control at the Control Systems Centre, University of Manchester, Manchester, UK, since 2002. He also holds a research visiting position with the Northeastern University (China), Huazhong University of Science and Technology (China), and Institute of Automation, Chinese Academy of Sciences. His research interest covers stochastic distribution control, fault detection and diagnosis, nonlinear control, and databased modeling for complex systems. Corresponding author of this paper.
E-mail: Hong.wang@manchester.ac.uk



CHAI Tian-You Received his Ph.D. degree in control theory and engineering from Northeastern University, Shenyang, China, in 1985.

Since then he has been with the Center of Automation, Northeastern University, where he became a professor in 1988 and an academician of Chinese Academy of Engineering in 2003. His research interest covers adaptive control, intelligent decoupling control, integrated plant control and system, and the development of control technologies with applications to various industrial processes.
E-mail: tychai@mail.neu.edu.cn



DING Jin-Liang Ph.D. candidate at the Key Laboratory of Integrated Automation for Process Industry, Ministry of Education, Northeastern University. He received his bachelor and master degrees from Northeastern University in 2001 and 2004, respectively. His research interest covers modeling, control and optimization, and FDD of the complex industrial process.
E-mail: jlding@mail.neu.edu.cn



BROWN Martin Received his bachelor degree in applied mathematics from the University of Bath and his Ph.D. degree from the University of Southampton in 1989 and 1993, respectively. He is currently a senior lecturer in the School of Electrical and Electronic Engineering, University of Manchester. His research interest covers system identification, systems biology and hybrid, and robotic control.
E-mail: Martin.Brown@manchester.ac.uk