Pergamon

Chemical Engineering Science, Vol. 50, No. 18, pp. 2861–2882, 1995

0009–2509/95 $9.50 + 0.00

0009-2509(95)00127-1

IMPLEMENTATION ISSUES CONCERNING THE EKF-BASED FAULT DIAGNOSIS TECHNIQUES

CHUEI-TIN CHANG† and JEN-WEN CHEN
Department of Chemical Engineering, National Cheng Kung University, Tainan, Taiwan 70101, Republic of China

(*Received* 29 *September* 1994; *accepted in revised form* 16 *February* 1995)

Abstract—The extended Kalman filter (EKF) is one of the most popular model-based techniques for fault detection and diagnosis. Although its effectiveness has been widely recognized, the practical applications of EKFs are still very limited. This is due to the fact that the estimates of EKF are often biased when the occurrence of multiple faults is possible. In this study, we have extended the findings of our previous research on *diagnostic observability* and *diagnostic resolution* concerning a set of parallel single-parameter EKFs (Chang *et al.*, 1993, *A.I.Ch.E. J.* **39**, 1146) to the multiple-parameter EKFs which are designed to identify more than one fault origin. The problems in implementing these EKFs, i.e. misdiagnosis due to biased estimates and heavy computation load due to the parallel configuration, have been solved with a selection strategy for proper combinations of sensor locations and EKF parameters. More importantly, simple procedures have been developed to quickly evaluate the performance of any given system and the reliability of the proposed approach has been repeatedly confirmed in extensive simulation studies without exceptions. As a result, it becomes feasible to construct appropriate fault monitoring schemes without extensive computational effort even for large and complex chemical processes.

INTRODUCTION

Due to the frequency and seriousness of chemical accidents that have occurred during recent years, the importance of incipient fault detection and diagnosis in complex process plants has become apparent. Previous studies in this area are mainly concerned with the developments of computer-aided monitoring systems for assisting operators to respond to emergency situations effectively (Davis *et al.*, 1987; Lamb *et al.*, 1987; Petti *et al.*, 1990). Various different techniques were suggested, e.g. signed directed graph (Kramer and Palowitch, 1987; Chang and Yu, 1990; Yu and Lee, 1991), parameter estimation (Isermann, 1984), expert system (Rich and Venkatasubramanian, 1987; Petti *et al.*, 1990) and neural networks (Venkatasubramanian and Chan, 1989; Watanabe *et al.*, 1989; Venkatasubramanian *et al.*, 1990; Hoskins *et al.*, 1991). In this research, a model-based approach was adopted for the development of fault detection and diagnosis methods. More specifically, detection of the changes in states and/or parameters of the mathematical model was achieved by implementing system identification techniques and diagnosis of the fault origins, e.g. equipment malfunctions and external disturbances, were accomplished on the basis of the physical interpretations of these changes and/or structural analysis of the model. As a direct result of this approach, the candidate faults in this research were limited to those that can be associated with parameter variations. In other words, sensor failures and failures that can change the structure of system models were not considered in our analysis.

There are a large number of related works published in the literature (Willsky, 1976; Isermann, 1984; Ljung, 1987). Among various different estimation techniques adopted in the past, the extended Kalman filter (EKF) is clearly one of the most popular methods, e.g. Hamilton *et al.* (1973), Park and Himmelblau (1983), Watanabe and Himmelblau (1983a, b; 1984), Dalle Molle and Himmelblau (1987), Li and Olson (1991) and Venkateswarlu *et al.* (1992). In essence, EKFs of one form or another were employed to estimate both the states and parameters of chemical engineering systems and, then, causes of abnormal system behaviors were identified accordingly. Although the effectiveness of EKFs has been widely recognized, its use in commercial units was in fact very limited. This is mainly due to a critical drawback of EKF, i.e. its inability to guarantee unbiased estimates (Watanabe and Himmelblau, 1984). Obviously, incorrect information about the system parameters and/or states can mislead diagnosis. To overcome this problem, we modified the traditional way in implementing the extended Kalman filters in an earlier study (Chang *et al.*, 1993). Instead of estimating all parameters simultaneously in a large EKF, several single-parameter EKFs were used in parallel. Although this approach served for the purpose of eliminating bias, there were still several unsettled issues requiring further attention. In particular, the computational effort needed to carry out the parallel parameter estimation scheme can be overwhelming and, more importantly, the possibility of multiple

†Corresponding author.

faults was totally ignored. The present paper is a report of our recent studies which address these practical issues in implementing the EKF-based fault diagnosis techniques.

The remainder of this paper is divided into eight parts. First, the specific steps to identify both the states and parameters with a conventional EKF are described and various causes of biased estimation reported in the literature are briefly reviewed. Secondly, the concept of *diagnostic observability* is introduced to characterize outcomes of the above EKF computations. Next, an analysis of the propagation mechanisms of state estimates in EKF is provided to facilitate understanding of the relations between the state variables and model parameters. Then, on the basis of structural analysis of the system model, a testing procedure is outlined for identifying diagnostically observable EKFs. After the issue of observability is settled, a composite parameter estimation scheme, which incorporates several EKFs in parallel, is proposed to overcome another practical problem, i.e. misdiagnosis due to insufficient and/or inappropriate measurement points. Subsequently, the concept of *diagnostic resolution* is developed as a criterion for selecting the best scheme among competing candidates and a simple tool for its representation and quantification, i.e. the diagnostic performance table, is also described. Furthermore, a systematic procedure for constructing this table is presented in detail. Finally, two application examples are provided to show the usefulness of the diagnostic performance table and to demonstrate the practical value of the proposed approach.

PRACTICAL PROBLEMS IN IMPLEMENTING THE EKFs

As mentioned before, the extended Kalman filter is used in this study as a tool for fault detection and diagnosis. Because parameter estimates are more sensitive to faults than those of the state variables, they are in general a better indication of the degradation of system performance. Since it is usually possible to associate all the assumed malfunctions with changes in the corresponding model parameters, all such parameters must be treated as augmented states in the corresponding EKF (Himmelblau, 1978). Specifically, let us consider a system model with the following general form:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, \theta, t) + \omega_{\mathbf{x}} \tag{1a}$$

and

$$\omega_{\mathbf{x}} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{1b}$$

in which

$$\mathbf{f}(\cdot) = [f_1(\cdot), f_2(\cdot), \ldots, f_n(\cdot)]^T$$
$$\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$$
$$\theta = [\theta_1, \theta_2, \ldots, \theta_m]^T$$
$$\omega_{\mathbf{x}} = [\omega_1, \omega_2, \ldots, \omega_n]^T$$

where x_i's are the state variables and θ_j's are the parameters of the system model, f_k's are nonlinear functions of x_i's and θ_j's, ω_k's represent the normally distributed random system noises and $\mathbf{Q}$ is the covariance matrix associated with $\omega_{\mathbf{x}}$.

In order to estimate the model parameters (θ_j's) in an EKF, one can treat them as state variables and augment the corresponding equations with eq. (1a), i.e.

$$\frac{\mathrm{d}\tilde{\mathbf{x}}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbf{x}\\ \theta\end{bmatrix} = \begin{bmatrix}\mathbf{f}\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\omega_{\mathbf{x}}\\ \omega_\theta\end{bmatrix} = \tilde{\mathbf{f}} + \tilde{\omega} \tag{2a}$$

and

$$\tilde{\omega} \sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{Q}}) \tag{2b}$$

where, ω_θ is a m-dimensional random vector with mean zero. For the sake of convenience, the components in $\tilde{\omega}$ are assumed to be independent and thus $\tilde{\mathbf{Q}}$ is a diagonal matrix. Also, without loss of generality, it is assumed in this study that the first s ($s \leqslant n$) state variables can be measured directly. In other words, the measurement model can be written as

$$\mathbf{z}_l = \mathbf{H}\tilde{\mathbf{x}}_l + \mathbf{v}_l = [\mathbf{I}|\mathbf{0}]\tilde{\mathbf{x}}_l + \mathbf{v}_l \tag{3a}$$

$$\mathbf{v}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \tag{3b}$$

$$l = 0, 1, 2, \ldots$$

where, $\mathbf{z}_l$, $\tilde{\mathbf{x}}_l$ and $\mathbf{v}_l$ are the system output vector, augmented state vector and measurement noise vector, respectively, at time t_l, $\mathbf{I}$ is a $s \times s$ identity matrix and $\mathbf{0}$ is a $s \times (m + n - s)$ matrix whose entries are all zeros. Also, $\mathbf{R}$ is assumed to be a diagonal matrix in this study.

From the above formulations, it is plain that they are consistent with the standard form and thus an EKF can be adopted directly to produce the estimates of parameters and state variables simultaneously. As a result of this approach to fault detection and diagnosis, it becomes necessary to consider the common problems often observed in implementing the Kalman filters. In particular, the phenomena of "apparent" and "true" divergence (Gelb, 1974) in the corresponding calculations must be studied in depth. The possibility of unpredictable bias in estimations (Watanabe and Himmelblau, 1984) is regarded as the main reason which has prevented the use of EKFs in many realistic fault monitoring systems. There are in fact a variety of causes for the failure of a Kalman filter to achieve its theoretical performance. According to Grewal and Andrews (1993), they can be grouped into five types, i.e. (i) natural behavior of the dynamic equations, (ii) unobservability with the given measurement, (iii) bad on-line data, (iv) numerical problems and (v) mismodelling. Type (i) causes can be attributed to significant system and/or measurement noises associated with equipments and, thus, the remedy for this type of problems is hardware improvement. The third type of causes are due to large, one of a kind, isolated errors, i.e. "blunders", in the sensor systems. The best way to overcome this problem is to implement prefiltering and data rejection methods in real time. Sources of type (iv) problems can usually be traced to truncation errors in computation. Possible counter measures include the use of higher (double) precision and more numerically stable methods in software. To simplify the arguments presented later in this paper, let us assume that these three types of causes, i.e. type (i), (iii) and (iv), have already been removed through hardware and/or software means. Therefore, only structural issues, i.e. those concerned with type (ii) and (v) causes, are addressed in this work. Specifically, under the assumptions stated above and that accurate system model, eqs (1a) and (1b), is available, our objective is to develop a practical strategy for constructing the appropriate augmented system equations, eqs (2a) and (2b), and measurement model, eqs (3a) and (3b), so that all possible combinations of multiple faults can be diagnosed correctly.

DIAGNOSTIC OBSERVABILITY

As mentioned previously, one of the reasons for the failure of a Kalman filter to produce unbiased estimates is due to the fact that the system itself is unobservable, i.e. type (ii) causes. However, in this study, we are interested in the opposite question: can the traditional observability criteria, e.g. see Gelb (1974), guarantee the correctness of parameter estimates? The answer to this question can be found readily in the following example.

Example 1

Let us use the simplest case, i.e. a linear constant-coefficient system, to generate the information we need. Consider the system model

$$\frac{dx_1}{dt} = -4.2x_1 + 5.3x_2 - x_3 \tag{4a}$$

$$\frac{dx_2}{dt} = 1.1x_1 - 7.1x_2 + x_4 \tag{4b}$$

where x_1 and x_2 are the state variables and x_3 and x_4 are the model parameters associated with possible faults. If only x_1 can be measured on-line, the observability matrix of the augmented system equations, i.e. those corresponding to eqs (2a) and (3a), can be determined to be

$$\begin{bmatrix} 1 & -4.2 & 23.47 & -164.453 \\ 0 & 5.3 & -59.89 & 549.610 \\ 0 & -1.0 & 4.20 & -23.470 \\ 0 & 0.0 & 5.30 & -59.890 \end{bmatrix}. \tag{5}$$

From the fact that this matrix is of full rank, one can conclude that the augmented system is observable.

The corresponding Kalman filter was then tested using numerical simulation. It was assumed that the system was operated at its normal steady state initially and the faults associated with x_3 and x_4 occurred simultaneously at $t = 40$. More specifically, the changes in these parameters were described with:

$$\frac{dx_3}{dt} = \frac{dx_4}{dt} = 0.02[u(t-40) - u(t-50)] \tag{6}$$

where, $u(t - \bar{t})$ is a unit step function defined by

$$u(t-\bar{t}) = \begin{cases} 1 & \text{if } t \geqslant \bar{t} \\ 0 & \text{otherwise} \end{cases}$$

Equations (4a), (4b) and (6) were integrated together to produce the transient behavior of the state variables. Measurement noise was simulated with a random number generator and then added to x_1 to obtain the simulated on-line measurement data. The covariance matrices $\tilde{\mathbf{Q}}$ and $\mathbf{R}$ used in the corresponding EKF were:

$$\tilde{\mathbf{Q}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 10^{-4} & 0 \\ 0 & 0 & 0 & 10^{-4} \end{bmatrix} \tag{7}$$

$$\mathbf{R} = [10^{-6}]. \tag{8}$$

The results of EKF estimation are presented in Fig. 1(a)–(d). One can see clearly that, although the estimates of the measured state variable x_1 are close to the true values, the estimates of another state variable, i.e. x_2, and the model parameters, i.e. x_3 and x_4, are erroneous. Furthermore, this situation cannot be improved by adjusting the entries of $\tilde{\mathbf{Q}}$ and $\mathbf{R}$. Thus, in this system, it is not possible to trace the changes of parameters with Kalman filter using only the measurement data of x_1. This phenomenon will be referred to as "diagnostically unobservable" in this paper.

If both x_1 and x_2 are selected as the measurement variables, the resulting system should still be observable in the traditional sense, i.e. the corresponding observability matrix is of full rank. However, contrary to the previous results, it can be demonstrated with numerical simulation that all the state variables and model parameters can be estimated correctly with a Kalman filter. In this case, the system should be regarded as "diagnostically observable".

At this point, the answer to our previous question should be quite clear. One can easily see that the traditional observability criteria cannot be adopted as the sufficient conditions for *diagnostic observability*. One can also observe from the above results that biased estimation is really a practical problem that occurs in the implementation of EKF for the purpose of simultaneous state and parameter identification. Since none of the available techniques address this issue adequately, it is thus necessary to develop an alternative approach.

PROPAGATION OF THE STATE ESTIMATES

In this study, it is our intention to develop qualitative criteria of diagnostic observability based only on structural information of the system model. The practical value of such an approach is quite evident. Since

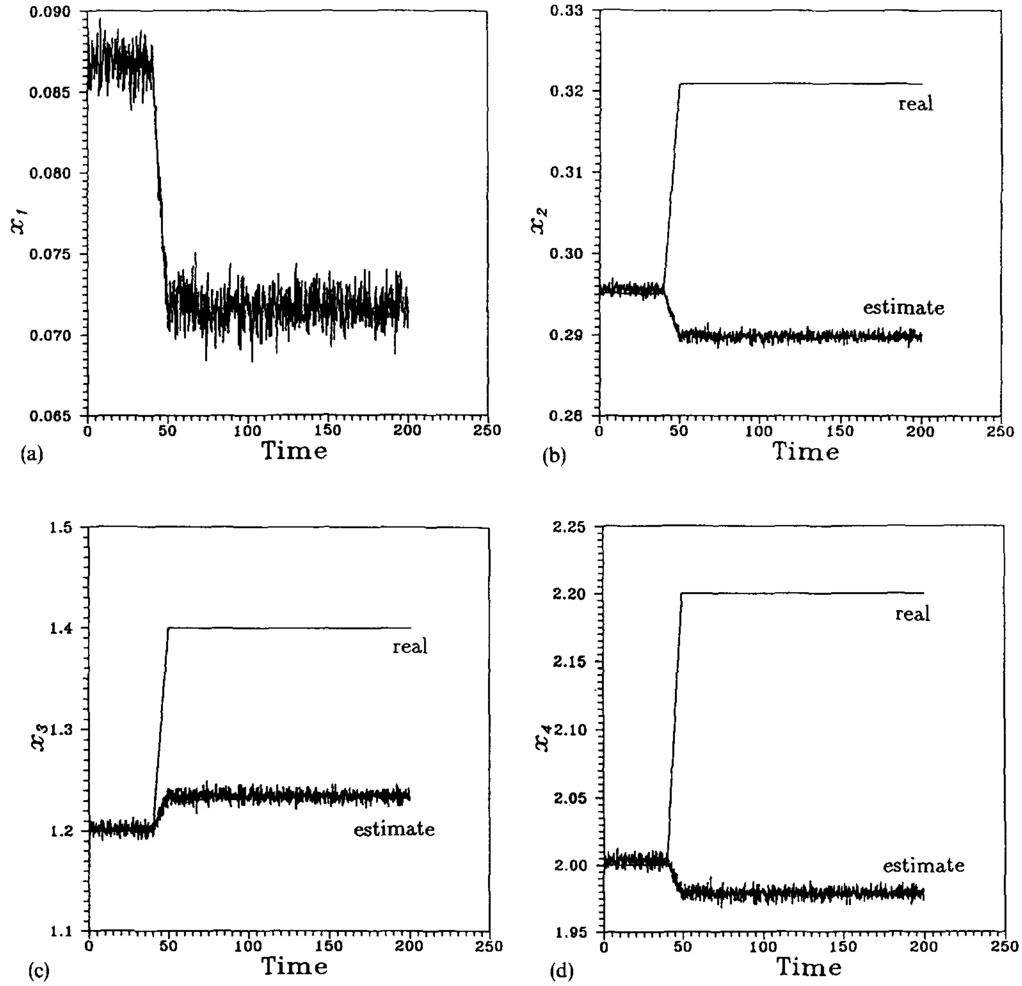


Fig. 1. EKF estimates of: (a) state variable x_1; (b) state variable x_2; (c) model parameter x_3; (d) model parameter x_4.

there is no need for quantitative calculations, such as checking the observability matrix, the results of EKF estimation can be characterized quickly in advance even for a large and complex process. Although, by definition, the qualitative methods developed in this work are not theoretically rigorous, their effectiveness can nevertheless be verified with extensive simulation studies.

To achieve our purpose, it is obvious that an understanding of the algorithm for generating the estimates is necessary. As mentioned previously, if an EKF is to be used for fault detection and diagnosis, its state estimate propagation equations must be formulated according to eq. (2a). At any sampling time (say t_{l-1}), these equations should be integrated numerically to determine the estimates at the next time step, i.e. t_l. Without loss of generality, let us assume that the implicit Euler method is used, i.e. the state estimate propagation equations can be approximated by

$$\begin{aligned}
&\hat{x}_1^{(-)}(t_l) - \hat{x}_1^{(+)}(t_{l-1}) \\
&\quad - f_1[\hat{x}_1^{(-)}(t_l), \dots, \hat{x}_n^{(-)}(t_l); \hat{\theta}_1^{(-)}(t_l), \dots, \hat{\theta}_m^{(-)}(t_l)]\,\Delta t = 0 \\
&\qquad \vdots \\
&\hat{x}_n^{(-)}(t_l) - \hat{x}_n^{(+)}(t_{l-1}) \\
&\quad - f_n[\hat{x}_1^{(-)}(t_l), \dots, \hat{x}_n^{(-)}(t_l); \hat{\theta}_1^{(-)}(t_l), \dots, \hat{\theta}_m^{(-)}(t_l)]\,\Delta t = 0 \\
&\hat{\theta}_1^{(-)}(t_l) - \hat{\theta}_1^{(+)}(t_{l-1}) = 0 \\
&\qquad \vdots \\
&\hat{\theta}_m^{(-)}(t_l) - \hat{\theta}_m^{(+)}(t_{l-1}) = 0
\end{aligned} \tag{9}$$

where $\hat{x}_i^{(-)}(t_l)$ $(i = 1, 2, \dots, n)$ and $\hat{\theta}_j^{(-)}(t_l)$ $(j = 1, 2, \dots, n)$ are the estimates of states and parameters at time t_l, which should be considered as the unknowns of eqs (9). On the other hand, $\hat{x}_i^{(+)}(t_{l-1})$ and $\hat{\theta}_j^{(+)}(t_{l-1})$ are the *updated* estimates of states and parameters at time t_{l-1}. Their values must be determined before solving eqs (9) with the state estimate update equations of EKF.

In previous studies, e.g. Watanabe and Himmelblau (1984), it is a widely accepted assumption that the model parameters are more sensitive (than the state variables) to incipient faults. Thus, if an EKF is used for the purpose of fault detection and diagnosis, it is a common practice to adopt a system noise covariance matrix, i.e. $\tilde{Q}$ in eq. (2b), with the form

$$\tilde{Q} = \begin{bmatrix} 0 & 0 & & \cdots & & 0 \\ 0 & \ddots & & & & \\ & & 0 & & & \vdots \\ \vdots & & & q_{n+1,n+1} & & \\ & & & & \ddots & 0 \\ 0 & & \cdots & & 0 & q_{n+m,n+m} \end{bmatrix} \tag{10}$$

As a result, the corrections in the updated estimates of the state variables are in general small, i.e.

$$\hat{x}_i^{(+)}(t_{l-1}) = \hat{x}_i^{(-)}(t_{l-1}) + \Delta x_i(t_{l-1}) \sim \hat{x}_i^{(-)}(t_{l-1})$$
$$i = 1, 2, \ldots, n. \tag{11}$$

On the other hand, the values of $\hat{\theta}_j^{(+)}(t_{l-1})$'s are allowed to deviate from $\hat{\theta}_j^{(-)}(t_{l-1})$'s, i.e.

$$\hat{\theta}_j^{(+)}(t_{l-1}) = \hat{\theta}_j^{(-)}(t_{l-1}) + \Delta\theta_j(t_{l-1}) \neq \hat{\theta}_j^{(-)}(t_{l-1})$$
$$j = 1, 2, \ldots, m. \tag{12}$$

Since the estimates of state variables and model parameters, i.e. $\hat{x}_i^{(-)}(t_{l-1})$'s and $\hat{\theta}_j^{(-)}(t_{l-1})$'s, were computed at time t_{l-2} and should be available at time t_{l-1}, eqs (9) can be written as

$$\begin{aligned} &\phi_1[\hat{x}_1^{(-)}(t_l), \ldots, \hat{x}_n^{(-)}(t_l); \hat{\theta}_1^{(+)}(t_{l-1}), \ldots, \hat{\theta}_n^{(+)}(t_{l-1})] = 0 \\ &\qquad\vdots \\ &\phi_n[\hat{x}_1^{(-)}(t_l), \ldots, \hat{x}_n^{(-)}(t_l); \hat{\theta}_1^{(+)}(t_{l-1}), \ldots, \hat{\theta}_n^{(+)}(t_{l-1})] = 0. \end{aligned} \tag{13}$$

At time t_{l-1}, one first needs to determine the values of $\hat{\theta}_j^{(+)}(t_{l-1})$'s and, then, eqs (13) can be solved accordingly. In other words, the state estimates at time t_l are mainly functions of the updated estimates of the model parameters at time t_{l-1}.

If the EKF performs satisfactorily, correct estimates of $\hat{\theta}_j^{(+)}(t_{l-1})$'s can be chosen to yield the following results:

$$\hat{x}_i^{(-)}(t_l) \sim \bar{x}_i \quad i = 1, 2, \ldots, s \tag{14}$$

where, $\bar{x}_i$'s are the measurement values of the state variables at time t_l. However, it is also a well-known fact that biased estimation is a phenomenon often encountered in the practical applications of EKF. Thus, the structure of the state estimate propagation equations, i.e. eqs (13), must be further analyzed to understand the *precedence order of influences* among the state variables, i.e. $\hat{x}_i^{(-)}(t_l)$'s, due to changes in the adjustable parameters $\hat{\theta}_j^{(+)}(t_{l-1})$'s.

STRUCTURAL ANALYSIS OF THE SYSTEM MODEL

There are in general two standard approaches that can be used to solve n nonlinear algebraic equations numerically. One can certainly try to obtain the n unknowns simultaneously using, for example, the Newton–Raphson method. On the other hand, one can also try to promote the computation efficiency by reordering the equations according to the model structure (Steward, 1965; Christensen and Rudd, 1969; Christensen, 1970; Stadtherr *et al.*, 1974). The result of reordering is called a *partition*. More specifically, a partition is the division of the set of equations into subsets, which we call *blocks*, so that each block in the partition is the smallest set of equations that must be solved simultaneously. After such a partition is established, the blocks can be solved one at a time in series.

In this work, the partition algorithm suggested by Steward (1965) is extended to determine the precedence order. Consequently, a systematic procedure can be developed for testing the diagnostic observability of any given system. To facilitate understanding of the rationale behind various steps in this proposed procedure, a series of examples are presented below. First of all, it is intuitively correct that, in a diagnostically observable system, the symptoms of faults must appear in the measurement data. Precedence order can be used as an aid to determine whether this criterion is satisfied. This fact can be demonstrated with a simple example:

Example 2

Let us consider the system of two storage tanks connected in series (see Fig. 2). The system model can be written as

$$A_1 \frac{dh_1}{dt} = q_i - q_1 - \Delta c l_1 \sqrt{h_1} \tag{15}$$

$$\frac{dq_1}{dt} = \frac{\pi d_1^2}{4\rho l_1}\left[\rho g h_1 - \frac{8 f_1 l_1 \rho q_1 |q_1|}{\pi^2 d_1^5}\right] \tag{16}$$

$$A_2 \frac{dh_2}{dt} = q_1 - q_2 \tag{17}$$

$$\frac{dq_2}{dt} = \frac{\pi d_2^2}{4\rho l_2}\left[\rho g h_2 - \frac{8(f_2 + \Delta f_2) l_2 \rho q_2 |q_2|}{\pi^2 d_2^5}\right] \tag{18}$$

where ρ is the density of liquid, h_k and A_k denote, respectively, the height of liquid level in and the cross-sectional area of tank k ($k = 1, 2$), q_k, d_k, l_k and f_k represent, respectively, the volumetric flow rate in and the diameter, length and friction factor of the outlet pipeline from tank k ($k = 1, 2$). Also notice that the parameters $\Delta c l_1$ and Δf_2 are associated with two assumed failures, i.e. leakage in tank 1 and partial blockage in the exit pipeline of tank 2, respectively.

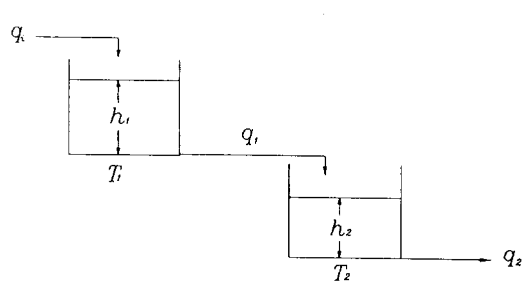


Fig. 2. The simplified flow diagram of the two-tank system in example 2.

As indicated previously, the state estimate propagation equations, eqs (13), can be solved if the values of model parameters are given. The corresponding precedence order can be determined with Steward's partitioning algorithm. Since the algorithm has been well documented in the literature, we choose not to present the detailed implementation steps for the sake of brevity. The result is represented with a structural matrix [see Fig. 3(a)] and also a precedence diagram [Fig. 3(b)]. Notice that ϕ_1–ϕ_4 in the structural matrix are corresponding to eqs (15)–(18), respectively. The equations ϕ_5 and ϕ_6 are:

$$\phi_5: \quad \Delta cl_1 - c_5 = 0 \tag{19}$$

$$\phi_6: \quad \Delta f_2 - c_6 = 0 \tag{20}$$

where c_5 and c_6 are the guess values of Δcl_1 and Δf_2, respectively.

From Fig. 3(a) and (b), one can see that all four state variables in this system are affected by a leak in tank 1, but partial blockage in pipeline 2 can only cause h_2 and q_2 behave abnormally. Thus, if h_1 and q_1 (or only one of them) are chosen as the measurement variables, it is certainly not possible to produce correct estimates of Δf_2 on the basis of the available on-line data.

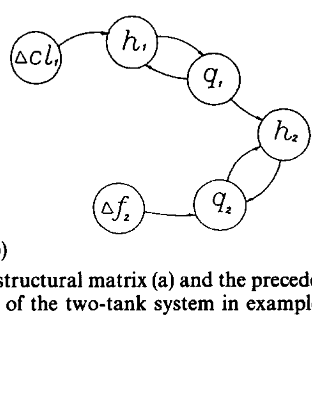
Although direct application of the Steward's algorithm yields a precedence order which is useful for identifying a special class of diagnostically unobservable systems, this approach is still limited in the sense that the correctness and uniqueness of the solutions to the state estimate propagation equations cannot be confirmed accordingly. This situation can be illustrated with another example:

Example 3

A simplified flow diagram of the system considered in this example is presented in Fig. 4(a). It is assumed that there are two possible faults, i.e. (i) a sudden change in the inlet flow rate and (ii) partial blockage in the pipeline between tank T_1 and tank T_2. The structural matrix of the system model can be found in Fig. 4(b). Notice that Δq_i and Δf_1 are the parameters associated with faults (i) and (ii), respectively. The definitions of other symbols used here are essentially the same as those adopted in example 2. Again, Steward's partitioning algorithm has been applied and the precedence diagram can be obtained accordingly [see Fig. 4(c)].

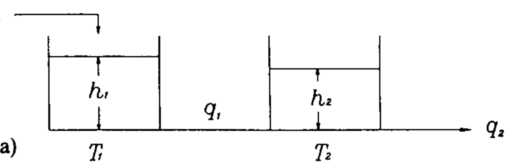
	Δcl_1	Δf_2	h_1	q_1	h_2	q_2
ϕ_5	×					
ϕ_6		×				
ϕ_1	×		×	×		
ϕ_2			×	×		
ϕ_3				×	×	×
ϕ_4		×			×	×

(a)



(b)

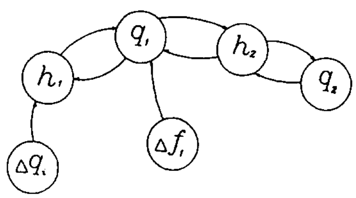
Fig. 3. The structural matrix (a) and the precedence diagram (b) of the two-tank system in example 2.



(a)

	h_1	q_1	h_2	q_2	Δf_1	Δq_i
ϕ_1	×	×				×
ϕ_2	×	×	×		×	
ϕ_3		×	×	×		
ϕ_4			×	×		
ϕ_5					×	
ϕ_6						×

(b)



(c)

Fig. 4. (a) The simplified flow diagram of the two-tank system in example 3. (b) The structural matrix of the two-tank system in example 3. (c) The precedence diagram of the two-tank system in example 3.

From Fig. 4(c), it can be observed that all four state variables, i.e. h_1, q_1, h_2 and q_2, are affected by the two parameters Δq_i and Δf_1. Thus, if at least one state variable can be measured on-line, the occurrence of faults should be detectable. However, one can also find in the precedence diagram that all state variables are connected to *both* parameters and they are interconnected via several feedback loops. In other words, all of them are within one block and thus must be solved simultaneously. Therefore, on the basis of this precedence order, one still cannot be certain whether a *unique* set of correct parameter values can be found to satisfy the requirements implied in eqs (14).

If the measurement variables are embedded in coupled, feedback loops, one can conclude from the above discussions that the precedence order obtained with the traditional approach is not really useful for the purpose of confirming diagnostic observability. Thus, additional tools must be developed for our purpose.

It should be noted that it may not be necessary to guess and iterate each variable in solving an irreducible set of equations. For sparse equation sets, it is often possible to reach a solution by guessing only a few of the variables. This is the so-called *tearing* technique recommended in the literature (Christensen and Rudd, 1969; Christensen, 1970; Stadtherr *et al.*, 1974), which can be used to determine an efficient iteration procedure. In this research, this method was also adopted as an aid for clarifying the cause-and-effect relations between the model parameters and the measurement variables in the state estimate propagation equations. Specifically, all s measurement variables were treated as the "tear-variables", i.e. the variables which are guessed, and s "tear equations", i.e. the equations used to check the guesses, were then chosen from eqs (13). These tear equations were not determined with conventional strategies, e.g. Stadtherr *et al.* (1974). A simple criterion was used instead, i.e. each equation must contain the corresponding tear variable. The advantage of such an approach can be demonstrated clearly with the following example:

Example 4

Let us reconsider the system described in Example 3. Assume that h_1 and h_2 are the measurement variables in this case and thus they should be regarded as the tear variables in the structural analysis. The tearing operation can be performed on the structural matrix presented in Fig. 4(b). Specifically, in each of the two columns associated with h_1 and h_2, all elements except an arbitrarily chosen one should be removed [see Fig. 5(a)]. The tear equations are corresponding to the rows in which these remaining elements are located, i.e. ϕ_1 and ϕ_3. Having determined the tear sets, one can then rearrange the structural matrix in Fig. 5(a) according to Steward's partitioning algorithm [see Fig. 5(b)]. The corresponding precedence order can be found in Fig. 5(c).

In the conventional process of solving algebraic equations, the tear variables are unknowns. Their values must be obtained through iterative procedure. In this work, however, the desired values of the measurement variables should satisfy the constraints stipulated in eqs (14). Thus, in solving the state estimate propagation equations of this example, the values of tear variables can be set directly to be the measurement values and the outputs h_1 and h_2 can then be calculated according to the precedence order in Fig. 5(c) and any given values of Δf_1 and Δq_i. Also notice that, in this computation process, q_2 should be regarded as a constant since it is affected only by the tear variable h_2.

The most appropriate parameter values should be chosen, of course, on the basis of eqs (14). From Fig. 5(c), one can see that a change in either of the parameters Δf_1 and Δq_i can cause variations in one or both of the output variables, h_1 and h_2. Since the two parameters in this case must be adjusted *simultaneously* in order to produce the two desired output values, it is therefore assumed that the chance for biased EKF estimation in this situation is low and the system should be diagnostically observable. This assumption has been verified with extensive simulation studies and the results are documented elsewhere (Chen, 1994).

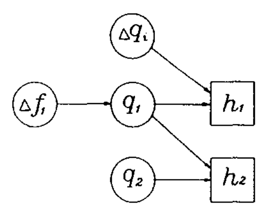
To facilitate later discussions, it is now necessary to classify the model parameters and measurement variables according to the precedence diagram just described. In particular, if a parameter is connected to one or more measurement variables, it is referred to as a "tunable parameter". On the other hand, if a measurement variable is connected to at least one parameter, then this variable is the "affected variable". If some of the tunable parameters can be assigned *independently* and the rest of the parameters can always be adjusted accordingly to produce the desired values for the affected variables, then there should be an infinite number of suitable parameter values that can satisfy eqs (14). As a result, the possibility of biased EKF estimation is extremely high and the system should also be regarded as diagnostically unobservable. The following example is used to demonstrate this special feature in EKF estimation.

(a)

	h_1	q_1	h_2	q_2	Δf_1	Δq_i
ϕ_1	×	×				×
ϕ_2		×			×	
ϕ_3		×	×	×		
ϕ_4				×		
ϕ_5					×	
ϕ_6						×

(b)

	Δf_1	Δq_i	q_1	q_2	h_1	h_2
ϕ_5	×					
ϕ_6		×				
ϕ_3	×		×			
ϕ_4				×		
ϕ_1		×	×		×	
ϕ_2			×	×		×



(c)

Fig. 5(a). The post-tearing matrix of the two-tank system in example 4. (b) The re-ordered post-tearing matrix of the two-tank system in example 4. (c) The post-tearing precedence diagram of the two-tank system in example 4.

Example 5

Let us again consider the system described in Example 3 and use h_2 and q_2 as the measurement variables this time. After tearing, the resulting precedence order can be obtained with Steward's partitioning algorithm (see Fig. 6). One can see that both Δq_i and Δf_1 are tunable parameters, but only h_2 is affected by these two parameters. Thus, the value of either one of the parameters can be assigned arbitrarily first and then the other parameter can always be adjusted to ensure output h_2 approaching its measurement value. Since the EKF does not have a prior knowledge concerning the actual variations in Δq_i and Δf_1, the possibility of obtaining the correct results is almost nil in the corresponding optimal estimation process. The prediction that this system is diagnostically unobservable has been supported by extensive simulation results presented elsewhere (Chen, 1994).

Although the precedence order obtained after tearing can be adopted as the basis for identifying diagnostically unobservable systems, the methods mentioned above are still difficult to apply when the system is large and complicated. Thus, a systematic procedure has been developed to overcome this problem. Since we are interested in keeping the affected variables at their measurement values, the corresponding equations in eqs (14) can be augmented with eqs (13) and our problem at hand can be transformed into one of determining the uniqueness of the solution to these augmented equations *after* tearing. For example, if this approach is used in example 4, the structural matrix in Fig. 5(a) should be changed into the matrix presented in Fig. 7(a). Notice that the equations ϕ_5 and ϕ_6 are replaced by

$$\phi'_5 = h_1 - \bar{h}_1 = 0 \qquad (21)$$

$$\phi'_6 = h_2 = \bar{h}_2 = 0. \qquad (22)$$

After executing the Steward's partitioning algorithm again, the precedence order for calculating the parameters can be obtained [see Fig. 7(b)]. From this result, one can clearly see that Δf_1 and Δq_i can be sequentially determined. Specifically, q_1 can be computed first by substituting h_2 and q_2 into ϕ_3; Δf_1 can then be calculated by substituting q_1 into ϕ_2; Δq_i can finally be determined by substituting h_1 and q_1 into ϕ_1. Thus, it can be concluded that the parameter

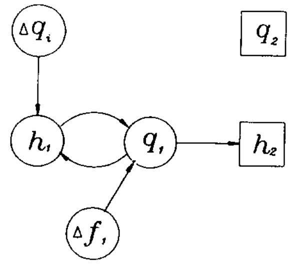


Fig. 6. The post-tearing precedence diagram of the two-tank system in example 5.

(a)

	h_1	q_1	h_2	q_2	Δf_1	Δq_i
ϕ_1	×	×				×
ϕ_2		×			×	
ϕ_3		×	×	×		
ϕ_4				×		
ϕ'_5	×					
ϕ'_6			×			

(b)

	h_1	h_2	q_2	q_1	Δf_1	Δq_i
ϕ'_5	×					
ϕ'_6		×				
ϕ_4			×			
ϕ_3		×	×	×		
ϕ_2				×	×	
ϕ_1	×			×		×

Fig. 7. (a) The augmented matrix of the two-tank system in example 4. (b) The re-ordered augmented matrix of the two-tank system in example 4.

values are unique and the corresponding EKF estimates should be unbiased.

On the other hand, if the proposed procedure is to be applied to the problem described in example 5, the resulting precedence order can be represented by the matrix given in Fig. 8. Notice that, since there is one affected variable in this case, it is only necessary to augment one equation, i.e. $\phi'_5 = h_2 - \bar{h}_2 = 0$, to the state estimate propagation equations. According to this precedence order, q_1 can be computed by substituting h_2 into ϕ_3. However, h_1, Δf_1 and Δq_i have to be determined with only two equations, i.e. ϕ_1 and ϕ_2. Thus, the solution is not unique and the system is diagnostically unobservable. In fact, since the number of variables is larger than that of equations in Fig. 8, this conclusion can be drawn without actually completing the Steward's algorithm. In implementing the procedure suggested by Steward, each variable must be assigned as the output of one equation and, in addition, each equation can only have one output which is distinct from that of another equation. Thus, these equations cannot be solved unless the value of one variable is fixed arbitrarily, which implies that there are an infinite number of solutions.

In some cases, the task of identifying unobservable systems is not as trivial as that described above. For instance, let us consider the fictitious augmented matrix presented in Fig. 9(a). Notice that, although the number of equations equals that of the variables, the system is still diagnostically unobservable. The main reason for this outcome is again that not all variables can be assigned as outputs. This special insight can be gained by rearranging the structural matrix [see Fig. 9(b)]. From the corresponding precedence order, one can see that it is necessary to determine three parameters (θ_1, θ_2 and θ_3) with only two equations (ϕ_1 and ϕ_2). This need is not immediately obvious from the matrix in Fig. 9(a).

	h_2	q_2	q_1	h_1	Δf_1	Δq_i
ϕ_5'	×					
ϕ_4		×				
ϕ_3	×		×			
ϕ_1			×	×		×
ϕ_2			×	×	×	

Fig. 8. The re-ordered augmented matrix of the two-tank system in example 5.

(a)

	θ_1	θ_2	θ_3	x_1	x_2	x_3	x_4	x_5
ϕ_1	×	×		×				
ϕ_2		×	×		×			
ϕ_3				×		×		
ϕ_4				×	×		×	
ϕ_5					×			×
ϕ_6'						×		
ϕ_7'							×	
ϕ_8'								×

(b)

	x_3	x_4	x_5	x_1	x_2	θ_1	θ_2	θ_3
ϕ_6'	×							
ϕ_7'		×						
ϕ_8'			×					
ϕ_3	×			×				
ϕ_4		×		×	×			
ϕ_5			×		×			
ϕ_1				×		×	×	
ϕ_2					×		×	×

Fig. 9. (a) The augmented matrix of a fictitious system. (b) The re-ordered augmented matrix of a fictitious system.

In order to systematically detect such structures in large and complex systems, the Steward's algorithm has been modified slightly in this study. Notice that this algorithm was originally designed to determine the precedence order in a set of *solvable* algebraic equations. Thus, the first task was to make a choice of one output variable for each equation such that no variable becomes the output of more than one equation, i.e. the outputs of all equations were chosen simultaneously in the beginning. Since this may not be possible in the present case, they are selected *one at a time* whenever an output is called for in the algorithm. For convenience, the original algorithm will be referred to as Algorithm I and the revised version will be named Algorithm II later in this paper. After applying Algorithm II to the augmented equations, three possible outcomes may be found:

- *Outcome* (i): The algorithm fails before deleting all variables. In the remaining system, the number of undeleted variables should be always larger than that of the equations in which these variables appear.
- *Outcome* (ii): All variables and equations are deleted.
- *Outcome* (iii): At least one equation remains after deleting all variables.

Outcome (i) implies that there are extra degrees of freedom for the variables and thus more than one set of parameter values can be identified. Outcome (ii) is an indication that the number of variables equals that of the constraints and outcome (iii) means that the former is less than the latter. In both cases, the optimal estimates produced by an EKF should also be the correct estimates.

The various testing techniques described above can be integrated into a single procedure. Since only structural information of the system model is required, implementation of this procedure should be simple and easy. Its specific steps are outlined below.

Procedure A

1. Construct a structural matrix according to the system model and the model parameter.
2. Tear the measurement variables from the structural matrix, i.e. remove all elements except one in each of the corresponding columns.
3. Apply Algorithm I to the post-tearing matrix to obtain the corresponding precedence order.
4. Identify the tunable parameters and the affected variables. If intunable parameters exist, then the system is diagnostically unobservable and the procedure should be terminated. Otherwise, go to the next step.
5. Remove the rows corresponding to the model parameters in the post-tearing matrix and add rows associated with the affected variables to form an augmented matrix.
6. Perform Algorithm II on the augmented matrix.
   (a) If outcome (i) is found, then the system is diagnostically unobservable.
   (b) If outcome (ii) or (iii) is found, then the system is diagnostically observable.

It has to be emphasized that, although structural analysis is qualitative in nature and thus not theoretically rigorous, the correctness of its predictions has been verified in numerous simulation studies. Let us now use the system described in Fig. 4(a) as an example again. Assume that there are five possible faults: (i) a sudden change in the inlet flow rate of tank T_1 (Δq_i); (ii) a leak in tank T_1 (Δcl_1); (iii) partial blockage in the pipeline between tanks T_1 and T_2 (Δf_1); (iv) a leak in T_2 (Δcl_2); (v) partial blockage in the exit pipeline of tank T_2 (Δf_2). Also assume that the probability of simultaneous occurrence of three or more faults is very low and therefore can be neglected. The outcomes of EKF estimation can be predicted easily with the proposed procedure. The results associated with all possible binary combinations of faults are presented in Table 1. Each row in this table corresponds to the set of measurement variables indicated

Table 1. The diagnostic observability of various EKFs for the two-tank system shown in Fig. 4

	Δcl_1 Δf_1	Δcl_1 Δcl_2	Δcl_1 Δf_2	Δcl_1 Δq_i	Δf_1 Δcl_2	Δf_1 Δf_2	Δf_1 Δq_i	Δcl_2 Δf_2	Δcl_2 Δq_i	Δf_2 Δq_i
h_1, q_1	√	√	√	★	★	★	√	★	√	√
h_1, h_2	√	√	√	★	√	√	√	★	√	√
h_1, q_2	√	√	√	★	√	√	√	√	√	√
q_1, h_2	★	√	√	★	√	√	★	★	√	√
q_1, q_2	★	√	★	★	√	√	★	√	√	√
h_2, q_2	★	★	√	★	★	√	★	√	★	√
h_1, q_1, h_2	√	√	√	★	√	√	√	★	√	√
h_1, q_1, q_2	√	√	√	★	√	√	√	√	√	√
h_1, h_2, q_2	√	√	√	★	√	√	√	√	√	√
q_1, h_2, q_2	★	√	√	★	√	√	★	√	√	√

in the first cell on the left and each column corresponds to the model parameters indicated in the top cell. The symbol "√" denotes a diagnostically observable system and the symbol "★" means that the corresponding system should be diagnostically unobservable. It must be noted that *every* entry in this table has been confirmed with numerical simulation. *All* predictions have been shown to be accurate. Also, the correctness of the suggested approach has been tested with numerous other system models as well. Due to the limitation of space, these studies are not reported here.

THE COMPOSITE PARAMETER ESTIMATION SCHEME

As indicated in the previous discussions, in designing an EKF-based fault monitoring system, one of the basic requirements is that the system must be diagnostically observable with respect to the model parameters included in the state estimate propagation equations. From structural analysis one can also conclude that the number of measurement variables s should be no less than that of the *all* possible faults p if only *one* EKF is adopted for diagnosis purpose. However, due to economical or technical reasons, it may be difficult to satisfy this condition when the number p is large. This situation can be improved with the composite parameter estimation scheme suggested below.

If one can assume that the probability of coexistence of more than p^* ($p^* < p$) faults is extremely low and thus negligible, then it is usually possible to construct several parallel EKFs, each with fewer than p measurement variables, to trace the variations in all combinations of p^* parameters changing simultaneously. The measurement data used in all these EKFs are the same, but the assumed faults in each of them are not identical to those in any other filter. Of course, the number of measurement variables s should always be larger than or equal to that of the augmented parameters in every EKF, i.e. $s \geq m$ and m is the number of augmented parameters, and the latter must be no less than the number p^*, i.e. $m \geq p^*$. A simple example is used here to illustrate its detailed implementation procedure and to demonstrate the advantages of adopting such an approach.

Example 6

Let us now consider only the first tank described in Fig. 2. Assume in this case that $p = 3$ and the three possible faults are: (i) a sudden change in the inlet flow rate; (ii) a leak in the tank; (iii) partial blockage in the outlet pipeline. The corresponding parameters are Δq_i, Δcl and Δf, respectively. Also, let us assume that the probability of simultaneous occurrence of two or more faults is negligible, i.e. $p^* = 1$, and the measurement variables are h and q, i.e. $s = 2$.

If the traditional approach is taken, all three parameters Δq_i, Δcl and Δf must be augmented with the state variables h and q in one EKF, as shown in

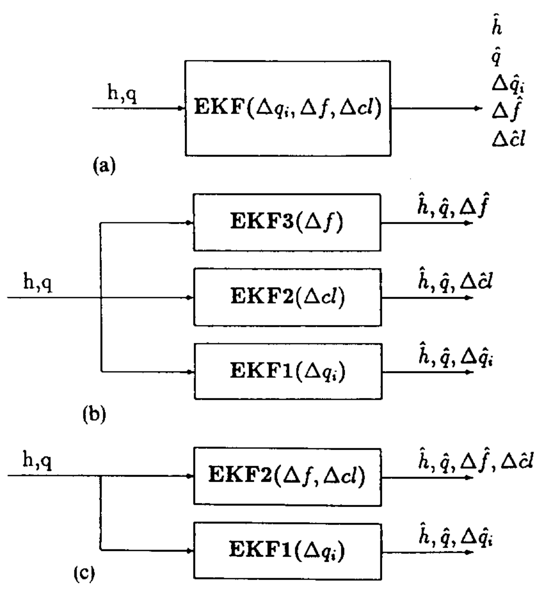


Fig. 10. (a) The conventional parameter estimation scheme for the one-tank system in example 6. (b) The parallel parameter estimation scheme for the one-tank system in example 6. (c) A composite parameter estimation scheme for the one-tank system in example 6.

Fig. 10(a). Thus, $m = 3$ in this case. It can be easily demonstrated with structural analysis that this EKF is diagnostically unobservable. Indeed, the results of simulation studies also support this prediction. The estimates of the state variables and model parameters corresponding to fault (iii) are presented in Fig. 11(a)–(e). It is apparent that these results are erroneous.

According to Chang *et al.* (1993), the above drawback can be remedied with several single-parameter EKFs implemented in parallel. Notice that, in this case, $m = p^* = 1$. The corresponding parameter estimation scheme is shown in Fig. 10(b). Using the procedure outlined in the previous section, it can be verified that all three EKFs are diagnostically observable. Further, a comparison of the estimates generated

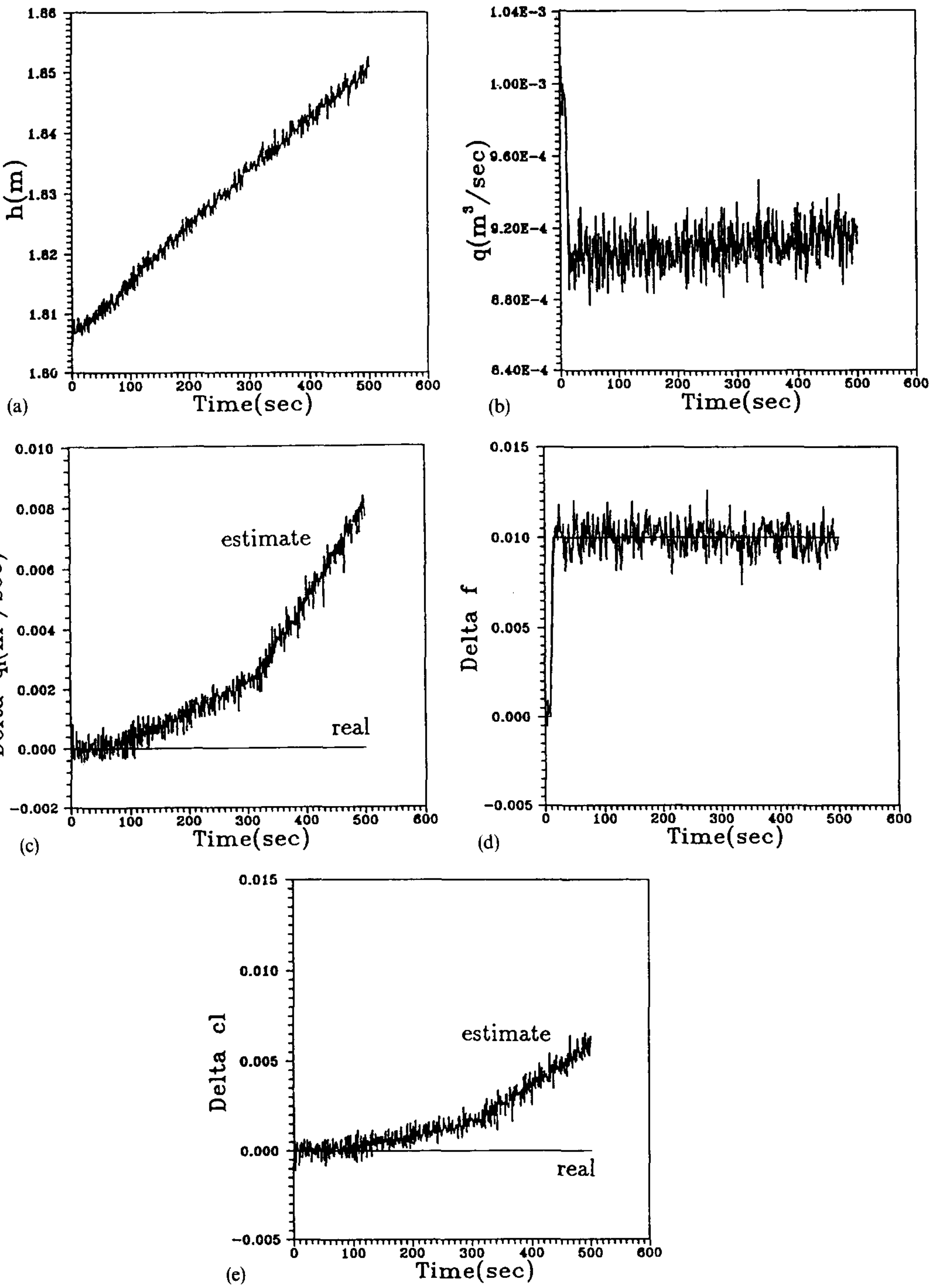


Fig. 11. Estimates generated by a conventional extended Kalman filter for the one-tank system in: (a) example 6 - h; (b) example 6 - q; (c) example 6 - Δq_i; (d) example 6 - Δf; (e) example 6 - Δcl.

from these EKFs and the on-line measurement data offers additional clues for identifying the correct fault origin. Again, simulation results corresponding to fault (iii) are used here to demonstrate the advantage of this approach [see Fig. 12(a)–(i)]. One can observe from Fig. 12(b) and (e) that the estimates of q produced in EKF1 and EKF2 deviate significantly from the measurement data. Consequently, the possibility of faults (i) and (ii) should be excluded from consideration. On the other hand, from the fact that the estimates of both measurement variables in EKF3 follow the transient on-line data closely, fault (iii) can be regarded as a candidate cause of the abnormal system symptoms. Since this fault, i.e. partial blockage in the outlet pipeline, is the only candidate, one can thus correctly conclude that it is the fault origin.

One of the drawbacks of the above parameter estimation scheme is that its computation load is usually very heavy. This problem is mainly due to the fact that a large number of EKFs must be implemented in parallel. In the present study, it has been shown that the amount of parallel calculations can be reduced by grouping some of the EKFs together. This modified configuration is referred to as the *composite* parameter estimation scheme. For example, the scheme in Fig. 10(b) can be changed to the one presented in Fig. (10c). Notice that it is not necessary for all EKFs to augment the same number of parameters. Also notice that, although $m > p^*$ in EKF2, it is still diagnostically observable. The same set of simulation data, i.e. those corresponding to fault (iii), has been used to assess the feasibility of this approach. Since the estimates of EKF1 here should be exactly the same as those obtained with the previous scheme in Fig. 10(b), only the results associated with EKF2 are provided in Fig. 13(a)–(d). One can observe from these results that (1) the estimates of h and q are in close agreement with their measurement values, (2) the estimates of Δf indicate that fault (iii) is a possible candidate and (3) the estimates of Δcl remain approximately unchanged at zero. Naturally, the diagnosis in this case must be directed toward the only possibility, i.e. partial blockage in the outlet pipeline. Thus, when compared with the configuration shown in Fig. 10(b), the composite parameter estimation scheme is superior due to the fact that it can be used to achieve the

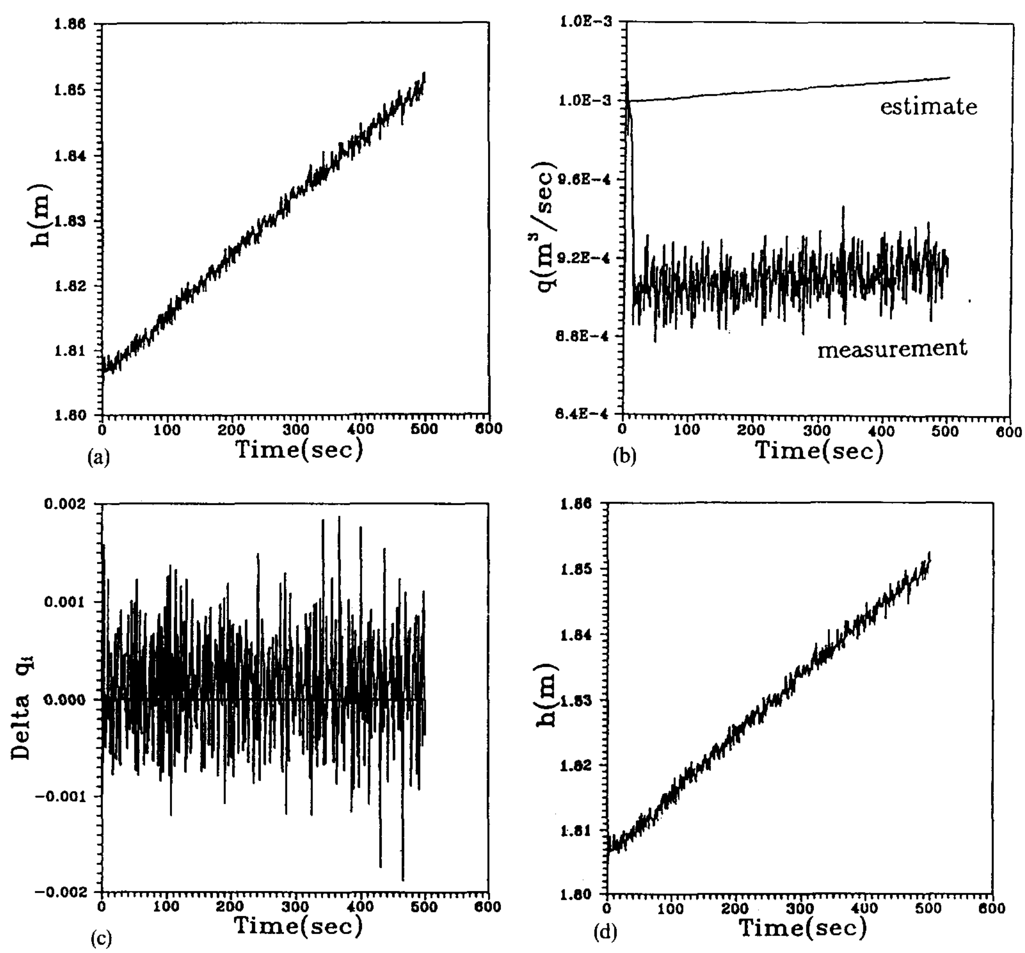


Fig. 12. Estimates generated, in the parallel parameter estimation scheme for the one-tank system in example 6, by EKF1 for (a) h, (b) q, (c) Δq_i, by EKF2 for (d) h, (e) q, (f) Δcl, and by EKF3 for (g) h, (h) q, and (i) Δf.

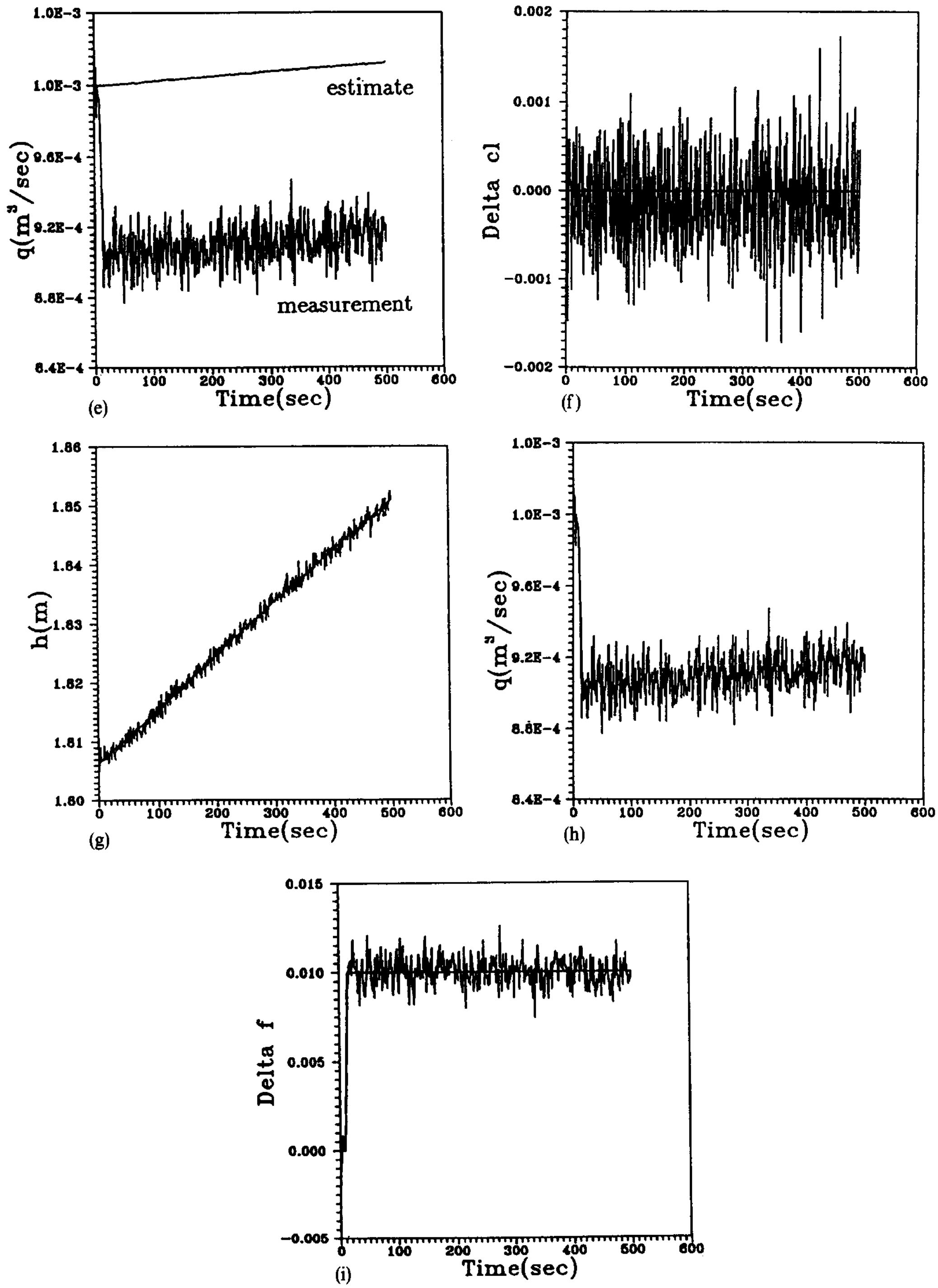


Fig. 12. (e)–(i).

same level of performance with less computation effort. In addition, there is one other advantage which is probably worth mentioning. Notice that the possibility of simultaneous occurrence of multiple faults was ignored completely in the previous studies, e.g. Chang *et al.* (1993), and this issue can be addressed readily within the framework of the composite parameter estimation scheme.

DIAGNOSTIC RESOLUTION

In the previous section, although we have established that the composite parameter estimation scheme does possess certain desirable features, the question of *how* to select the augmented parameters in each EKF is still unanswered. Nevertheless, it is quite evident that, if these selections are to be made pro-

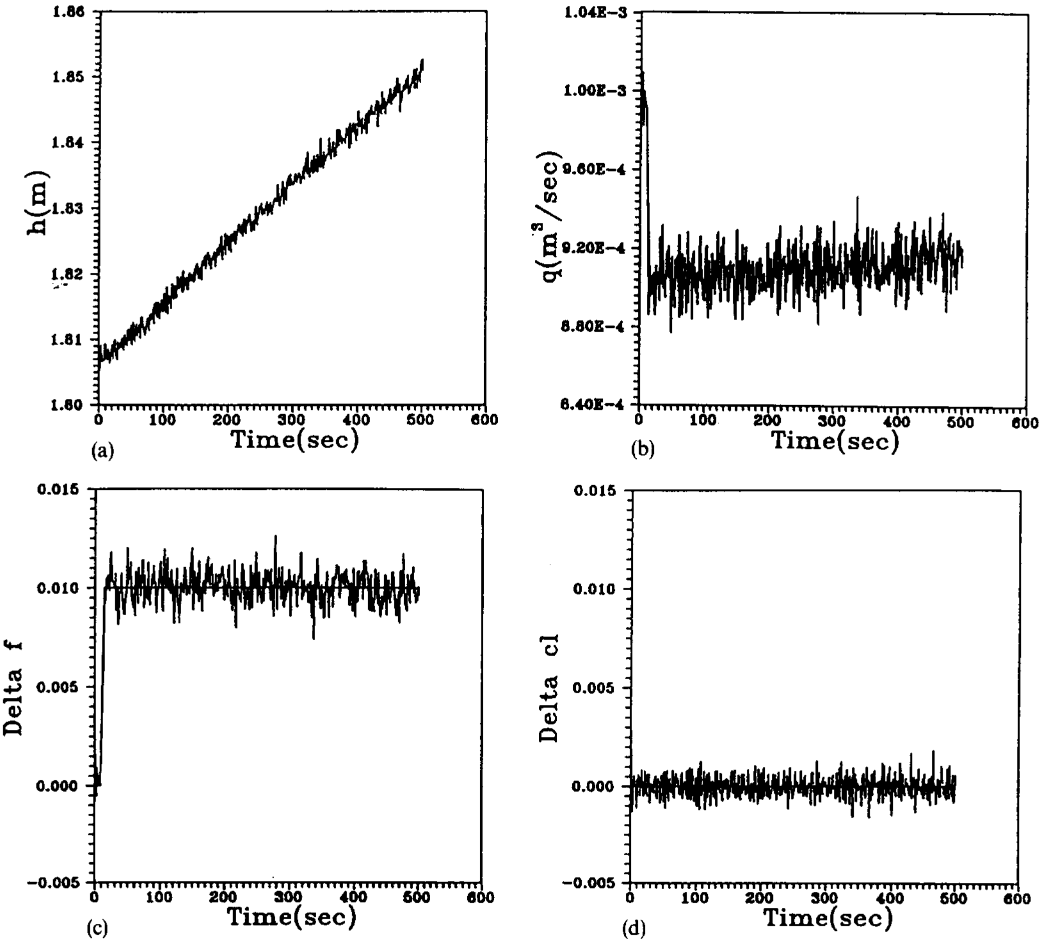


Fig. 13. Estimates generated by EKF2 in the composite parameter estimation scheme shown in: (a) Fig. 12(c) - h; (b) Fig. 12(c) - q; (c) Fig. 12(c) - Δf; (d) Fig. 12(c) - Δcl.

perly, a set of logical criteria must first be developed to assess the performance of any given scheme.

Apparently, *diagnostic observability* must be adopted as one of the selection criteria. This requirement implies that, if the occurrence of faults causes a subset of the augmented parameters in one of the EKFs to vary, the corresponding estimates should be correct. However, in the fault diagnosis process, one naturally does not have a prior knowledge about the actual causes of abnormal system behavior. As a result, one is bound to face the task of analyzing the results produced by EKFs in which some of the fault-affected parameters are *not* augmented. Since there is a mismatch between the assumed model and the actual system behavior in each of these EKFs, these results should be disregarded. Also, as indicated before, mismodelling is one of the main causes of divergence, i.e. type (v) cause. From the standpoint of fault diagnosis, divergence in this situation is in fact desirable because model mismatch can be easily detected by comparing the measurement data and the estimates. However, this desirable feature cannot be guaranteed under all circumstances. Misdiagnosis may therefore be possible if diagnostic observability is the only restriction imposed upon the composite parameter estimation scheme.

Diagnostic resolution is the second criterion proposed in this work for evaluating the performance of a fault-monitoring system. In the literature, this term usually refers to a measure of uniqueness in identifying the correct fault origin among all possible candidates (Kramer and Polawitch, 1987). In order to represent such an abstract concept explicitly, *a diagnostic performance table* has been designed. This table can be best illustrated with an example.

Example 7

The single-tank system described in example 6 is adopted here. In the present example, assume that liquid level h is the only measurement variable and that the parallel scheme in Fig. 10(b) is adopted for diagnosing faults. Assume also that the faults that may occur in this system are the same as those described in example 6. Let us again consider the results of EKF estimation when fault (iii), i.e. partial blockage in the outlet pipeline, occurs. It can be demonstrated with simulation studies that the estimates of h follow the measurement values closely in all three EKFs.

Consequently, none of the candidates can be excluded from the list of potential fault origins and misdiagnosis becomes inevitable due to low resolution. Conclusions associated with all other possible scenarios can also be obtained in a similar fashion by studying the corresponding simulation data. These outcomes of fault diagnosis can then be summarized in a diagnostic performance table as shown in Table 2a.

In general, the chosen measurement variables are indicated at the upper-left corner of a diagnostic performance table. For each possible combination of actual faults, simulated data can be generated by introducing a change in each of the corresponding parameters. The entries in each row of the table are obtained from the same set of simulated measurement data corresponding to the parameters indicated at the left side of the row. Each entry in this table is a conclusion drawn from analyzing the results of implementing one EKF to the above data. The model parameters augmented in this EKF are indicated on the top of the corresponding column. The symbols used as entries in this table are defined below:

- If *all* the parameters associated with the actual faults are included as the augmented parameters in an EKF, then two types of outcomes are possible, i.e.
  - —$\surd$: The estimates of the augmented parameters and the state variables match their actual transient behaviors, i.e. the EKF must be diagnostically observable.
  - —★: The EKF fails to trace the variations in the augmented parameters, i.e. the EKF is diagnostically unobservable.
- If some of the parameters associated with the actual faults are *not* included as the augmented parameters in an EKF, then the corresponding results can also be classified into two categories:
  - —o: The estimates of the measurement variables follow the on-line data closely. In other words, mismodelling cannot be detected and the corresponding faults cannot be excluded from the list of candidates.
  - —blank: The estimates of some of the state variables do not match their measurement values or the estimates of all augmented parameters remain unchanged. In such cases, mismatch between the actual system behavior and the EKF model can be identified and thus the corresponding results can be disregarded.

The diagnostic error "o" is due to the fact that the effects of the actual faults on the measured state variables can be replaced by those of the faults assumed in the EKF. If the estimates of other unmeasured state variables are different from their actual behaviors, the corresponding errors can often be avoided by redistributing the sensor locations or placing additional measurement points. For instance, if both state variables in the present example are measured, some of the entries "o" can be removed (see Table 2b).

Table 2. Diagnostic performance of a parallel parameter estimation scheme (Fig. 10b)

(a) One-tank system with one measurement variable

h	Δq_i	Δcl	Δf
Δq_i	$\surd$	o	o
Δcl	o	$\surd$	o
Δf	o	o	$\surd$

(b) One-tank system with two measurement variables

h, q	Δq_i	Δcl	Δf
Δq_i	$\surd$	o	
Δcl	o	$\surd$	
Δf			$\surd$

Notice also that several cells in Table 2b are still filled with "o". A detailed analysis of the simulation results reveals that, in these cases, the effects of the actual faults on *all* state variables can be reproduced exactly by the anticipated faults. This special relation exists between the faults associated with Δq_i and Δcl in the present example. In particular, the abnormal transient behaviors of both state variables h and q, which are caused by a decrease in the inlet flow rate, may be mistakenly regarded as due to tank leaks and, similarly, the opposite situation may also occur. From a structural viewpoint, the two parameters Δq_i and Δcl are interchangeable in the system model. As a result, these mistakes cannot be eliminated even by measuring more state variables.

Obviously, an ideal composite parameter estimation scheme should be one associated with a diagnostic performance table that contains only "$\surd$" and blanks. Assuming that informations about all possible combinations of actual faults are available, i.e. the entries in the first column on the left are given in advance, one can develop such schemes by appropriately selecting the measurement variables and the augmented parameters in EKFs so that the number of "★" and "o" can be minimized. In this research, diagnostic observability is considered to be a basic condition that all EKFs in the composite parameter estimation scheme have to satisfy. In addition, these EKFs should be selected in such a way that the parameters associated with each row of the diagnostic performance table must be the subset of the augmented parameters in at least one EKF. As a result, the possibility of "★" can be eliminated completely. On the other hand, although the number of "o" can be reduced by changing the sensor number and/or distribution, some of the corresponding errors may still be unavoidable due to inherent structural property of the system model. Thus, diagnostic performance is essentially dependent only upon resolution, i.e. the degree of uniqueness achieved in diagnosis, which is considered in this work to be proportional to the number of blanks in the diagnostic performance table.

CONSTRUCTION OF THE DIAGNOSTIC PERFORMANCE TABLE

As indicated previously, the diagnostic performance table can be considered as a performance measure of any EKF-based parameter estimation scheme and thus can be adopted as an aid in decision making when several competing designs are available. Although this table can be constructed by analyzing the simulation data, the numerical effort for generating such data is often overwhelming. In order to overcome this difficulty, it has been our intention to develop a simple procedure for predicting the outcomes of fault diagnosis with only *structural* information of the system model.

Since every EKF in the composite parameter estimation scheme is required to be diagnostically observable, the locations of "$\surd$" in the diagnostic performance table can be determined by comparing the fault-affected parameters with the augmented parameters in each EKF. Specifically, if the former forms a subset of the latter, the corresponding cell should be filled with a "$\surd$".

The other entries in the table must be either "o" or blank. Notice that the augmented parameters associated with each EKF in the composite parameter estimation scheme should be considered as the *tunable parameters.* If some of the actual fault-affected parameters are not the tunable parameters associated with an EKF, then the variables affected by the former may not be the same as those by the latter. For convenience, the former affected variables are referred to as class I variables and the latter as class II variables. Naturally, one question arises: can the tunable parameters be properly adjusted so that the resulting values of class I variables can meet the requirements of eqs (14) ? If the answer is yes, then a diagnostic mistake "o" is sure to result. Otherwise, model mismatch can be identified by comparing the estimates and the measurement data. In this work, a systematic procedure has been developed to answer the above question. To facilitate understanding of the proposed procedure, an example is presented as follows.

Example 8

Let us now consider the two-tank system described in example 3. Assume that the measurement variables selected in this system are h_1, h_2 and q_2. Three representative cases can be identified:

- Assume that two faults, i.e. (i) a leak in tank T_1 and (ii) partial blockage in the pipeline between T_1 and T_2, occur simultaneously during operation. Thus, the fault-affected parameters should be Δcl_1 and Δf_1, respectively. If the augmented parameters of one of the EKFs in the composite parameter estimation scheme are Δcl_1 and Δcl_2, then they should be regarded as the tunable parameters for this particular EKF.

  After applying Algorithm I to the post-tearing structural matrix, the precedence order presented in Fig. 14(a) can be obtained. Notice that, for illustration purpose, both fault-affected parameters and tunable parameters are included in Fig. 14(a). The former parameters are marked with asterisks in this figure. One can easily identify from this precedence order that class I and class II variables are identical, i.e. h_1 and h_2. As indicated before, the values of class I variables h_1 and h_2 must be forced to meet the requirements of eqs (14) by properly adjusting the tunable parameters Δcl_1 and Δcl_2. Since there are two parameters and two constraints, convergence in EKF estimation can generally be expected. As a result, the diagnostic mistake "o" should be unavoidable.
- Next, let us consider the case when the same EKF is used to process the transient data caused by simultaneous occurrence of two abnormal events, i.e. (i) a leak in tank $T_1(\Delta cl_1)$ and (ii) partial blockage in the exit pipeline of tank $T_2(\Delta f_2)$. From the corresponding precedence order [Fig. 14(b)], one can identifiy that h_1 and q_2 are class I variables and h_1 and h_2 are class II variables.

  Since q_2 is not connected to any of the tunable parameters, it is apparently not possible to trace

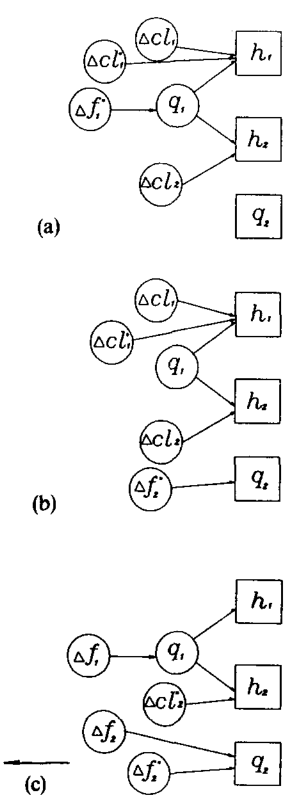


Fig. 14. The post-tearing precedence diagram of the two-tank system in example 8. (a) Case 1, (b) case 2, (c) case 3.

its changes with the present EKF. Thus, one can conclude that the estimates of the measurement variables should diverge and the corresponding entry in the diagnostic performance table must be a blank.

- If the actual faults are (i) a leak in tank T_2 (Δcl_2) and (ii) partial blockage in the exit pipeline of T_2 (Δf_2) and the tunable parameters associated with another EKF are Δf_1 and Δf_2, the precedence order shown in Fig. 14(c) can be determined. Thus, h_2 and q_2 are class I variables and h_1, h_2 and q_2 are class II variables in this case, i.e. class I is a subset of class II. One can also see that, by means of adjusting the tunable parameter Δf_1, the value of the affected variable h_2 cannot be varied independently without causing a change in the variable h_1 as well. In other words, one should actually be concerned about all constraints associated with class II variables in this situation. Since the total number of constraints is greater than that of the tunable parameters, the resulting EKF estimates must diverge and the corresponding entry in the diagnostic performance table should be a blank.

It should be emphasized that all predictions presented in example 8 have been verified with thorough numerical simulation studies. For the sake of brevity, the results of these studies are not included in this paper. From the three different cases discussed in this example, one can see that the precedence order obtained after the tearing operation is quite helpful in securing a solution to our problem at hand, i.e. to determine whether class I variables can be forced to meet the requirements of eqs (14) by adjusting the tunable parameters. Generally speaking, the diagnostic error "o" is committed if

- class I variables are a subset of class II variables, and
- the number of constraints associated with class II variables is not greater than that of the tunable parameters.

Otherwise, the corresponding entry in the diagnostic performance table should be blank. Notice that Algorithm II can be adopted to verify the validity of the second condition. Specifically, outcomes (i) and (ii) are indications of a model structure which facilitates convergence in EKF estimation and thus should be assigned a "o" and, on the other hand, outcome (iii) should be associated with a blank in the table.

Under the assumptions that all possible combinations of faults are given and that all the EKFs chosen in the composite scheme are diagnostically observable, the techniques introduced in this section have been integrated into a procedure for constructing the corresponding diagnostic performance table. The specific steps to determine the entry in each cell of the table are outlined below.

Procedure B

1. Compare the fault-affected parameters with the augmented parameters associated with the selected cell. If the former does not form a subset of the latter, go to the next step. Otherwise, fill the cell with a "$\surd$".
2. Determine the class I variables.
   - Construct a structural matrix according to the system model and the fault-affected parameters.
   - Tear the measurement variables from the structural matrix, i.e. remove all but one element in each of the columns corresponding to the measurement variables.
   - Apply Algorithm I on the post-tearing matrix to obtain the corresponding precedence order and identify accordingly the class I variables, i.e. the measurement variables which are connected to the fault-affected parameters.
3. Attach the tunable parameters, i.e. the augmented parameters of the EKF under consideration, to the above precedence order and identify the class II variables, i.e. the measurement variables connected to the tunable parameters.
4. If some of the class I variables are not class II variables, then the cell should be blank. Otherwise, go to the next step.
5. Build an agumented matrix corresponding to the tunable parameters and class II variables.
   - Remove the rows and columns corresponding to the fault-affected parameters in the post-tearing matrix.
   - Add columns associated with the augmented parameters of the EKF under consideration.
   - Add rows corresponding to class II variables.
6. Perform Algorithm II on the augmented matrix.
   - If outcome (i) or (ii) is found, then the cell should be filled with a "o".
   - If outcome (iii) is found, then the cell should be left as blank.

APPLICATION EXAMPLES

Two application examples of the diagnostic performance table are provided here. The first is concerned with the simple two-tank system which has been discussed repeatedly throughout this paper and the other is a complex two-phase reaction process. The former example is designed to illustrate the implementation procedure and to show the usefulness of the information generated with the proposed approach. The latter is meant to demonstrate the practical value of our techniques and the potential for other realistic applications.

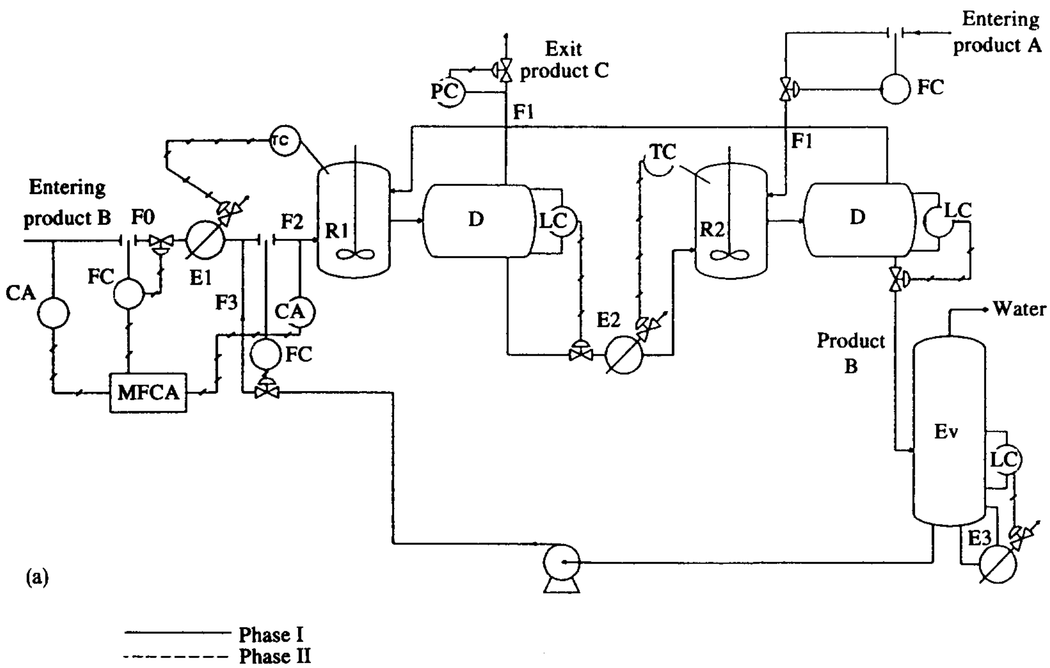
Example 9

Again the system described in Fig. 4(a) is considered here. Let us assume that there are five (5) possible abnormal events, i.e. a leak in tank T_1 or T_2, partial blockage in the pipeline between T_1 and T_2 or the exit pipeline of T_2, or a sudden change in the inlet flow rate of T_1. Their corresponding model parameters are Δcl_1, Δcl_2, Δf_1, Δf_2, and Δq_i, respectively. Let us neglect the possibility of three or more events occurring simultaneously. In other words, $p = 5$ and $p^* = 2$. All diagnostically observable EKFs with two parameters, i.e. $m = 2$, have been selected in the parameter estimation scheme. Notice that, due to the requirement for diagnostic observability, the EKF which incorporates the augmented parameters Δq_i and Δcl_1 are not included in the scheme.

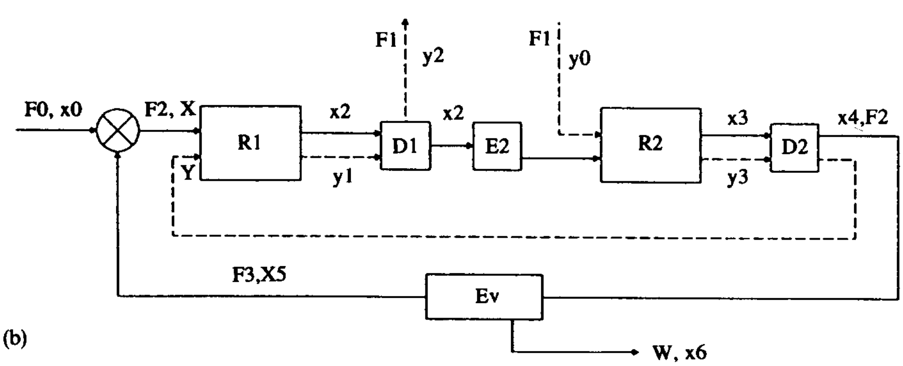
Let us first consider the case when h_1 and q_2 are chosen as the measurement variables, i.e. $s = 2$. The resulting diagnostic performance table (Table 3a) can be obtained easily with Procedure B. Notice that none of the entries are blanks. Thus, if diagnostic observability is the only condition imposed upon the parameter estimation scheme, its diagnostic performance may not be acceptable.

The resolution of diagnosis can be improved by adding more measurements. As shown in Table 3b, the number of misdiagnosis, i.e. "o", can be reduced significantly by introducing an additional measurement variable h_2. On the other hand, if h_2 is replaced by another measurement variable q_1, then the diagnostic performance can be enhanced further. The number of "o" is decreased from 34 in Table 3b to 21 (see Table 3c). As a matter of fact, the diagnostic resolution associated with this choice, i.e. h_1, q_1 and q_2, is the highest among all four possible combinations of three measurement variables. Notice also that, although the undesirable entries "o" cannot be eliminated completely in this case, almost all combinations of faults can be uniquely identified with the aid of *fault patterns*. One can observe from Table 3c that there are in fact only two groups of indistinguishable fault origins, i.e. (i) those associated with rows 4, 5 and 7 and (ii) those corresponding to rows 10 and 11.

Finally, it must be noted that *every* entry in the above three tables has been confirmed with numerical



(a)



(b)

Fig. 15. (a) The process flow diagram of a two-phase reaction process. (b) The process block diagram of a two-phase reaction process.

Table 3. Diagnostic performance of a composite parameter-estimation scheme
(a) Two-tank system with two measurement variables

h_1, q_2	Δcl_1 Δf_1	Δcl_1 Δcl_2	Δcl_1 Δf_2	Δf_1 Δcl_2	Δf_1 Δf_2	Δf_1 Δq_i	Δcl_2 Δf_2	Δcl_2 Δq_i	Δf_2 Δq_i
$\Delta cl_1, \Delta f_1$	√	o	o	o	o	o	o	o	o
$\Delta cl_1, \Delta cl_2$	o	√	o	o	o	o	o	o	o
$\Delta cl_1, \Delta f_2$	o	o	√	o	o	o	o	o	o
$\Delta f_1, \Delta cl_2$	o	o	o	√	o	o	o	o	o
$\Delta f_1, \Delta f_2$	o	o	o	o	√	o	o	o	o
$\Delta f_1, \Delta q_i$	o	o	o	o	o	√	o	o	o
$\Delta cl_2, \Delta f_2$	o	o	o	o	o	o	√	o	o
$\Delta cl_2, \Delta q_i$	o	o	o	o	o	o	o	√	o
$\Delta f_2, \Delta q_i$	o	o	o	o	o	o	o	o	√
Δq_i	o	o	o	o	o	√	o	√	√
Δcl_1	√	√	√	o	o	o	o	o	o
Δf_1	√	o	o	√	√	√	o	o	o
Δcl_2	o	√	o	√	o	o	√	√	o
Δf_2	o	o	√	o	√	o	√	o	√

(b) Two-tank system with three measurement variables—(i)

h_1, h_2, q_2	Δcl_1 Δf_1	Δcl_1 Δcl_2	Δcl_1 Δf_2	Δf_1 Δcl_2	Δf_1 Δf_2	Δf_1 Δq_i	Δcl_2 Δf_2	Δcl_2 Δq_i	Δf_2 Δq_i
$\Delta cl_1, \Delta f_1$	√	o		o		o		o	
$\Delta cl_1, \Delta cl_2$	o	√		o		o		o	
$\Delta cl_1, \Delta f_2$			√						o
$\Delta f_1, \Delta cl_2$	o	o		√		o		o	
$\Delta f_1, \Delta f_2$					√				
$\Delta f_1, \Delta q_i$	o	o		o		√		o	
$\Delta cl_2, \Delta f_2$							√		
$\Delta cl_2, \Delta q_i$	o	o		o		o		√	
$\Delta f_2, \Delta q_i$			o						√
Δq_i	o	o	o	o		√		√	√
Δcl_1	√	√	√	o		o		o	o
Δf_1	√	o		√	√	√		o	
Δcl_2	o	√		√		o	√	√	
Δf_2			√		√		√		√

(c) Two-tank system with three measurement variables—(ii)

h_1, q_1, q_2	Δcl_1 Δf_1	Δcl_1 Δcl_2	Δcl_1 Δf_2	Δf_1 Δcl_2	Δf_1 Δf_2	Δf_1 Δq_i	Δcl_2 Δf_2	Δcl_2 Δq_i	Δf_2 Δq_i
$\Delta cl_1, \Delta f_1$	√					o			
$\Delta cl_1, \Delta cl_2$		√						o	
$\Delta cl_1, \Delta f_2$			√						o
$\Delta f_1, \Delta cl_2$				√	o		o		
$\Delta f_1, \Delta f_2$				o	√		o		
$\Delta f_1, \Delta q_i$	o					√			
$\Delta cl_2, \Delta f_2$				o	o		√		
$\Delta cl_2, \Delta q_i$		o						√	
$\Delta f_2, \Delta q_i$			o						√
Δq_i	o	o	o			√		√	√
Δcl_1	√	√	√			o		o	o
Δf_1	√			√	√	√	o		
Δcl_2		√		√	o		√	√	
Δf_2			√	o	√		√		√

simulation. The predictions are all correct without exception.

Example 10

Let us consider the process shown in Fig. 15(a) in which reactants A and B are converted to product C by the reaction

$$A + B \rightarrow C.$$

The reactants are fed to the process in two different immiscible phases (A is in phase I and B is in phase II), and the reaction takes place in the mixer of the mixer–settler pairs. The objective of the process is to convert a fixed percentage of A into product C. The mathematical model of this system can be found in Himmelblau and Bischoff (1968):

$$(1-\alpha)V\frac{\mathrm{d}y_1}{\mathrm{d}t} = F_1(Y - y_1) - k_1\alpha V x_1 y_1 \quad (22)$$

$$\alpha V\frac{\mathrm{d}x_1}{\mathrm{d}t} = F_0 x_0 + F_3 x_5 - F_2 x_1 - 2k_1\alpha V x_1 y_1 \quad (23)$$

$$V_1'\frac{\mathrm{d}y_2}{\mathrm{d}t} = F_1(y_1 - y_2) \quad (24)$$

$$V_2'\frac{\mathrm{d}x_2}{\mathrm{d}t} = F_2(x_1 - x_2) \quad (25)$$

$$V_3\frac{\mathrm{d}\xi}{\mathrm{d}t} = F_2(x_2 - \xi) \quad (26)$$

$$(1-\alpha)V\frac{\mathrm{d}y_3}{\mathrm{d}t} = F_1(y_0 - y_3) - k_2\alpha V x_3 y_3 \quad (27)$$

$$\alpha V\frac{\mathrm{d}x_3}{\mathrm{d}t} = F_2(\xi - x_3) - 2k_2\alpha V x_3 y_3 \quad (28)$$

$$V_1'\frac{\mathrm{d}Y}{\mathrm{d}t} = F_1(y_3 - Y) \quad (29)$$

$$V_2'\frac{\mathrm{d}x_4}{\mathrm{d}t} = F_2(x_3 - x_4) \quad (30)$$

$$(v_0 + S\lambda)\frac{\mathrm{d}x_5}{\mathrm{d}t} = F_2(x_4 - x_5) - W[(\alpha_1 - 1)x_5 + \beta] \quad (31)$$

$$S\frac{\mathrm{d}\lambda}{\mathrm{d}t} = F_2 - W - F_3 \quad (32)$$

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \frac{K_1}{S}(F_2 - W - F_3) + K_2\lambda \quad (33)$$

$$\frac{\mathrm{d}F_0}{\mathrm{d}t} = \frac{F_2 - F_0}{x_0 - X}\frac{\bar{X} - X}{\tau} \quad (34)$$

$$\frac{\mathrm{d}F_3}{\mathrm{d}t} = -\frac{\mathrm{d}F_0}{\mathrm{d}t}. \quad (35)$$

Most of the notation used in this model is presented in the block diagram of the process shown in Fig. 15(b). Additional explanations can also be found in the Notation section at the end of this paper.

To verify the correctness of proposed procedures, it is necessary to carry out extensive simulation studies corresponding to every entry in the diagnostic performance table. For the purpose of cutting down the computational load, only four parameters, x_0, F_1, y_0 and k_2, were allowed to vary in this example. The rest of the parameters were assumed to be constants and their values can be found in Himmelblau and Bischoff (1968). It was also assumed that the probability of three or more faults occurring simultaneously in this process is very low and thus can be excluded from consideration. As a result, the possible abnormal system transients are associated with 10 combinations of parameters, i.e. $(\Delta x_0, \Delta F_1)$, $(\Delta x_0, \Delta y_0)$, $(\Delta x_0, \Delta k_2)$, $(\Delta F_1, \Delta y_0)$, $(\Delta F_1, \Delta k_2)$, (Δx_0), (ΔF_1), (Δy_0) and (Δk_2). To cover all 10 possibilities, the first six pairs of the above parameters were augmented in six separate EKFs. These EKFs were found to be diagnostically observable using Procedure A. Procedure B was then applied to determine the entries in the diagnostic performance table (see Table 4). Notice that, with this parameter estimation scheme, any of the possible

Table 4. Diagnostic performance of a composite parameter estimation scheme in a two-phase reaction process with six measurement variables

y_2, F_0 F_3, x_5 λ, Y	ΔF_1 Δx_0	ΔF_1 Δy_0	ΔF_1 Δk_2	Δx_0 Δy_0	Δx_0 Δk_2	Δy_0 Δk_2
$\Delta F_1, \Delta x_0$	√					
$\Delta F_1, \Delta y_0$		√				
$\Delta F_1, \Delta k_2$			√			
$\Delta x_0, \Delta y_0$				√		
$\Delta x_0, \Delta k_2$					√	
$\Delta y_0, \Delta k_2$						√
ΔF_1	√	√	√			
Δx_0	√			√	√	
Δy_0		√		√		√
Δk_2			√		√	√

combinations of faults can be uniquely identified. It should also be emphasized that the scenarios corresponding to *all* entries in the diagnostic performance table have been simulated numerically. It was found that *every* prediction is correct and *no* exceptions can be found.

CONCLUSIONS

Several important issues concerning the implementation of EKF-based fault diagnosis techniques are addressed in this paper. First, the concept of *diagnostic observability* is introduced to characterize the phenomenon of bias in simultaneous state and parameter estimation. Second, on the basis of structural analysis of the system model, a testing procedure is outlined for identifying diagnostically observable EKFs. Third, a composite parameter estimation scheme, which incorporates several EKFs in parallel, is proposed to overcome the problem of insufficient measurement data. Then, the concept of *diagnostic resolution* is developed as a criterion for selecting the best scheme among competing candidates and a simple tool for its representation and quantification, i.e. the diagnostic performance table, is also described. Finally, a systematic procedure for constructing this table is presented in detail.

It must be emphasized that, although the approach adopted in this study is not theoretically rigorous, the correctness of its results has been supported by numerous simulation studies. Since only structural information of the system model is required, application of the above procedures should be simple and easy. One can thus conclude that the proposed techniques are practical even for large and complex chemical processes.

Acknowledgement—This work is supported by the National Science Council of the ROC under grant NSC84-2214-E006-008.

NOTATION

A_i	the cross-sectional area of the tank i (example 2–example 9)
cl_i	the parameter that characterizes tank i leaks (example 2–example 9)
d_i	the diameter of pipeline i (example 2–example 9)
f_i	the friction factor for the flow in pipeline i (example 2–example 9)
F_1	flow rate of phase I (example 10)
F_0, F_2, F_3	flow rates of phase II (example 10)
h_i	the height of liquid level in tank i (example 2–example 9)
l_i	the length of pipeline i (example 2–example 9)
k_1, k_2	reaction-rate constants in reactors and 2 (example 10)
K_1, K_2	adjustable constants in the level controller of evaporator (example 10)
q_i	the volumetric flowrate in pipeline i (example 2–example 9)
S	cross-sectional area of the evaporator (example 10)
t	time
X	concentration of B entering reactor $R_1(X = F_0x_0 + F_3x_5/F_2)$ (example 10)
$\bar{X}$	reference value of X (example 10)
x_0	entering concentration of B in phase II (example 10)
Y	concentration of A entering reactor 1 (example 10)
y_0	entering concentration of A is phase I (example 10)
W	flow rate of vapour from the evaporator (example 10)
V	total volume of reactors 1 and 2 (example 10)
v_0	total volume of liquid in the evaporator and in the heat exchanger E_3 (example 10)
V_3	volume of heat exchanger E_2 (example 10)
V'	total volume of decanters 1 and 2 (example 10)
V_1', V_2'	volume of the individual decanter $(V_1' = V_2' = V'/2)$ (example 10)

Greek letters

α	volume fraction occupied by phase II in each reactor (example 10)
α_1, β	constants in the model of evaporator $(x_6 = \alpha_1 x_5 + \beta)$ (example 10)
λ	variation from the reference level in the evaporator (example 10)
ξ	concentration of B leaving exchanger E_2 (example 10)
ρ	the density of liquid (example 2–example 9)
τ	time constant of the multifunctional controller (example 10)

REFERENCES

Chang, C. C. and Yu, C. C., 1990, On-line fault diagnosis using the signed directed graph. *Ind. Engng Chem. Res.* **29**, 1290.

Chang, C. T., Mah, K. N. and Tsai, C. S., 1993, A simple design strategy for fault monitoring systems. *A.I.Ch.E. J.* **39**, 1146.

Chen, J. W., 1994, Studies on fault monitoring techniques—implementation problems of the extended Kalman filter. MS Thesis, National Cheng Kung University, Tainan, Taiwan, ROC.

Christensen, J. H., 1970, Structuring of process optimization. *A.I.Ch.E. J.* **16**, 177.

Christensen, J. H. and Rudd, D. F., 1969, Structuring design computations. *A.I.Ch.E. J.* **15**, 94.

Dalle Molle, D. T. and Himmelblau, D. M., 1987, Fault detection in a single-stage evaporator via parameter estimation using the Kalman filter. *Ind. Engng Chem. Res.* **26**, 2482.

Davis, J. F., Shum, S. K., Chandrasekaran, B. and Punch, W. F., III, 1987, A task-oriented approach to malfunction diagnosis in complex processing plants. NSF–AAAI workshop on AI in Process Engineering, Columbia University, New York.

Gelb, A., 1974, *Applied Optimal Estimation.* MIT Press, Cambridge, MA.

Grewal, M. S. and Andrews, A. P., 1993, *Kalman Filtering—Theory and Practice.* Prentice-Hall, Englewood Cliffs, NJ.

Hamilton, J. C., Seborg, D. E. and Fisher, D. G., 1973, An experimental evaluation of Kalman filter. *A.I.Ch.E. J.* **19**, 901.

Himmelblau, D. M., 1978, *Fault Detection and Diagnosis in Chemical and Petro-chemical Processes.* Elsevier, Amsterdam, The Netherlands.

Himmelblau, D. M. and Bischoff, K. B., 1968, *Process Analysis and Simulation-Deterministic System*, p. 308 Wiley, New York.

Hoskins, J. C., Kalivur, K. M. and Himmelblau, D. M., 1991, Fault diagnosis in complex chemical plants using artificial neural networks. *A.I.Ch.E. J.* **37**, 137.

Isermannn, R., 1984, Process fault detection based on modeling and estimation methods—a survey. *Automatica* **20**, 387.

Kramer, M. A. and Palowitch, Jr., B. L., 1987, A rule-based approach to fault diagnosis using the signed directed graph. *A.I.Ch.E. J.* **33**, 1067.

Lamb, D. E., Dhurjati, P. and Chester, D. L., 1987, Development of an expert system for fault identification in a commercial scale chemical process. NSF-AAAI Workshop on AI in Process Engineering. Columbia University, New York.

Li, R. and Olson, J. H., 1991, Fault detection and diagnosis in a closed-loop nonlinear distillation process: application of extended Kalmam filters. *Ind. Engng Chem. Res.* **30**, 898.

Ljung, L., 1987, *System Identification: Theory for the User.* Prentice-Hall, Englewood Cliffs, NJ.

Park, S. and Himmelblau, D. M., 1983, Fault detection and diagnosis via parameter estimation in lumped dynamic systems. *Ind. Engng Chem. Process Des. Dev.* **22**, 482.

Petti, T. F., Klein, J. and Dhurjati, P. S., 1990, Diagnostic model processor: using deep knowledge for process fault diagnosis. *A.I.Ch.E. J.* **36**, 565.

Rich, S. H. and Venkatasubramanian, V., 1987, Model-based reasoning in diagnostic expert systems for chemical process plants. *Comput. Chem. Engng* **11**, 111.

Stadtherr, W. A., Gifford, A. W. and Scriven, L. E., 1974, Efficient solution of sparse sets of design equations. *Chem. Engng Sci.* **29**, 1025.

Steward, D. V., 1965, Partitioning and tearing systems of equations. *SIAM J.* **B2**, 345.

Venkatasubramanian, V. and Chan, K., 1989, A neural network methodology for process fault diagnosis. *A.I.Ch.E. J.* **35**, 1993.

Venkatasubramanian, V., Vaidyanathan, R. and Yamamoto, Y., 1990, Process fault detection and diagnosis using neural network—I. steady-state processes. *Comput. Chem. Engng* **14**, 699.

Venkateswarlu, C. H., Gangiah, K. and Rao, M. B., 1992, Two-level methods for incipient fault diagnosis in nonlinear chemical processes. *Comput. Chem. Engng* **16**, 463.

Watanabe, K. and Himmelblau, D. M., 1983a, Fault diagnosis in nonlinear chemical processes. Part I. Theory. *A.I.Ch.E. J.* **29**, 243.

Watanabe, K. and Himmelblau, D. M., 1983b, Fault diagnosis in nonlinear chemical processes. Part II. Application to a chemical reactor. *A.I.Ch.E. J.* **29**, 250.

Watanabe, K. and Himmelblau, D. M., 1984, Incipient fault diagnosis of nonlinear processes with multiple causes of faults. *Chem. Engng Sci.* **39**, 491.

Watanabe, K., Matsuura, M. A., Kubota, M. and Himmelblau, D. M., 1989, Incipient fault diagnosis of chemical process via artifical neural networks. *A.I.Ch.E. J.* **35**, 1803.

Willsky, A. S., 1976, A survey of design methods for failure detection in dynamic systems. *Automatica* **12**, 601.

Yu, C. C. and Lee, C., 1991, Fault diagnosis based on qualitative/quantitative process knowledge. *A.I.Ch.E. J.* **37**, 617.