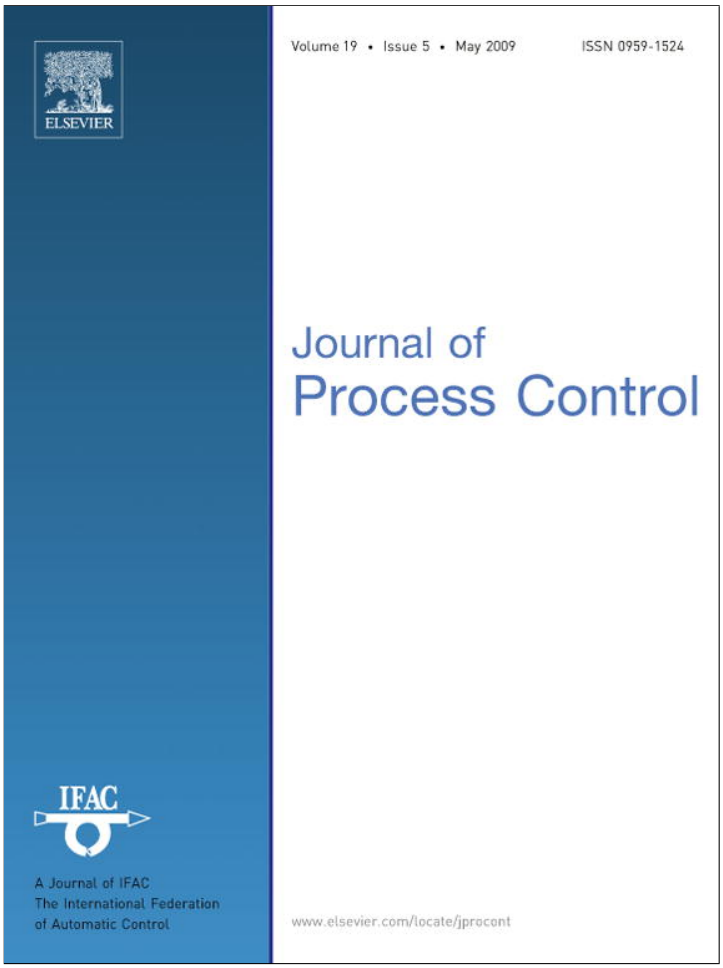

Journal of Process Control 19 (2009) 842–858

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Development of fault diagnosis strategies based on qualitative predictions of symptom evolution behaviors

Jung Yang Chen, Chuei-Tin Chang *

Department of Chemical Engineering, National Cheng Kung University, Tainan 70101, Taiwan, ROC

ARTICLE INFO

Article history:
Received 20 December 2007
Received in revised form 16 October 2008
Accepted 4 November 2008

Keywords:
Fault diagnosis
Fuzzy inference rule
Qualitative simulation
Signed–directed graph
Symptom evolution sequence

ABSTRACT

A SDG-based reasoning procedure is presented in this paper to qualitatively predict all possible symptom patterns and also their progression sequences caused by fault propagation in any given process system. These intrinsic features of the symptom evolution behaviors are then captured with IF–THEN rules in a two-layer fuzzy inference system. The proposed diagnostic system can be used to identify not only the locations of fault origins but also their magnitude levels with relatively high resolution. Numerical simulation studies have been carried out to verify the feasibility and effectiveness of the proposed approach.

1. Introduction

Due to the need to minimize operating costs while maintaining economic production scale, the chemical plants built in recent years are in general much larger and more complex than they used to be and, furthermore, their processing units are often designed to be operated under more extreme conditions. Thus, the development of hazard identification and risk reduction measures for such processes has become an important issue of industry-wide concern. To this end, the on-line fault diagnosis system should be considered as an useful tool. Notice that many different methods have already been proposed in the literature, e.g., the state estimator [1], the expert system [2] the neural network [3], the signed–directed graph (SDG) [4], the principal component analysis (PCA) [5], fuzzy system [6] and the frequency-domain analysis [7], etc. Generally speaking, these methods could be classified into three distinct groups [8–10], i.e., the model based approaches, the knowledge based approaches, and the data-analysis based approaches.

The SDG-based fault diagnosis strategy is the focus of present study. In essence, the digraph models have been used in the past to qualitatively characterize the causal relations among faults, failures and their effects [8,9,11,12]. The advantage of this modeling approach is mainly due to the fact that the SDG can almost always be constructed on the basis of simple causal relations. On the other hand, the more accurate mathematical models and the more case-specific knowledge bases are required to be built from the measurement data and operational experiences obtained in the course of *every* possible accident. This need is often not satisfiable. Thus, a qualitative model such as the digraph is actually more useful in these circumstances.

Although the SDG models are easy to develop, they are essentially static in nature. As a result, the available fault identification techniques are implemented mostly on the basis of steady-state symptoms, e.g., Maurya et al. [11]. Notice that the effects of fault(s) and/or failure(s) usually propagate throughout the entire system sequentially and dynamically. A series of intermediate events may occur before the inception of catastrophic consequences. Thus, the performance of a diagnosis scheme should be evaluated not only in terms of its correctness but also its timeliness. To enhance the diagnostic efficiency, it becomes necessary to consider the precedence order (in time) between the fault propagation effects implied in every input–output connection in digraph.

Extensive studies have been carried out in recent years to develop fault identification techniques by incorporating both the eventual symptoms and also their *occurrence order* into a fuzzy inference system (FIS) [13–16]. Basically, this approach can be implemented in two stages, i.e., (1) the off-line preparation stage and (2) the on-line implementation stage. In the preparation stage, a SDG-based qualitative simulation procedure is used to predict the fault propagation paths (FPPs) and the symptom occurrence orders (SOOs) resulting from various fault origins. The corresponding candidate patterns of on-line symptoms are then translated into IF–THEN inference rules for assessing the occurrence possibilities of fault origins. In the next stage, the on-line measurement data are normalized and used as inputs to FIS for computing all

* Corresponding author. Tel.: +886 6 275 7575x62663; fax: +886 6 234 4496.
E-mail address: ctchang@mail.ncku.edu.tw (C.-T. Chang).

0959-1524/$ - see front matter
doi:10.1016/j.jprocont.2008.11.006

corresponding occurrence indices in real time. This fault diagnosis approach has been implemented successfully in a number of loop-free processes [14] and also in systems with feedback and/or feed forward control loops [13,15,16]. Finally, an extended version of qualitative simulation method has been developed in Chen and Chang [16] to differentiate not only the locations of fault origins but also their magnitude levels. As a result, the FIS can be made to function effectively even in the multiple-fault scenarios.

Despite the aforementioned advancements, there is still room for significant improvement. Notice that, for every possible fault origin, essentially *all* the IF–THEN rules derived from corresponding candidate patterns are supposed to be included in the FIS and these rules are fired indiscriminately in any on-line application. This practice could reduce diagnostic resolution if the symptom patterns associated with two or more fault origins happen to be identical. On the other hand, it should also be noted that some of these predicted candidate patterns may not show up in an actual fault propagation scenario and the observable ones can be found to appear one-at-a-time in a definite evolution sequence. This concept of *pattern evolution sequence* (PES) has been briefly brought up in Chen and Chang [15], but was utilized in an ad hoc fashion without specific guidelines. Clearly, there is a need to develop a reliable approach to compute the total number of pattern evolution sequences and to generate a complete list of such sequences automatically. The intrinsic features of the symptom evolution behaviors can then be captured accordingly with the fuzzy inference rules. Thus, in order to further enhance the diagnostic performance of FIS, these PES-related issues have been thoroughly addressed in the present study.

The rest of this paper is organized as follows. For illustration purpose, the SDG-based qualitative simulation procedure for predicting the fault propagation behaviors is reviewed and also the special representations of simulation results, i.e., FPPs and SOOs, are described in Section 2. The intrinsic features embedded in the symptom evolution behaviors are analyzed in the next section. Also included are the mathematical formulas and computer algorithms for generating all possible candidate patterns and pattern evolution sequences. In Section 4, both the single-layer and two-layer inference structures are outlined. The IF–THEN rules in the former system are constructed only on the basis of candidate patterns, while the PESs are also incorporated in the latter. Finally, extensive numerical simulation results are presented at the end of this paper to demonstrate the feasibility and effectiveness of the proposed approach.

2. Fault propagation mechanisms

By definition, an accident is a rarely-occurred unplanned event or a sequence of such events. Some of the catastrophic accidents may not be experienced even in a long-existing plant. Thus, in any realistic system, it is obviously not feasible to collect and analyze the historical data of all possible scenarios. As a result, it becomes necessary to *predict* the fault propagation behaviors with qualitative simulation techniques. A brief review of the SDG-based approach [4,12,17–21] is presented in the sequel:

2.1. Qualitative simulation procedure

The effects of a fault/failure in a process system can be easily simulated on the basis of a SDG model. It should be noted that the fault origins can usually be associated with the primal nodes, i.e., nodes with no inputs. A set of five values, i.e., $\{-10,-1,0,+1,+10\}$, may be assigned to each node to *qualitatively* represent deviation from the normal value of corresponding variable. "0" means that it is at the normal steady state. The negative values are used to denote the lower-than-normal states and the positive values signify the opposite. The absolute values of non-zero deviations, i.e., 1 or 10, can be interpreted qualitatively as "small" and "large", respectively. Notice also that the causal relation between the deviations in two variables can be characterized with a directed arc and the corresponding gain. Again each gain may assume one of the aforementioned five qualitative values, i.e., 0, ±1 and ±10. The output value of any arc can be computed with the gain and its input value according to the following equation:

$$v_{out} = \begin{cases} g \times v_{in} & \text{if } -10 \leqslant g \times v_{in} \leqslant +10 \\ +10 & \text{if } g \times v_{in} > +10 \\ -10 & \text{if } g \times v_{in} < -10 \end{cases} \quad (1)$$

where g, v_{in} and v_{out} denote respectively the gain, input and output values. It is obvious that the deviation values of all variables affected by one or more fault origin can always be computed with this formula, but the time at which each deviation occurs is indeterminable. Without the reference of time in the SDG-based simulation results, it can nonetheless be safely assumed that *the change in an input variable should always occur earlier than those in its outputs.*

2.2. Fault propagation path

Due to the unique information structure generated with the above approach, a special representation has been designed to characterize the predicted fault propagation behaviors. This representation is referred to as the *fault propagation path* (FPP). For the sake of brevity, only the FPPs associated with the tree-shaped SDGs are presented here. It should be emphasized that *all* other digraph configurations, e.g., feedback loops and feed forward loops, can be converted to equivalent trees and the following analysis is still applicable in these cases [16].

To fix idea, let us consider the SDG given in Fig. 1a. The fault propagation path associated with a "small" disturbance $D(+1)$ can be found in Fig. 1b. Notice first that the structures of SDG and FPP in this case are identical. Each node in FPP represents a previously nonexistent fault effect. Every effect is specified with a qualitative value +1 or −1, which can be computed according to Eq. (1). The precedence order of two different effects is specified with the connecting symbol $\prec$, i.e., the effect on its left must occur earlier than that on the right. The sequence of conditions on the same propagation path should be interpreted as the order of occurrence (in time) of different effects resulting from the given fault origin. However, the order of two distinct events located on two *separate* branches should be considered as indeterminable.

On the other hand, the FPP resulting from a "large" disturbance is obtained in this study by assuming that a *finite* time constant is needed to characterize the transient response of an output variable

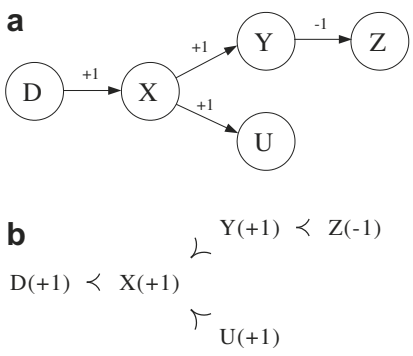

Fig. 1. (a) A fictitious tree-shaped SDG model; (b) The corresponding fault propagation path resulting from $D(+1)$.

to the disturbance in its input. Thus, an additional constraint is introduced to facilitate an accurate description of the fault propagation behaviors in this situation, i.e., *the smaller deviation of a process variable must occur before reaching a larger one of the same variable*. By incorporating this requirement, the branches in the tree-shaped FPP in Fig. 1b can be transformed to the ones given in Figs. 2a and b, respectively. This composite propagation mechanism will later be expressed with a simplified format, i.e., Fig. 2c, throughout this paper.

2.3. Symptom occurrence order

Notice that, in practice, not all the events included in FPPs can be monitored with existing on-line instruments. Thus, the FPPs should be further reduced by merging every pair of measured variable and its measurement signal in the propagation paths and then eliminating the unmeasured ones. This resulting symptom occurrence order (SOO) can be used as the basis for developing the fuzzy inference system (FIS) for fault diagnosis.

3. Symptom evolution behaviors

3.1. Candidate patterns

If all symptoms in a SOO can be observed simultaneously, then it is certainly logical to confirm the existence of corresponding fault origin(s). However, since the fault propagation behaviors are dynamical in nature, the resulting on-line measurements should vary with time during the incipient stage. In this work, the collection of on-line symptoms observed at any instance in the fault propagation process is referred to as a *candidate pattern*. It is obvious that every candidate pattern can be considered as an evidence for fault identification. Thus, it is important to enumerate all possibilities and assign each one of them an appropriate confidence level.

In a previous study, Chen and Chang [16] derived a set of formulas to compute the total number of candidate patterns (N_{CP}) associated with any tree-shaped composite SOO. For the sake of completeness, these formulas are given below:

$$N_{CP} = \mathcal{N}\left\{\mathbf{P}^{(0)}(m,n_0)\right\} = \binom{m+n_0-1}{m} + \sum_{\tilde{j}_1=1}^{m}\binom{m-\tilde{j}_1+n_0-1}{m-\tilde{j}_1}\prod_{\tilde{\imath}_1=1}^{B_0}\mathcal{N}\left\{\mathbf{P}^{(0,\tilde{\imath}_1)}(\tilde{j}_1,n_{0,\tilde{\imath}_1})\right\} \quad (2)$$

where $\mathbf{P}^{(0)}(m,n_0)$ denotes the initial path of a tree-shaped composite SOO having m disturbance levels and n_0 measurement nodes; $\mathbf{P}^{(0,\tilde{\imath}_1)}(\tilde{j}_1,n_{0,\tilde{\imath}_1})$ denotes the $\tilde{\imath}_1$th branch path connecting to the end of $\mathbf{P}^{(0)}(m,n_0)$ with $\tilde{j}_1$ disturbance levels and $n_{0,\tilde{\imath}_1}$ measurement nodes; B_0 denotes the total number of these branch paths. Finally, notice that $\mathcal{N}\{\cdot\}$ is a counting operator and the counting operation is carried out recursively, i.e.,

$$\mathcal{N}\left\{\mathbf{P}^{(0,i_1,i_2,\ldots,i_k)}(j_k,n_{0,i_1,i_2,\ldots,i_k})\right\} = \binom{j_k+n_{0,i_1,i_2,\ldots,i_k}-1}{j_k} + \sum_{\tilde{j}_{k+1}=1}^{j_k}\binom{j_k-\tilde{j}_{k+1}+n_{0,i_1,i_2,\ldots,i_k}-1}{j_k-\tilde{j}_{k+1}} \times \prod_{\tilde{\imath}_{k+1}=1}^{B_{0,i_1,i_2,\ldots,i_k}}\mathcal{N}\left\{\mathbf{P}^{(0,i_1,i_2,\ldots,i_k,\tilde{\imath}_{k+1})}(\tilde{j}_{k+1},n_{0,i_1,i_2,\ldots,i_k,\tilde{\imath}_{k+1}})\right\} \quad (3)$$

where $k = 1, 2, \ldots$ and $\mathbf{P}^{(0,i_1,i_2,\ldots,i_k,\tilde{\imath}_{k+1})}(\tilde{j}_{k+1},n_{0,i_1,i_2,\ldots,i_k,\tilde{\imath}_{k+1}})$ is the $\tilde{\imath}_{k+1}$th branch path (with $\tilde{j}_{k+1}$ disturbance levels and $n_{0,i_1,i_2,\ldots,i_k,\tilde{\imath}_{k+1}}$ measurement nodes) connecting to the end of path $\mathbf{P}^{(0,i_1,i_2,\ldots,i_k)}(j_k,n_{0,i_1,i_2,\ldots,i_k})$ (with j_k disturbance levels and $n_{0,i_1,i_2,\ldots,i_k}$ measurement nodes). If there are no further branches connected to the end of the branch path $\mathbf{P}^{(0,i_1,i_2,\ldots,i_k)}(j_k,n_{0,i_1,i_2,\ldots,i_k})$, i.e., $B_{0,i_1,i_2,\ldots,i_k}=0$, then

$$\prod_{\tilde{\imath}_{k+1}=1}^{0}[\cdot] = 1 \quad (4)$$

As an example, let us again consider the tree-shaped SDG presented in Fig. 1a. The FPPs resulting from a large and a small disturbance, i.e., $D(+1)$ and $D(+10)$, are given in Figs. 1b and 2c, respectively. Notice that the value of m should be set to 1 for the SOO caused by $D(+1)$ in the former case and 2 in the latter case. Let us further assume that all nodes except the disturbance itself are observable and, thus, $n_0 = 1$, $n_{0,1} = 2$, $n_{0,2} = 1$, $B_0 = 2$ and $B_{0,1} = B_{0,2} = 0$. The total number of candidate patterns can be computed according to Eqs. (2)–(4), i.e.,

- $m = 1$:

$$N_{CP} = \binom{1}{1} + \sum_{\tilde{j}_1=1}^{1}\binom{1-\tilde{j}_1}{1-\tilde{j}_1}\prod_{\tilde{\imath}_1=1}^{2}\mathcal{N}\left\{\mathbf{P}^{(0,\tilde{\imath}_1)}(\tilde{j}_1,n_{0,\tilde{\imath}_1})\right\} = 1+\mathcal{N}\left\{\mathbf{P}^{(0,1)}(1,2)\right\}\mathcal{N}\left\{\mathbf{P}^{(0,2)}(1,1)\right\} = 1+(2+1\times1)(1+1\times1) = 7$$

- $m = 2$:

$$N_{CP} = \binom{2}{2} + \sum_{\tilde{j}_1=1}^{2}\binom{2-\tilde{j}_1}{2-\tilde{j}_1}\prod_{\tilde{\imath}_1=1}^{2}\mathcal{N}\left\{\mathbf{P}^{(0,\tilde{\imath}_1)}(\tilde{j}_1,n_{0,\tilde{\imath}_1})\right\} = 1+\mathcal{N}\left\{\mathbf{P}^{(0,1)}(1,2)\right\}\mathcal{N}\left\{\mathbf{P}^{(0,2)}(1,1)\right\} + \mathcal{N}\left\{\mathbf{P}^{(0,1)}(2,2)\right\}\mathcal{N}\left\{\mathbf{P}^{(0,2)}(2,1)\right\} = 1+(2+1\times1)(1+1\times1) + (3+2\times1+1\times1)(1+1\times1+1\times1) = 25$$

All corresponding patterns are listed in Tables 1 and 2 respectively. In the large disturbance scenario, since two distinct deviation levels of the same variable are expressed with separate

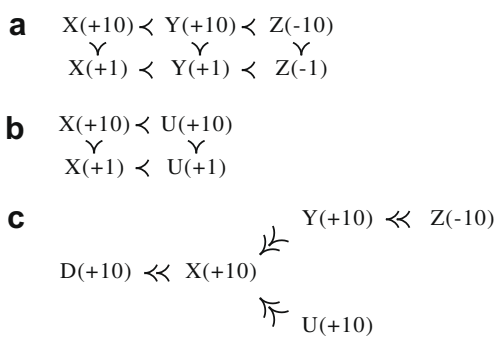

Fig. 2. (a) The fault propagation mechanism caused by $D(+10)$ along a branch of the tree-shaped SDG in Fig. 1a; (b) The fault propagation mechanism caused by $D(+10)$ along another branch of the tree-shaped SDG in Fig. 1a; (c) The composite fault propagation mechanism.

Table 1
Candidate patterns derived from the FPP in Fig. 1b.

No.	X	Y	Z	U
1	0	0	0	0
2	1	0	0	0
3	1	1	0	0
4	1	1	−1	0
5	1	0	0	1
6	1	1	0	1
7	1	1	−1	1

Table 2
Candidate patterns derived from the composite FPP in Fig. 2c.

No.	X	Y	Z	U
1	0	0	0	0
2	1	0	0	0
3	1	1	0	0
4	1	1	−1	0
5	1	0	0	1
6	1	1	0	1
7	1	1	−1	1
8	10	0	0	0
9	10	1	0	0
10	10	1	−1	0
11	10	0	0	1
12	10	1	0	1
13	10	1	−1	1
14	10	10	0	0
15	10	10	−1	0
16	10	10	0	1
17	10	10	−1	1
18	10	10	−10	0
19	10	10	−10	1
20	10	0	0	10
21	10	1	0	10
22	10	1	−1	10
23	10	10	0	10
24	10	10	−1	10
25	10	10	−10	10

nodes in the composite SOO, a reconciliation criterion is needed to determine their net effects at various instances. This criterion is stipulated in our work according to the definition of precedence order, i.e., *the latest symptom should always override all the previous ones in a candidate pattern.* The candidate patterns in Table 2 are obtained by enumerating all possible patterns according to Figs. 2a and b and then applying the aforementioned overriding principle.

3.2. Pattern evolution sequences

It is important to point out that not only the collection of current on-line symptoms (i.e., the candidate patterns) but also their *evolution sequence* in the past could be used as evidences for fault identification. Let us consider the FPP presented in Fig. 1b and the corresponding candidate patterns in Table 1. It can be deduced by inspection that the disturbance $D(+1)$ can result in *only one* of the following three evolution sequences of symptom patterns:

$$\mathbf{PES_1}: \quad 1 \prec 2 \prec 3 \prec 4 \prec 7 \tag{5}$$

$$\mathbf{PES_2}: \quad 1 \prec 2 \prec 3 \prec 6 \prec 7 \tag{6}$$

$$\mathbf{PES_3}: \quad 1 \prec 2 \prec 5 \prec 6 \prec 7 \tag{7}$$

Notice that every pattern in the above equations is represented with a numerical label and the corresponding symptoms can be found in Table 1. Each series of patterns is referred to in this paper as a *pattern evolution sequence* (PES). Thus, the task of fault diagnosis can also be viewed as that of identifying a match between the historical record of on-line symptoms and one of the PESs. Finally, it can be observed from the above PESs that the number of *occurred* symptoms is increased one-at-a-time in the candidate patterns. In other words, the fewer the number of symptoms included in a pattern the earlier it can be observed on-line. As a result, the aforementioned PESs can be expressed in an alternative form according to the precedence order of newly observed symptoms, i.e.,

$$\mathbf{PES_1}: \quad X(+1) \prec Y(+1) \prec Z(-1) \prec U(+1) \tag{8}$$

$$\mathbf{PES_2}: \quad X(+1) \prec Y(+1) \prec U(+1) \prec Z(-1) \tag{9}$$

$$\mathbf{PES_3}: \quad X(+1) \prec U(+1) \prec Y(+1) \prec Z(-1) \tag{10}$$

A set of generalized formulas have been derived in this study for computing the total number of pattern evolution sequences (N_{PES}) corresponding to any tree-shaped SOO with a single disturbance level, i.e., $m = 1$. By using the notations in Eqs. (2)–(4) to characterize the tree branches in a SOO, these formulas can be written as

$$N_{PES} = \mathscr{K}\left\{\mathbf{P}^{(0)}(1, n_0)\right\} \tag{11}$$

where, $\mathscr{K}\{\cdot\}$ is a counting operator for a branch path in SOO. This operation should be carried out recursively, i.e.,

$$\begin{aligned}\mathscr{K}&\left\{\mathbf{P}^{(0,i_1,i_2,\ldots,i_k)}(1, n_{0,i_1,i_2,\ldots,i_k})\right\} \\ &= \mathscr{R}\left\{\mathbf{P}^{(0,i_1,i_2,\ldots,i_k)}(1, n_{0,i_1,i_2,\ldots,i_k})\right\} \\ &\quad \times \prod_{\tilde{i}_{k+1}=1}^{B_{0,i_1,i_2,\ldots,i_k}} \mathscr{K}\left\{\mathbf{P}^{(0,i_1,i_2,\ldots,i_k,\tilde{i}_{k+1})}(1, n_{0,i_1,i_2,\ldots,i_k,\tilde{i}_{k+1}})\right\}\end{aligned} \tag{12}$$

In this equation, the value of $\mathscr{R}\{\cdot\}$ is determined according to the following formula:

$$\mathscr{R}\left\{\mathbf{P}^{(0,i_1,i_2,\ldots,i_k)}(1, n_{0,i_1,i_2,\ldots,i_k})\right\} = \frac{\left[\sum_{\tilde{i}_{k+1}=1}^{B_{0,i_1,\ldots,i_k}} \mathscr{L}\left\{\mathbf{P}^{(0,i_1,\ldots,i_k,\tilde{i}_{k+1})}(1, n_{0,i_1,\ldots,i_k,\tilde{i}_{k+1}})\right\}\right]!}{\prod_{\tilde{i}_{k+1}=1}^{B_{0,i_1,\ldots,i_k}} \mathscr{L}\left\{\mathbf{P}^{(0,i_1,\ldots,i_k,\tilde{i}_{k+1})}(1, n_{0,i_1,\ldots,i_k,\tilde{i}_{k+1}})\right\}!} \tag{13}$$

where the operator $\mathscr{L}\{\cdot\}$ in the above equation is

$$\begin{aligned}\mathscr{L}&\left\{\mathbf{P}^{(0,i_1,i_2,\ldots,i_k)}(1, n_{0,i_1,i_2,\ldots,i_k})\right\} \\ &= n_{0,i_1,i_2,\ldots,i_k} + \sum_{\tilde{i}_{k+1}=1}^{B_{0,i_1,i_2,\ldots,i_k}} \mathscr{L}\left\{\mathbf{P}^{(0,i_1,i_2,\ldots,i_k,\tilde{i}_{k+1})}(1, n_{0,i_1,i_2,\ldots,i_k,\tilde{i}_{k+1}})\right\}\end{aligned} \tag{14}$$

If there are no further branches connected to the end of the branch path $\mathbf{P}^{(0,i_1,i_2,\ldots,i_k)}(n_{0,i_1,i_2,\ldots,i_k})$, i.e., $B_{0,i_1,i_2,\ldots,i_k} = 0$, then Eqs. (12) and (14) can be evaluated, respectively, with Eq. (4) and the following equation:

$$\sum_{\tilde{i}_{k+1}=1}^{0} [\cdot] = 0 \tag{15}$$

Notice that a derivation is presented in Appendix A to show the validity of the aforementioned formulas. Let us again use the FPP in Fig. 1b as an example. The total number of PESs can be calculated with Eqs. (11)–(15), i.e., $N_{PES} = \frac{(2+1)!}{2!1!} = 3$, which is exactly the same as that determined by inspection. In order to provide additional details in implementing the proposed calculation steps, let us consider the more complicated SOO in Fig. 3 and the corresponding candidate patterns in Table 3. The specific computation procedure in this case can be expressed as follows:

$$N_{PES} = \frac{(2+1+1)!}{2!1!1!} = 12$$

It should also be noted that the total PES number associated with a multi-level SOO can be determined with essentially the same approach. However, it is extremely cumbersome to express the corre-

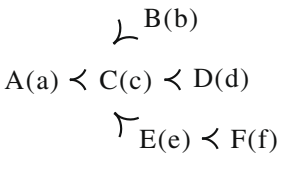

Fig. 3. A fictitious SOO.

Table 3
Candidate patterns derived from the SOO in Fig. 3.

No.	A	B	C	D	E	F
1	0	0	0	0	0	0
2	a	0	0	0	0	0
3	a	0	c	0	0	0
4	a	b	c	0	0	0
5	a	0	c	d	0	0
6	a	b	c	d	0	0
7	a	0	c	0	e	0
8	a	b	c	0	e	0
9	a	0	c	d	e	0
10	a	b	c	d	e	0
11	a	0	c	0	e	f
12	a	b	c	0	e	f
13	a	0	c	d	e	f
14	a	b	c	d	e	f

sponding computation procedure with explicit formulas in this situation. A PES generation algorithm has thus been developed to generate all possible sequences and also to compute N_{PES} on the basis of a given composite SOO. This algorithm and a pseudo code are included in an electronic annex (see Annex 1 in the online version of this article). In the case of the tree-shaped SOO in Fig. 2c, a total of 91 sequences can be identified. They are listed in another electronic annex (see Annex 2) according to the candidate patterns defined in Table 2.

4. Fuzzy inference system

The intrinsic features of the aforementioned symptom variation behaviors are described with IF–THEN rules in the present work. Each rule is used to evaluate a confidence measure toward confirming the corresponding fault origin. A single-layer fuzzy inference system can be built solely on the basis of candidate patterns, while the diagnostic resolution can be further enhanced in a two-layer system with additional information extracted from the pattern evolution sequences. The detailed construction procedures of these two types of inference systems are presented as follows:

4.1. Single-layer inference structure

Every candidate pattern can be encoded into an IF–THEN rule to evaluate the existence potential (or occurrence index *cs*) of the corresponding fault origin. Specifically, the premises of this rule are constructed on the basis of the qualitative deviation values of the symptoms in the given pattern. These deviations are translated into linguistic values according to an interpretation function F_{in}, i.e.,

$$F_{in}(\delta_j) = \begin{cases} \text{LN} & \text{if } \delta_j = -10 \\ \text{SN} & \text{if } \delta_j = -1 \\ \text{ZE} & \text{if } \delta_j = 0 \\ \text{SP} & \text{if } \delta_j = +1 \\ \text{LP} & \text{if } \delta_j = +10 \end{cases} \tag{16}$$

In the above equation, δ_j denotes the deviation value of the jth measurement ($j = 1,2,\ldots,N_M$); LN, SN, ZE, SP and LP denote respectively the linguistic values of −10, −1, 0, +1 and +10.

If the on-line symptoms are identical to those in a SOO, then it is highly possible that they are caused by the corresponding fault origin. To assert such a belief, the conclusion "*cs* is OCR" should be used in the inference rule. Here, OCR is the linguistic value of the occurrence index *cs* reflecting the highest confidence level in confirming the existence of the root cause(s). On the other hand, it is reasonable to disregard the possibility of a fault if none of the symptoms in the corresponding SOO can be observed. Thus, the conclusion in the inference rule for this scenario should be "*cs* is NOC", where NOC is the linguistic value representing the lowest confidence. The conclusions of the remaining rules should be uncertain. Naturally, the confidence level of a particular candidate pattern in confirming the existence of the root cause(s) should be proportional to the number of matched (or occurred) symptoms in SOO. In this study, the latter value ℓ is used directly as a qualitative measure of confidence. Since the events associated with different values of the same variable may be included as different nodes in a SOO, the latest among them always overrides all the previous ones in the on-line measurements. To account for the overridden symptoms implied in a candidate pattern, the following formula is used in this work for computing the confidence level ℓ:

$$\ell = \sum_{j=1}^{N_M} c(\delta_j) \tag{17}$$

Table 4
Fuzzy inference rules constructed according to the candidate patterns in Table 2.

No.	IF				THEN
	X	Y	Z	U	cs
1	ZE	ZE	ZE	ZE	NOC
2	SP	ZE	ZE	ZE	UCT_1
3	SP	SP	ZE	ZE	UCT_2
4	SP	SP	SN	ZE	UCT_3
5	SP	ZE	ZE	SP	UCT_2
6	SP	SP	ZE	SP	UCT_3
7	SP	SP	SN	SP	UCT_4
8	LP	ZE	ZE	ZE	UCT_2
9	LP	SP	ZE	ZE	UCT_3
10	LP	SP	SN	ZE	UCT_4
11	LP	ZE	ZE	SP	UCT_3
12	LP	SP	ZE	SP	UCT_4
13	LP	SP	SN	SP	UCT_5
14	LP	LP	ZE	ZE	UCT_4
15	LP	LP	SN	ZE	UCT_5
16	LP	LP	ZE	SP	UCT_5
17	LP	LP	SN	SP	UCT_6
18	LP	LP	LN	ZE	UCT_6
19	LP	LP	LN	SP	UCT_7
20	LP	ZE	ZE	LP	UCT_4
21	LP	SP	ZE	LP	UCT_5
22	LP	SP	SN	LP	UCT_6
23	LP	LP	ZE	LP	UCT_6
24	LP	LP	SN	LP	UCT_7
25	LP	LP	LN	LP	*OCR*

Table 5
PESs derived from the SOO in Fig. 3.

No.	IF						THEN
	A	B	C	D	E	F	
1	ZE	ZE	ZE	ZE	ZE	ZE	cp_1 is OCR
2	$F_{in}(a)$	ZE	ZE	ZE	ZE	ZE	cp_2 is OCR
3	$F_{in}(a)$	ZE	$F_{in}(c)$	ZE	ZE	ZE	cp_3 is OCR
4	$F_{in}(a)$	$F_{in}(b)$	$F_{in}(c)$	ZE	ZE	ZE	cp_4 is OCR
5	$F_{in}(a)$	ZE	$F_{in}(c)$	$F_{in}(d)$	ZE	ZE	cp_5 is OCR
6	$F_{in}(a)$	$F_{in}(b)$	$F_{in}(c)$	$F_{in}(d)$	ZE	ZE	cp_6 is OCR
7	$F_{in}(a)$	ZE	$F_{in}(c)$	ZE	$F_{in}(e)$	ZE	cp_7 is OCR
8	$F_{in}(a)$	$F_{in}(b)$	$F_{in}(c)$	ZE	$F_{in}(e)$	ZE	cp_8 is OCR
9	$F_{in}(a)$	ZE	$F_{in}(c)$	$F_{in}(d)$	$F_{in}(e)$	ZE	cp_9 is OCR
10	$F_{in}(a)$	$F_{in}(b)$	$F_{in}(c)$	$F_{in}(d)$	$F_{in}(e)$	ZE	cp_{10} is OCR
11	$F_{in}(a)$	ZE	$F_{in}(c)$	ZE	$F_{in}(e)$	$F_{in}(f)$	cp_{11} is OCR
12	$F_{in}(a)$	$F_{in}(b)$	$F_{in}(c)$	ZE	$F_{in}(e)$	$F_{in}(f)$	cp_{12} is OCR
13	$F_{in}(a)$	ZE	$F_{in}(c)$	$F_{in}(d)$	$F_{in}(e)$	$F_{in}(f)$	cp_{13} is OCR
14	$F_{in}(a)$	$F_{in}(b)$	$F_{in}(c)$	$F_{in}(d)$	$F_{in}(e)$	$F_{in}(f)$	cp_{14} is OCR

where $c(\delta_j)$ denotes the number of symptoms *occurred* in the jth measured variable when its current value is δ_j. In other words, ℓ is the total number of nodes in SOO which have been confirmed with measurement data.

A second interpretation function F_{out} can be defined accordingly to determine the linguistic values of occurrence index, i.e.,

$$F_{out}(\ell) = \begin{cases} \text{NOC} & \text{if } \ell = 0 \\ \text{OCR} & \text{if } \ell = \ell_{max} \\ \text{UCT}_\ell & \text{otherwise} \end{cases} \quad (18)$$

where ℓ_{max} is the confidence level associated with the fully-developed candidate pattern of the given SOO. As an example, let us consider the candidate patterns in Table 2. These patterns can be converted to the fuzzy inference rules presented in Table 4 with the aforementioned interpretation functions.

Table 6
Second-layer inference rules for the 4th pattern of the 1st PES in Eq. (16).

No.	IF				THEN
	$cp1^m$	$cp2^m$	$cp3^m$	$cp7^m$	$pes1$
1	OCR	OCR	OCR	OCR	UCT_3
2	OCR	OCR	NOC	OCR	UCT_2
3	OCR	NOC	OCR	OCR	UCT_2
4	OCR	NOC	NOC	OCR	UCT_1

4.2. Two-layer inference structure

As mentioned before, every candidate pattern is encoded into an inference rule in the single-layer system to evaluate the occurrence index of the corresponding fault origin. Thus, a set of different scenarios may not be distinguishable if some of the corresponding symptom patterns are the same. In these circumstances, the additional insights concerning PESs can be utilized to enhance diagnostic resolution. The two-layer inference structure is used in the present study for this purpose. A measure of agreement between the *current* on-line symptoms and every candidate pattern can be computed in the first layer. For illustration conve-

Table 7
Candidate patterns caused by fault origins $m_1(+1)$ and $m_3(+1)$ in Example 1.

No.	s_8	s_7
Origin 1: $m_1(+1)$		
1	0	0
2	1	0
3	1	1
4	10	10
Origin 2: $m_3(+1)$		
1	0	0
2′	0	1
3	1	1
4	10	10

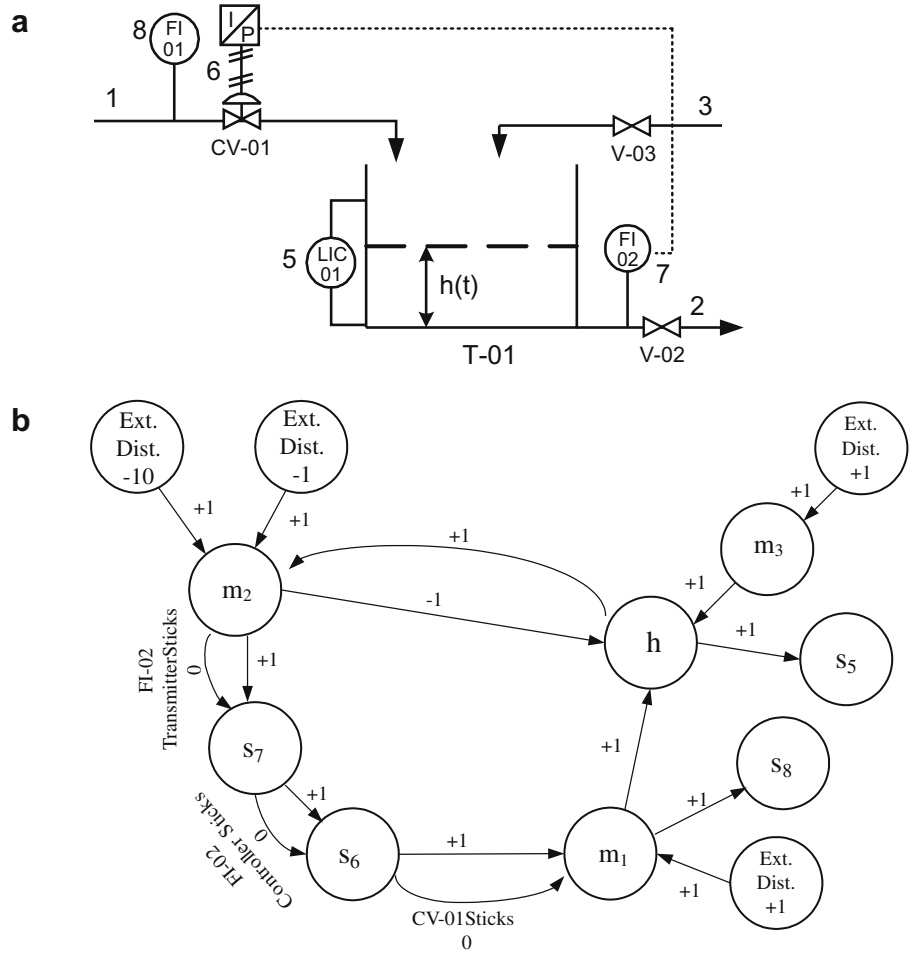

Fig. 4. (a) The process flow diagram of a single-tank storage system with feed-forward level-control loop and (b) the SDG model of single-tank storage system.

nience, let us consider the SOO in Fig. 3 and the corresponding candidate patterns in Table 3. At any instance, the *agreement measures* in this example (say cp_1–cp_{14}) can be evaluated with the IF–THEN rules given in Table 5. The maximum values of these measures $(cp_1^m, cp_2^m, \ldots, cp_{14}^m)$ and their occurrence times $(t_1^m, t_1^m, \ldots, t_{14}^m)$ are updated and also recorded every time a new batch of on-line measurements are taken. It should also be noted that twelve pattern evolution sequences can be generated according to the proposed algorithm, i.e,

$$\begin{aligned}
&\mathbf{PES}_1: \quad 1 \prec 2 \prec 3 \prec 7 \prec 11 \prec 13 \prec 14\\
&\mathbf{PES}_2: \quad 1 \prec 2 \prec 3 \prec 5 \prec 9 \prec 13 \prec 14\\
&\mathbf{PES}_3: \quad 1 \prec 2 \prec 3 \prec 5 \prec 6 \prec 10 \prec 14\\
&\mathbf{PES}_4: \quad 1 \prec 2 \prec 3 \prec 7 \prec 9 \prec 13 \prec 14\\
&\mathbf{PES}_5: \quad 1 \prec 2 \prec 3 \prec 5 \prec 9 \prec 10 \prec 14\\
&\mathbf{PES}_6: \quad 1 \prec 2 \prec 3 \prec 7 \prec 9 \prec 10 \prec 14\\
&\mathbf{PES}_7: \quad 1 \prec 2 \prec 3 \prec 7 \prec 11 \prec 12 \prec 14\\
&\mathbf{PES}_8: \quad 1 \prec 2 \prec 3 \prec 4 \prec 8 \prec 12 \prec 14\\
&\mathbf{PES}_9: \quad 1 \prec 2 \prec 3 \prec 4 \prec 6 \prec 10 \prec 14\\
&\mathbf{PES}_{10}: \quad 1 \prec 2 \prec 3 \prec 7 \prec 8 \prec 12 \prec 14\\
&\mathbf{PES}_{11}: \quad 1 \prec 2 \prec 3 \prec 7 \prec 8 \prec 10 \prec 14\\
&\mathbf{PES}_{12}: \quad 1 \prec 2 \prec 3 \prec 4 \prec 8 \prec 10 \prec 14
\end{aligned} \tag{19}$$

Naturally, the occurrence times of the maximum agreement measures must be consistent with the precedence order given in one of these PESs. For example, if the first PES in Eq. (19) is the correct one, then the following inequality constraint must be satisfied

$$t_1^m \leqslant t_2^m \leqslant t_3^m \leqslant t_7^m \leqslant t_{11}^m \leqslant t_{13}^m \leqslant t_{14}^m \tag{20}$$

If, at any time, this constraint is violated, the corresponding outputs of the first-layer inference system, i.e., cp_1^m–cp_{14}^m, should all be reset to zero to exclude the possibility of $\mathbf{PES}_1$.

The second-layer inference system is used for comparing the time profile of on-line measurement data with each PES, and then determining a measure of closeness between these two trends. The inference rules for computing this measure can be constructed on the basis of the outputs obtained from the first layer. For example, let us consider the fourth pattern of the 1st PES in Eq. (19), i.e., pattern 7 in Table 3. The corresponding *closeness measure* ($pes1$) can be obtained with the rules given in Table 6. Notice that the rationale for adopting rule 1 in this table is essentially the same as that for the inference rule used in a single-layer system to determine the occurrence index associated with the 7th pattern in Table 3. On the other hand, the remaining rules in Table 6 are used to reduce the closeness measure under the condition that one or more previous pattern in PES is unobservable. It should be stressed that all other second-layer inference rules for the PESs in Eq. (19) can be synthesized with essentially the same approach. A complete listing is omitted here for the sake of conciseness.

Finally, it should be noted that the occurrence index of a fault origin can be determined by taking the largest value among all closeness measures. In the case, this value is

$$cs = \max\{pes1, pes2, pes3, \cdots, pes12\} \tag{21}$$

5. Case studies

To verify the effectiveness of the proposed approach, extensive numerical simulation studies have been carried out in this work. Since the thrust of this work is to incorporate the intrinsic features of symptom evolution behaviors into the fuzzy inference system so as to enhance diagnostic resolution, the examples presented below are mainly used to demonstrate the importance of considering pattern evolution sequences in fault diagnosis. The on-line measurement data of all fault propagation scenarios were generated with *SIMULINK* [22]. These data were then used in Sugeno's inference procedure with the fuzzy-logic module of MATLAB toolbox [23].

5.1. A single-tank level-control system with coupled feed-forward and feedback loops

Let us first consider the level-control system presented in Fig. 4a and the corresponding SDG model in Fig. 4b. A "feed forward" control strategy is adopted here based on the assumption that only the disturbances affecting the outlet flow rate m_2 directly are important. The fluctuations in downstream pressure and adjustments of the outlet valve V-02 could be considered as typical examples of such disturbances. In this control system, the liquid level is supposed to be maintained at a desired height by adjusting the input flow according to on-line measurement of the flow rate at outlet. Although such a control strategy is clearly not realistic,

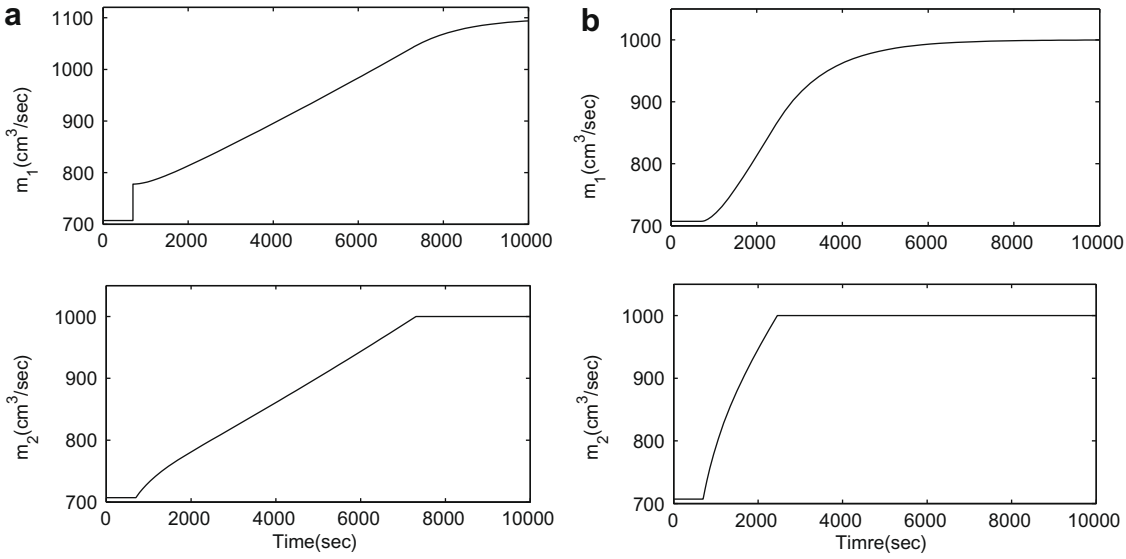

Fig. 5. (a) The simulated dynamic responses of $m_1(+1)$ and (b) the simulated dynamic responses of $m_3(+1)$.

the present example is nonetheless adopted mainly to demonstrate the potential benefits of the proposed two-layer inference structure. Notice that there are three interconnected "loops" in the digraph, i.e., (1) a negative feed forward loop with two paths

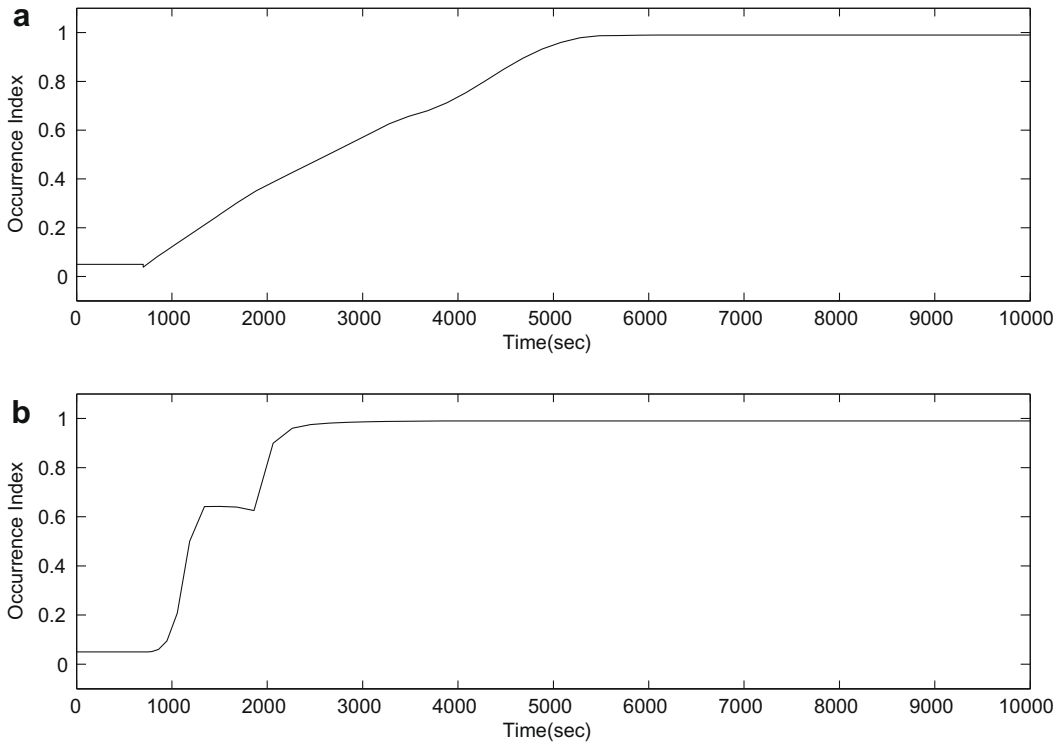

Fig. 6. The diagnosis results obtained with a single-layer inference system concerning the assumed fault origin $m_1(+1)$. (a) Actual fault origin: $m_1(+1)$; (b) Actual fault origin: $m_3(+1)$.

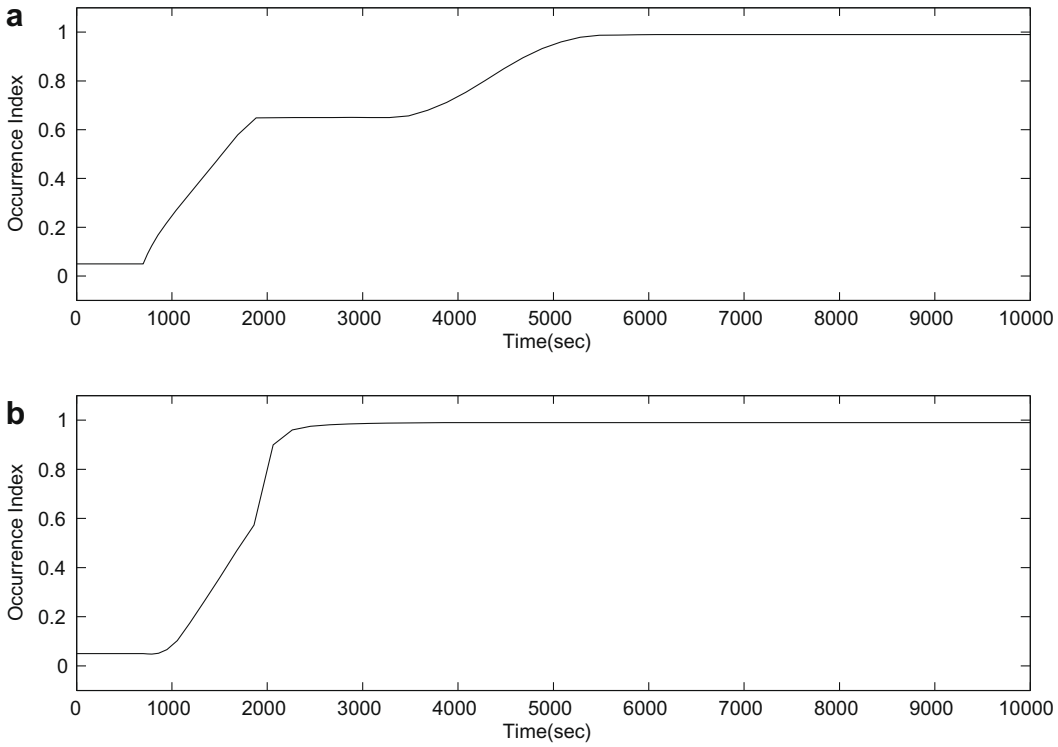

Fig. 7. The diagnosis results obtained with a single-layer inference system concerning the assumed fault origin $m_3(+1)$. (a) Actual fault origin: $m_1(+1)$; (b) Actual fault origin: $m_3(+1)$.

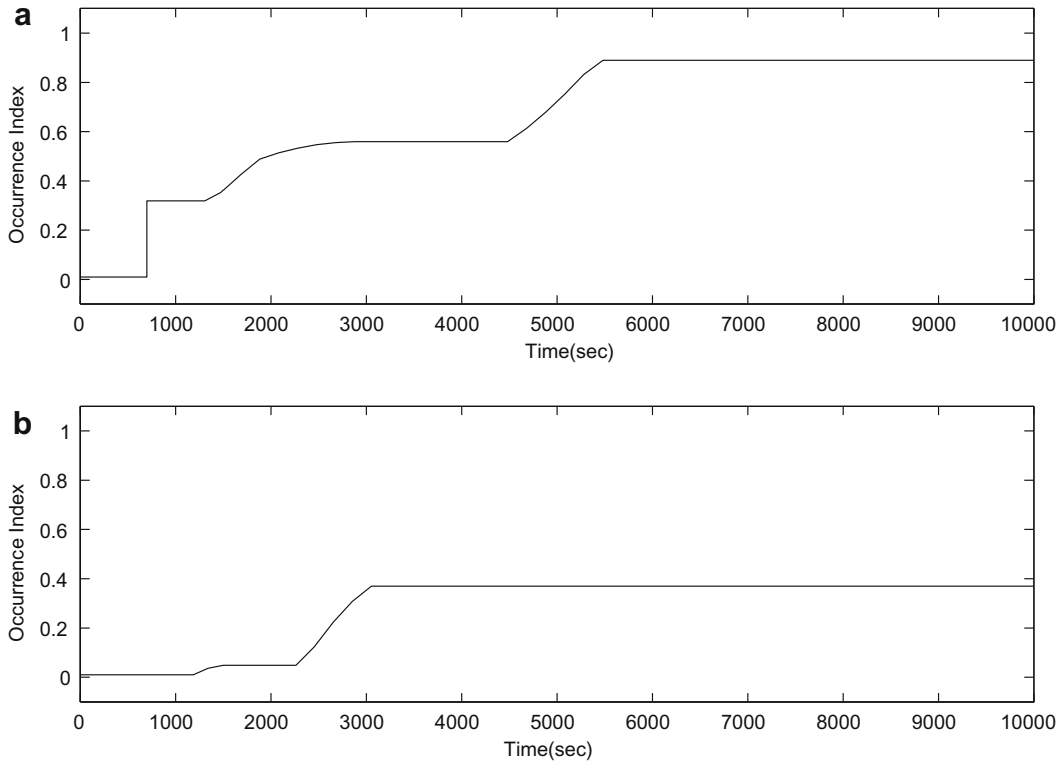

Fig. 8. The diagnosis results obtained with a two-layer inference system concerning the assumed fault origin $m_1(+1)$. (a) Actual fault origin: $m_1(+1)$ and (b) actual fault origin: $m_3(+1)$.

($m_2 \rightarrow h$ and $m_2 \rightarrow s_7 \rightarrow s_6 \rightarrow m_1 \rightarrow h$), (2) a positive feedback loop ($m_2 \rightarrow s_7 \rightarrow s_6 \rightarrow m_1 \rightarrow h \rightarrow m_2$), and (3) a negative feedback loop ($m_2 \rightarrow h \rightarrow m_2$). The mathematical model and the parameter values used in the numerical simulation studies are listed in Appendix B.1 and Table B.1 respectively. It is assumed in this example that flow sensors are installed on lines 1 and 2 in this system and, thus, the

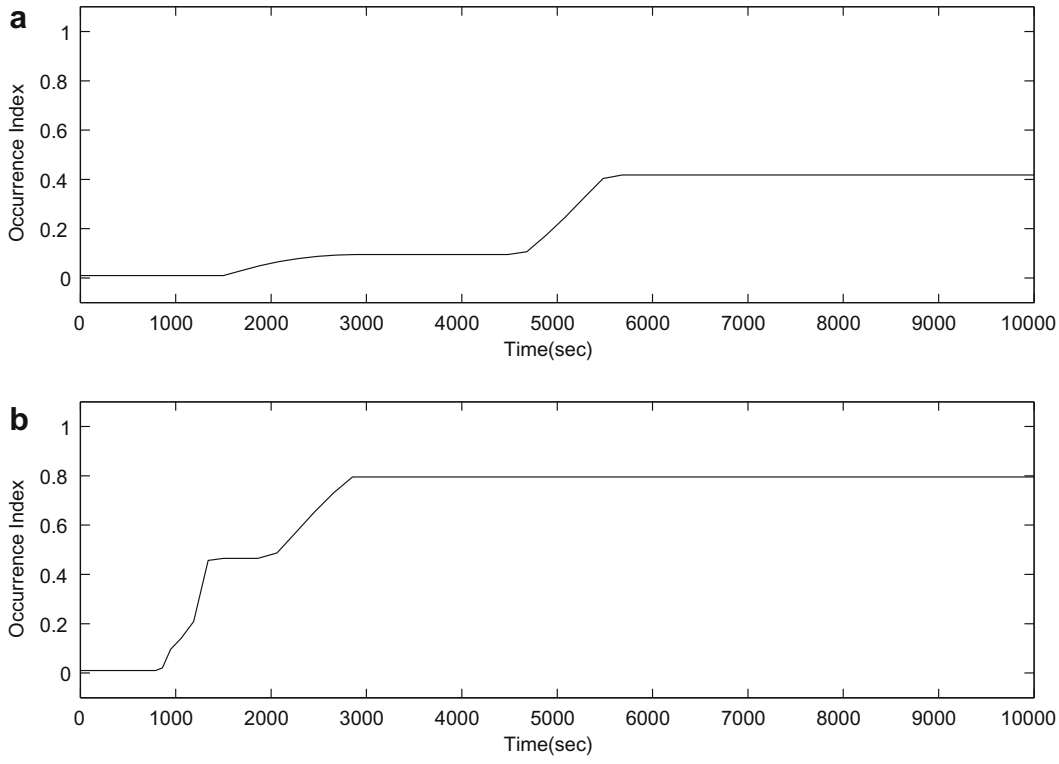

Fig. 9. The diagnosis results obtained with a two-layer inference system concerning the assumed fault origin $m_3(+1)$. (a) Actual fault origin: $m_1(+1)$; (b) Actual fault origin: $m_3(+1)$.

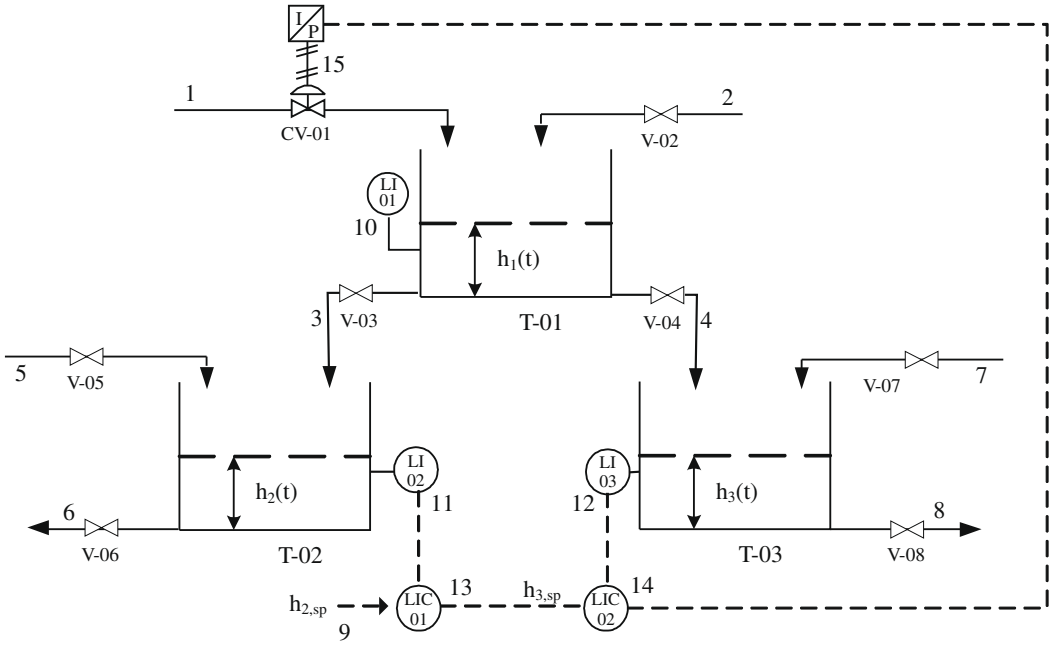

Fig. 10. The process flow diagram of a three-tank storage system with parallel cascade-control loops.

measurement signals s_8 and s_7 can be considered to be available for fault diagnosis purpose.

Two fault propagation scenarios are studied in this example. A partial malfunction in the control valve CV-01 is introduced in the first case to cause an increase in the inlet flow rate in line 1, i.e., $m_1(+1)$, while the initially-closed hand valve V-03 on line 3 is accidentally opened in the second case to produce an additional input flow, i.e., $m_3(+1)$. The SOOs in the aforementioned two scenarios can therefore be predicted with the qualitative simulation techniques [16], i.e.,

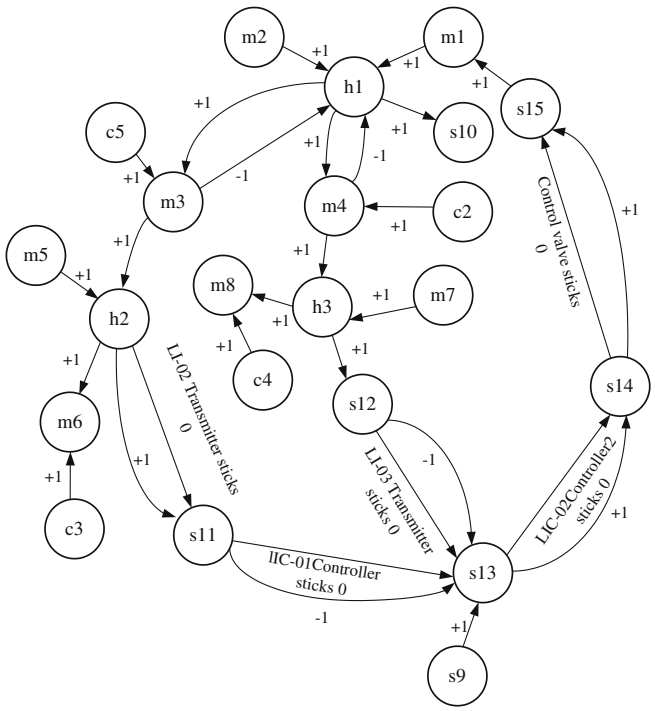

Fig. 11. The SDG model of three-tank storage system.

- Origin 1:

$$s_8(+1) \prec s_7(+1) \prec \cdots \prec [s_7(+10), s_8(+10)]_f$$

- Origin 2:

$$s_7(+1) \prec s_8(+1) \prec \cdots \prec [s_7(+10), s_8(+10)]_f$$

Notice that the final states of the sensor outputs, i.e., $s_7(+10)$ and $s_8(+10)$, are the same in both SOOs. The corresponding candidate patterns are presented in Table 7. Since the SOO in each case forms a simple single path, there should be only one corresponding PES. These two PESs can be written respectively as: $1 \prec 2 \prec 3 \prec 4$ (origin 1) and $1 \prec 2' \prec 3 \prec 4$ (origin 2). Notice that patterns *2* and $2'$ in these two sequences are not the same.

The aforementioned two faults were simulated numerically in two separate runs by setting the valve coefficient C_1 to 1.1 at 700 s (scenario 1) and the flow rate m_3 to 250 cm^3/s at 700 s (scenario 2). The simulated on-line measurement data are presented in Fig. 5. It can be observed that the final steady-state values of each measurement in the two cases are almost the same. In addition, notice that there

Table 8
Candidate patterns caused by fault origins $m_5(+10)$ and $m_4(-10)$ in Example 2.

No.	s_{11}	s_{12}
Origin 1: $m_5(+10)$		
1	0	0
2	1	0
3	1	−1
4	10	0
5	10	−1
6	10	−10
7	1/10	−10
Origin 2: $m_4(-10)$		
1	0	0
$2'$	0	−1
3	1	−1
$4'$	0	−10
$5'$	1	−10
6	10	−10
7	1/10	−10

exists a discontinuity in the first derivative of m_2 in each scenario. This is due to the facts that the outlet flow is gravity driven and the tank is overflowed at the time when the discontinuity occurs.

The single-layer inference results are shown in Figs. 6 and 7. It is obvious that the given two fault origins are indistinguishable with this approach. The misjudgement can be mainly attributed to the almost identical final steady-state measurements in the two cases. On the other hand, the diagnostic resolution can be significantly improved with the two-layer strategy. More satisfactory diagnosis results in both scenarios can be found in Figs. 8 and 9, respectively. The fault origins are differentiated by taking into consideration of the additional information embedded in the PESs. Specifically, the event $m_1(+1)$ must occur before the increase in liquid level and $m_2(+1)$ in the first scenario, while the fault origin $m_3(+1)$ in the second scenario must first cause an increase in liquid level and $m_2(+1)$ before the change in the flow rate of line 1, i.e., $m_1(+1)$, becomes observable.

5.2. A three-tank storage system with parallel cascade-control loops

The second example is taken from Chen et al. [24]. The flow diagram of this storage system and the corresponding SDG model are shown respectively in Figs. 10 and 11. A mathematical model has been built to facilitate simulation studies (see Appendix B.2). The

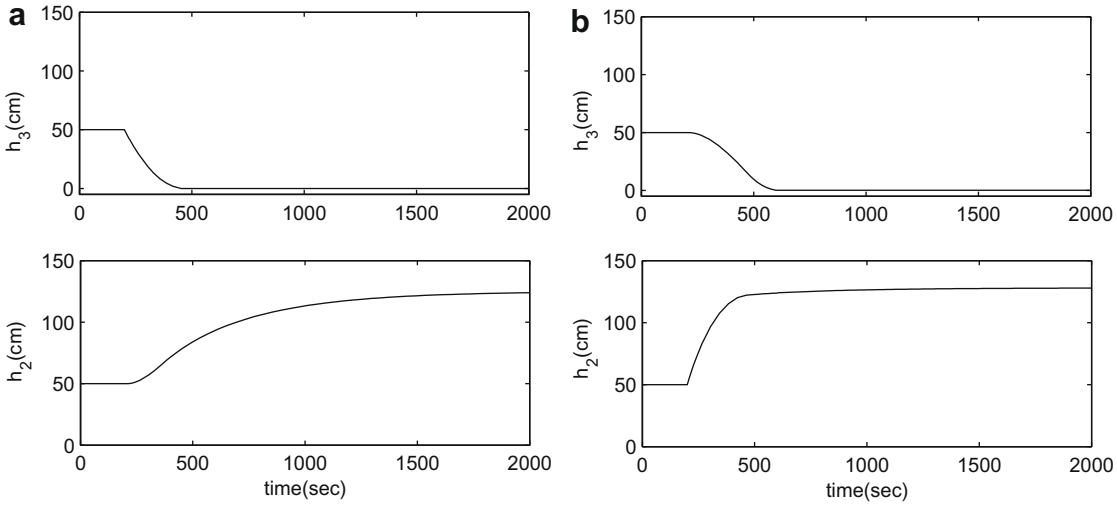

Fig. 12. (a) The simulated dynamic responses of $m_4(-10)$ and (b) the simulated dynamic responses of $m_5(+10)$.

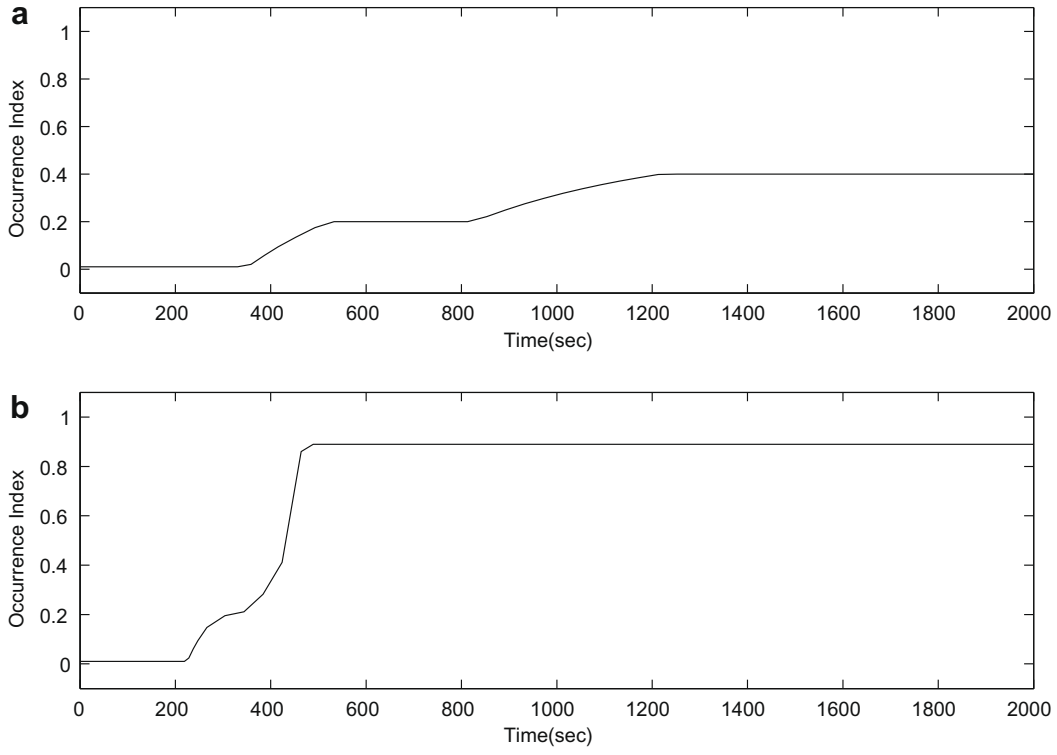

Fig. 13. The diagnosis results obtained with a two-layer inference system concerning the assumed fault origin $m_5(+10)$. (a) Actual fault origin: $m_4(-10)$ and (b) actual fault origin: $m_5(+10)$.

model parameters are provided in Table B.2 in the same appendix. Only two on-line measurements are assumed to be available in this example. Specifically, they are the outputs of level sensors on tank 2 and tank 3, i.e., s_{11} and s_{12}.

Two fault origins are studied here: (1) a large additional flow in line 5, i.e., $m_5(+10)$, and (2) a sudden blockage in line 4 causing its flow to stop, i.e., $m_4(-10)$. The corresponding composite SOOs are given below:

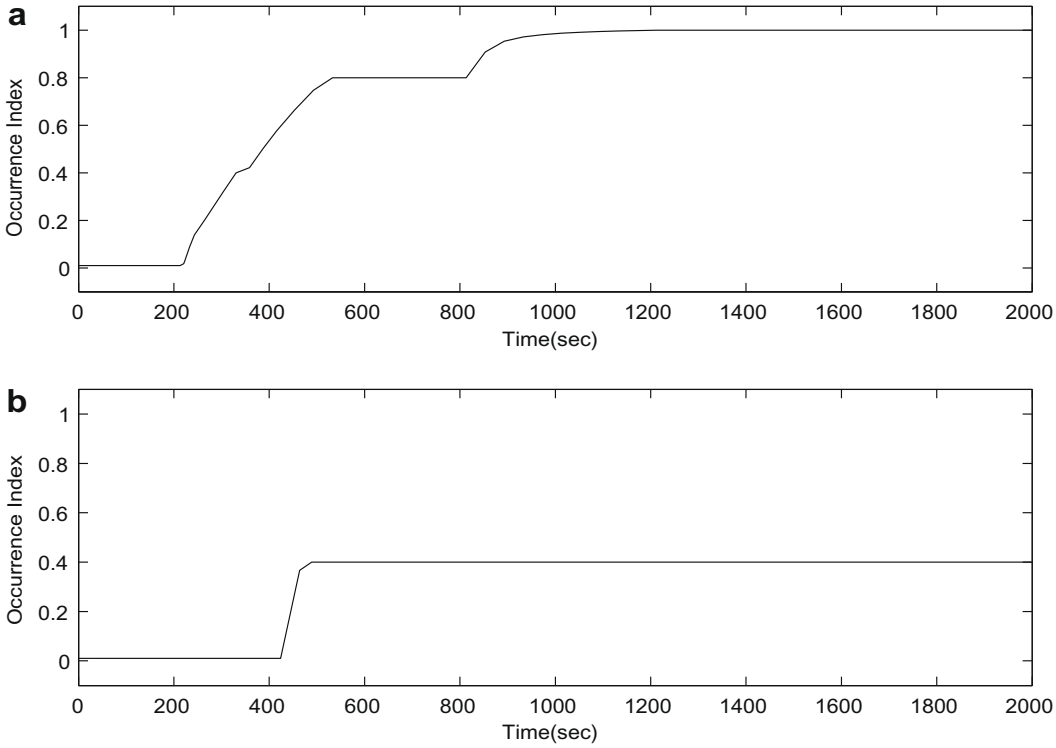

Fig. 14. The diagnosis results obtained with a two-layer inference system concerning the assumed fault origin $m_4(-10)$. (a) Actual fault origin: $m_4(-10)$ and (b) actual fault origin: $m_5(+10)$.

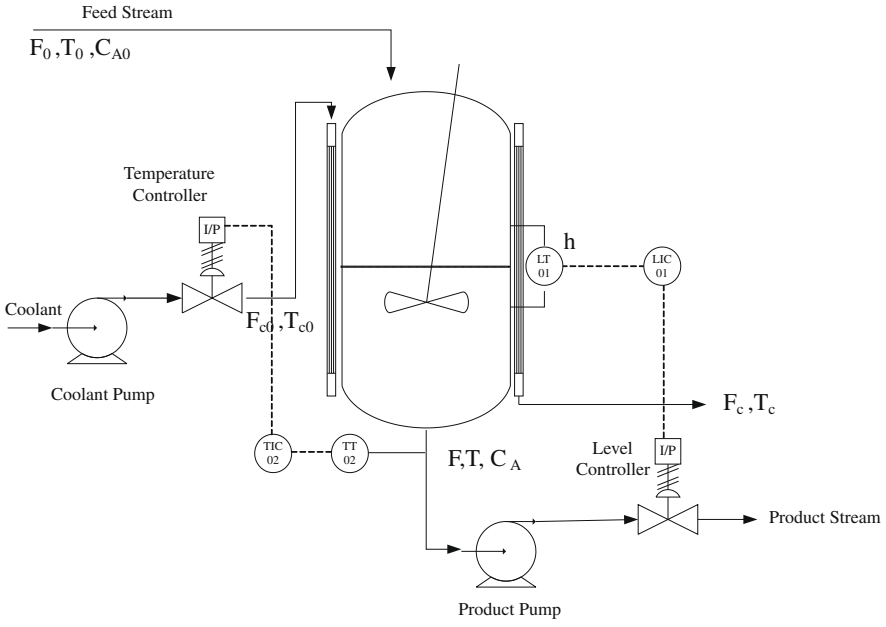

Fig. 15. The process flow diagram of a CSTR reactor system with level and temperature control loops.

- Origin 1:

$$s_{11}(+10) \prec\prec s_{12}(-10) \prec\prec \cdots \prec\prec [s_{11}(+1/+10), s_{12}(-10)]_f$$

- Origin 2:

$$s_{12}(-10) \prec\prec s_{11}(+10) \prec\prec \cdots \prec\prec [s_{11}(+1/+10), s_{12}(-10)]_f$$

The candidate patterns in these two scenarios can be identified accordingly (see Table 8). By applying the PES generation algorithm, the following sequences can be constructed:

- Origin 1:

$1 \prec 2 \prec 3 \prec 5 \prec 6 \prec 7$
$1 \prec 2 \prec 4 \prec 5 \prec 6 \prec 7$

- Origin 2:

$1 \prec 2' \prec 3 \prec 5' \prec 6 \prec 7$
$1 \prec 2' \prec 4' \prec 5' \prec 6 \prec 7$

The aforementioned faults were simulated in two separate runs by setting m_5 to be *200* cm^3/s at 200 s and c_2 to be *0* at 200 s, respectively. The simulated on-line measurements are given in Fig. 12. Again it can be observed that the steady-state measurements in both cases are almost the same. As a result, the two fault origins cannot be differentiated with the single-layer inference scheme. The diagnostic performance can be enhanced with a two-layer FIS, which is clearly demonstrated in the simulation results shown in Figs. 13 and 14.

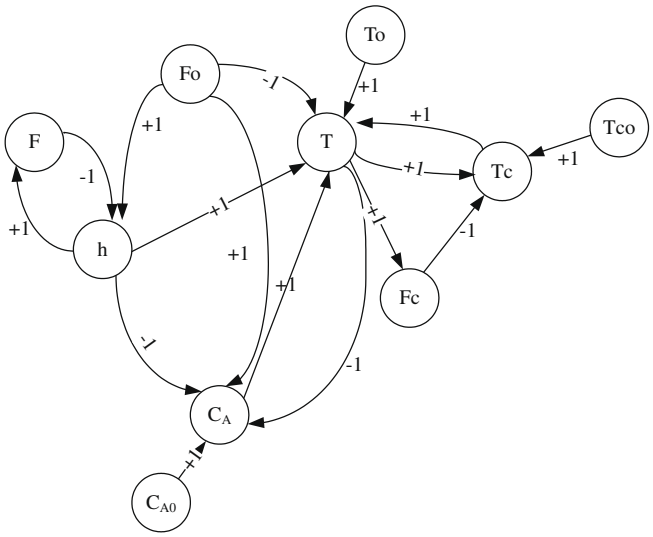

Fig. 16. The SDG model of CSTR reactor system.

a

$$C_A(+1) \prec T(+1) \prec Tc^{(1)}(-1) \cdots \prec [C_A(-1/0/+1), T(0), T_c(0)]_f$$

b

$$Tc(+1) \prec T(+1) \prec C_A(-1) \cdots \prec [C_A(-1/0/+1), T(0), T_c(0)]_f$$

Fig. 17. (a) The SOO resulting from fault origin $C_{A0}(+1)$ and (b) the SOO resulting from fault origin $T_{c0}(+1)$.

5.3. A CSTR with level and temperature control loops

The final example discussed in this paper is concerned with an exothermic CSTR reactor with its temperature and level control loops (see Fig. 15) [25]. It is assumed that there are three measurable process variables, the temperature of cooling water (T_c), the temperature and reactant concentration at the outlet of CSTR (T and C_A). For simplicity, it is further assumed that the variations in these variables are always accurately reflected in their measurements and, therefore, it is not necessary to distinguish a measured variable from its measurement signal in the digraph model. The resulting SDG is given in Fig. 16. The mathematical model and its parameters are presented in Appendix B.3 and Table B.3, respectively.

The fault origins considered in this example are: (1) a moderate disturbance in the input concentration, i.e., $C_{A0}(+1)$, and (2) a small increase in the upstream temperature of cooling water, i.e., $T_{c0}(+1)$. The corresponding SOOs and candidate patterns can be found in Fig. 17 and Table 9 respectively. Notice that, in the latter table, pattern *7* and pattern 5′ are essentially the same. In addition, since T_c is affected by a disturbance in two opposite directions via T (see Fig. 18), its value in pattern *6* is really indeterminable and, thus, all three possibilities (i.e., −1, *0* and +1) are adopted. This indiscriminate assignment will seriously undermine the diagnostic capability of single-layer FIS. On the other hand, one can still establish the PESs resulting from the above two fault origins, i.e.,

- Origin 1:

$1 \prec 2 \prec 3 \prec 5 \prec 6 \prec 7$
$1 \prec 2 \prec 3 \prec 4 \prec 6 \prec 7$

- Origin 2:

$1 \prec 2' \prec 3' \prec 4' \prec 5'$

These two fault propagation scenarios were simulated by setting the value of C_{A_0} to be *0.97* lb-mol/ft^3 at *3* h and T_{c0} to be *540* °R at *3* h, respectively, in two separate runs. Due to the aforementioned deficiencies in determining the definite candidate patterns, the diagnosis results obtained with the single-layer system were in fact erratic. However, from the simulated measurement data given in Figs. 18 and 19, one can clearly see that the initial transients of C_A and T_c are quite different in the two scenario. This unique feature provides the two-layer FIS with a good opportunity to improve diagnostic resolution. Indeed, more satisfactory diagnosis results can be obtained on the basis of the same measurement data and these results are presented in Figs. 20 and 21.

Table 9
Candidate patterns caused by fault origins $C_{A0}(+1)$ and $T_{c0}(+1)$ in Example 3.

No.	C_A	T	T_c
Origin 1: $C_{A0}(+1)$			
1	0	0	0
2	1	0	0
3	1	1	0
4	1	1	1
5	1	1	−1
6	1	1	−1/0/+1
7	−1/0/+1	0	0
Origin 2: $T_{c0}(+1)$			
1	0	0	0
2′	0	0	1
3′	0	1	1
4′	−1	1	1
5′	−1/0/+1	0	0

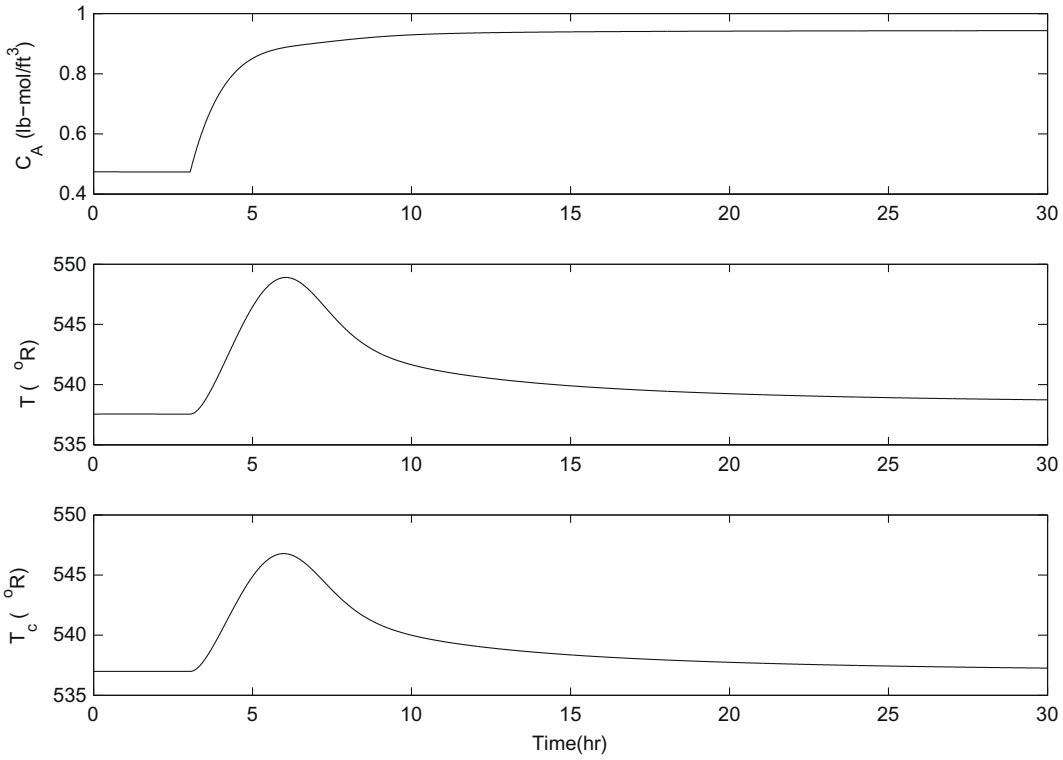

Fig. 18. The simulated dynamic responses of $C_{A0}(+1)$.

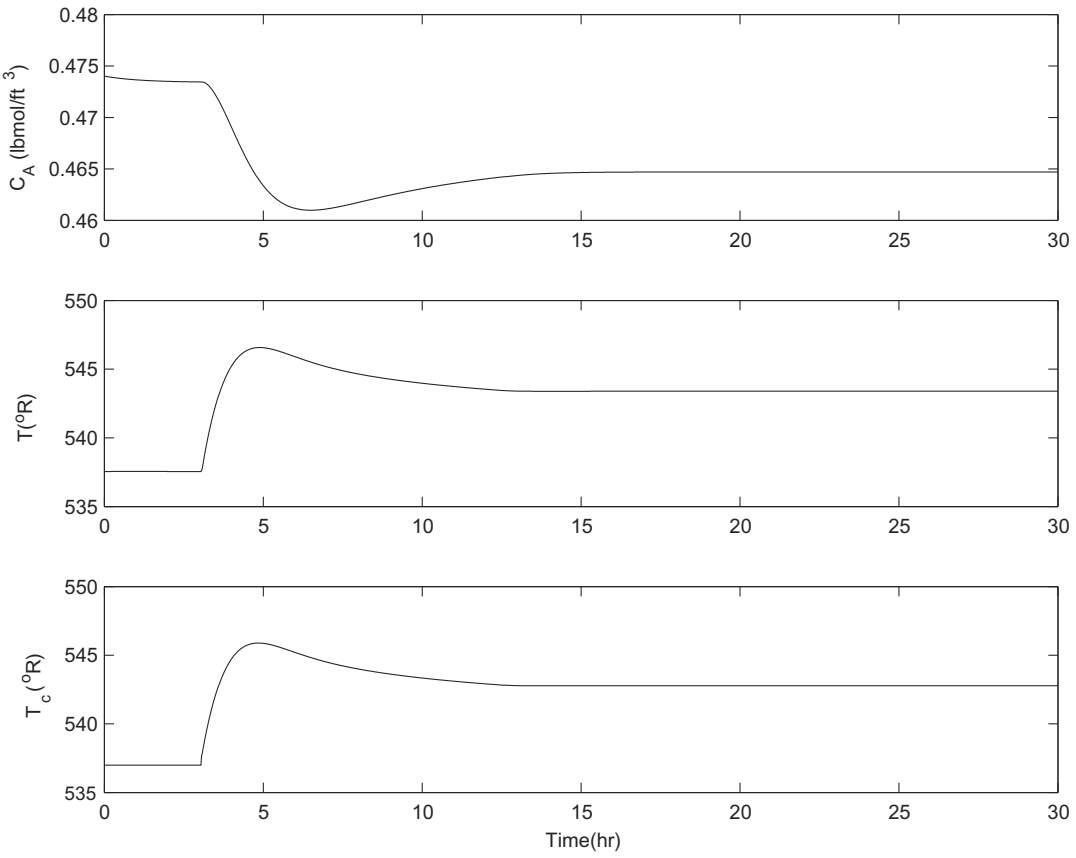

Fig. 19. The simulated dynamic responses of $T_{c0}(+1)$.

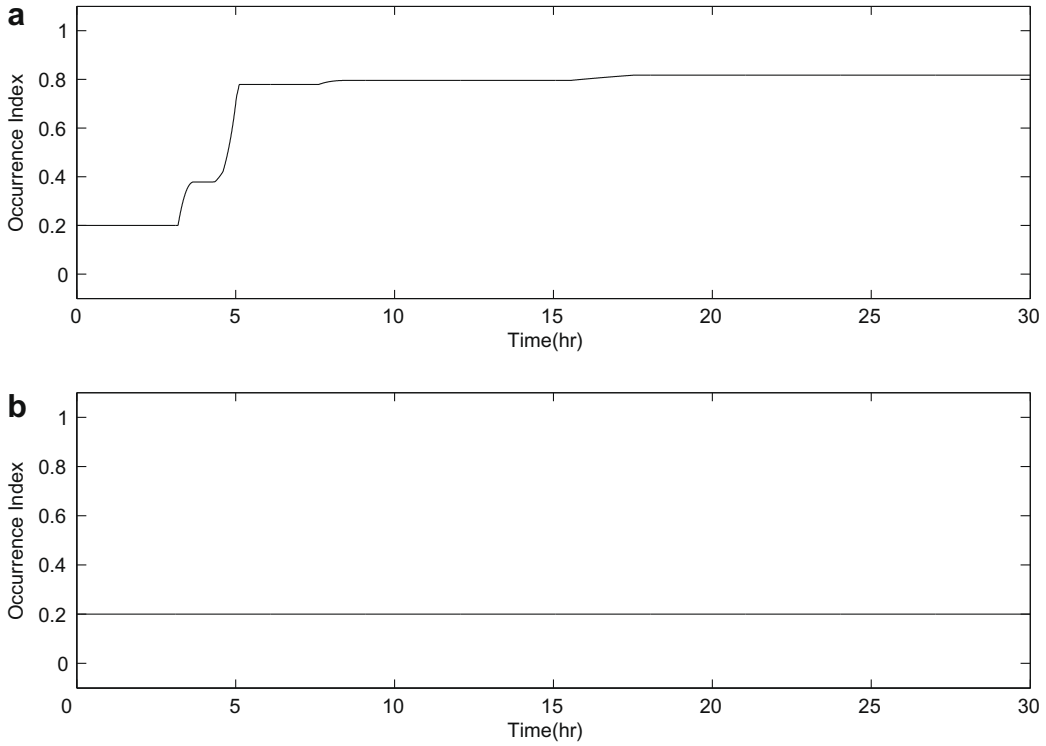

Fig. 20. The diagnosis results obtained with a two-layer inference system concerning the assumed fault origin $C_{A0}(+1)$. (a) Actual fault origin: $C_{A0}(+1)$; (b) Actual fault origin: $T_{c0}(+1)$.

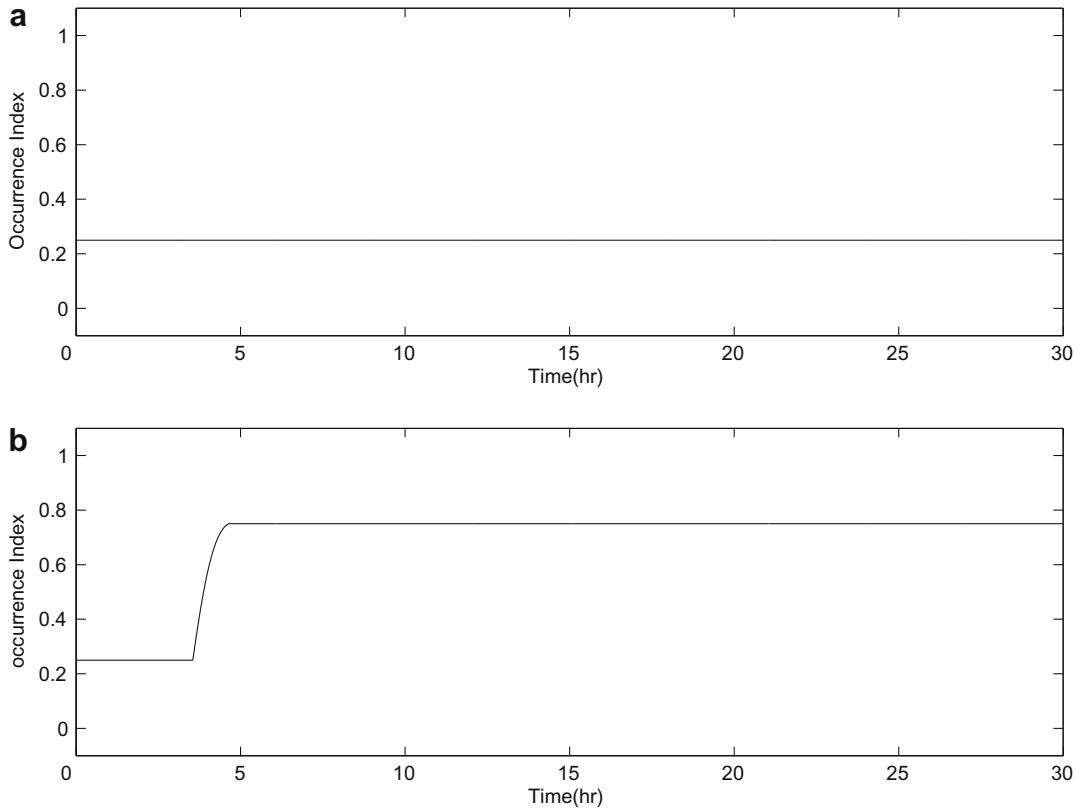

Fig. 21. The diagnosis results obtained with a two-layer inference system concerning the assumed fault origin $T_{c0}(+1)$. (a) Actual fault origin: $C_{A0}(+1)$; (b) Actual fault origin: $T_{c0}(+1)$.

6. Conclusions

The fault propagation paths (FPPs) and symptom occurrence orders (SOOs) caused by fault origins with one or more possible magnitude level are characterized systematically in this study according to (1) the degree of deviation of each process variable from its normal state and (2) the occurrence order of the abnormal deviations. In particular, a SDG-based reasoning procedure is proposed to qualitatively predict all possible symptom patterns and also their progression sequences. These intrinsic features of the symptom evolution behaviors are then captured with IF–THEN rules in a two-layer fuzzy inference system. This system can be used to identify not only the locations of fault origins but also their magnitude levels with relatively high resolution. The feasibility and effectiveness of the proposed diagnostic strategy are clearly demonstrated in extensive simulation studies.

Appendix A. Derivation of Eqs. (11)–(15)

Let us first consider the simplest tree-shaped SOO, in which two first-tier branch paths, $\mathbf{P}^{(0,i_1)}(1,n_{0,i_1})$ ($i_1 = 1,2$), are connected to the end of initial path $\mathbf{P}^{(0)}(1,n_0)$. Although the occurrence order of symptoms along $\mathbf{P}^{(0)}(1,n_0)$ should be unique, the order of events on the two branched paths, i.e., $\mathbf{P}^{(0,1)}(1,n_{0,1})$ and $\mathbf{P}^{(0,2)}(1,n_{0,2})$, is not defined exactly. If the two separate symptom occurrence orders specified in the branches are ignored, the total number of *all* possible evolution sequences should be $(n_{0,1}+n_{0,2})!$. However, since the precedence order within a branch path is fixed, the corresponding permutations should be excluded, i.e., the total PES number in this case should be

$$N_{PES} = \frac{(n_{0,1}+n_{0,2})!}{n_{0,1}!n_{0,2}!} \tag{A.1}$$

Next let us consider the case in which B_0 first-tier branches are connected to the initial path. The corresponding PES number can be determined on the basis of the same rationale, i.e.,

$$N_{PES} = \left(\sum_{i_1=1}^{B_0} n_{0,i_1}\right)! \Big/ \prod_{i_1=1}^{B_0} (n_{0,i_1}!) \tag{A.2}$$

The above derivation can be extended to SOOs with both first- and second-tier branches. For illustration convenience, let us consider a special case when $B_0 = B_{0,1} = B_{0,2} = 2$. After the occurrence of the last symptom in $\mathbf{P}^{(0,1)}(1,n_{0,1})$, the number of subsequent pattern evolution sequences can be determined on the basis of Eq. (A.1), i.e.,

$$N_{PES}^{(0,1)} = \frac{(n_{0,1,1}+n_{0,1,2})!}{n_{0,1,1}!n_{0,1,2}!} \tag{A.3}$$

Notice that the SOO associated with one of these subsequent sequences can be augmented with that along $\mathbf{P}^{(0,1)}(1,n_{0,1})$ to form a single-path pseudo SOO. There are of course $\frac{(n_{0,1,1}+n_{0,1,2})!}{n_{0,1,1}!n_{0,1,2}!}$ such pseudo SOOs. Similarly, the same approach can be used to compute the number of PESs which may appear after all symptoms in $\mathbf{P}^{(0,2)}(1,n_{0,2})$, i.e.,

$$N_{PES}^{(0,2)} = \frac{(n_{0,2,1}+n_{0,2,2})!}{n_{0,2,1}!n_{0,2,2}!} = \binom{n_{0,2,1}+n_{0,2,2}}{n_{0,2,2}} \tag{A.4}$$

and to construct the corresponding pseudo SOOs. Therefore, the aforementioned tree-shaped SOO can be viewed as an ensemble of different SOOs with two first-tier pseudo branches. The total number of all corresponding PES number should be

$$N_{PES} = \frac{(L_{0,1}+L_{0,2})!}{L_{0,1}!L_{0,2}!} N_{PES}^{(0,1)} N_{PES}^{(0,2)} \tag{A.5}$$

where $L_{0,1}$ and $L_{0,2}$ denote, respectively, the lengths of pseudo SOOs associated with $\mathbf{P}^{(0,1)}(1,n_{0,1})$ and $\mathbf{P}^{(0,2)}(1,n_{0,2})$. More specifically,

$$L_{0,1} = n_{0,1} + n_{0,1,1} + n_{0,1,2} \tag{A.6}$$

$$L_{0,2} = n_{0,2} + n_{0,2,1} + n_{0,2,2} \tag{A.7}$$

Finally, it should be noted that the generalized version of Eq. (A.5) can be derived with exactly the same procedure. These formulas are expressed in more compact form with the operators $\mathscr{K}\{\cdot\}$, $\mathscr{R}\{\cdot\}$ and $\mathscr{L}\{\cdot\}$ defined in Eqs. (9)–(11).

Appendix B. Process models

B.1. Mathematical model used for simulating single-tank system with feed forward control

$$A\frac{dh}{dt} = m_1 + m_3 - m_2 \tag{B.1}$$

$$m_2 = C_2\sqrt{h_1} \tag{B.2}$$

$$m_1 = C_1(m_{1s} + K_{cv}K_c[(m_{2_{set}} - m_2)]) \tag{B.3}$$

Notice that the model parameters in the above equations can be found in Table B.1.

B.2. Mathematical model used for simulating three tank system with parallel cascade control

$$A_1\frac{dh_1}{dt} = m_1 + m_2 - m_3 - m_4 \tag{B.4}$$

$$A_2\frac{dh_2}{dt} = m_3 + m_5 - m_6 \tag{B.5}$$

$$A_3\frac{dh_3}{dt} = m_4 + m_7 - m_8 \tag{B.6}$$

$$m_3 = c_5 * C_1\sqrt{h_1} \tag{B.7}$$

$$m_4 = c_2 * C_2\sqrt{h_1} \tag{B.8}$$

$$m_6 = c_3 * C_3\sqrt{h_2} \tag{B.9}$$

$$m_8 = c_4 * C_4\sqrt{h_3} \tag{B.10}$$

$$h_{3,set} = h_{3,set(s)} + K_{c1}\left[(h_{2,set} - h_2) + \frac{1}{\tau_{I1}}\int_0^t (h_{2,set} - h_2)\right] \tag{B.11}$$

$$m_1 = c_1 * \left(m_{1s} + K_{cv}K_{c2}\left[(h_{3,set} - h_3) + \frac{1}{\tau_{I2}}\int_0^t (h_{3,set} - h_3)\right]\right) \tag{B.12}$$

Table B.1
Model parameters used in Example 1.

Parameter	Definition	(Steady state) value
A	Cross-section area of tank	5000 cm^2
h	Height of liquid level in tank	50 cm
m_1	Input flow rate	707.11 cm^3/s
m_2	Output flow rate	707.11 cm^3/s
m_3	Additional input flow rate	0 cm^3/s
C_1	Input control valve constant	1
C_2	Proportional constant	100 $cm^{2.5}/s$
K_{cv}	Control valve gain	15 cm^3/s %
$m_{2_{set}}$	Set point of output flow rate	707.11 cm^3/s
K_c	Proportional gain of level controller	0.0667

Table B.2
Model parameters used in Example 2.

Parameter	Definition	(Steady state) value
A_1	Cross-section area of tank 1	2000 cm^2
h_1	Height of liquid level in tank 1	100 cm
A_2	Cross-section area of tank 2	2000 cm^2
h_2	Height of liquid level in tank 2	50 cm
A_3	Cross-section area of tank 1	2000 cm^2
h_3	Height of liquid level in tank 1	50 cm
m_1	Input flow rate of tank 1	1500 cm^3/s
m_3	Output flow rate of tank 1	750 cm^3/s
m_4	Output flow rate of tank 1	750 cm^3/s
m_6	Output flow rate of tank 2	750 cm^3/s
m_8	Output flow rate of tank 3	750 cm^3/s
m_2	Additional input flow rate to tank 1	0 cm^3/s
m_5	Additional input flow rate to tank 2	0 cm^3/s
m_7	Additional input flow rate to tank 3	0 cm^3/s
C_1,C_2	Proportional constants	150 $cm^{2.5}/s$
C_3,C_4	Proportional constants	106.07 $cm^{2.5}/s$
$c_1 \sim c_5$	Proportional constants	1
K_{cv}	Control valve gain	30 cm^3/s %
$h_{2,set}$	Set point of the level height in tank 2	50 cm
$h_{3,set}$	Set point of the level height in tank 3	50 cm
K_{c1}	Proportional gain of level controller1	2.48
K_{c2}	Proportional gain of level controller2	1
τ_{I1}	Integral time of level controller1	50×10^{-4} s
τ_{I2}	Integral time of level controller2	2×10^{-4} s

Table B.3
Model parameters used in Example 3.

Parameter	Definition	(Steady state) value
h	Height of liquid level in reactor	48 ft
V	Reactor volume	4800 ft^3
C_{A_0}	Reactant concentration in feed	0.47 lb-mol/ft^3
T	Reactor temperature	537 °R
F_0	Feed flow rate	4000 ft^3/h
T_0	Feed temperature	530 °R
C_A	Reactant concentration in reactor	0.474 lb-mol/ft^3
T_c	Outlet Coolant temperature	537 °R
F_c	Coolant flow rate	4836 ft^3/h
V_j	Volume of jacket	385 ft^3
k_0	Frequency factor	7.08×10^{10}/h
E	Activation energy	30,000 Btu/lb-mol
R	Universal gas constant	1.99 Btu/lb-mol °R
U	Overall heat transfer coefficient	150 Btu/h ft^2 °R
A	Heat transfer area	25,000 ft^2
T_{c0}	Inlet coolant temperature	530 °R
ΔH	Heat of reaction	−30,000 Btu/lb-mol
C_p	Heat capacity (process side)	0.72 Btu/lb_m °R
C_j	Heat capacity (coolant side)	1 Btu/lb_m °R
ρ	Density of process mixture	50 lb_m/ft^3
ρ_j	Density of coolant	62.3 lb_m/ft^3
A_r	Cross-section area of reactor	100 ft^2
K_c^H	Proportional gain of level controller	10
K_c^T	Proportional gain of temperature controller	80
τ_I^H	Integral time of level controller	89.286 h
τ_I^T	Integral time of temperature controller	0.6557 h
h_{set}	Set point of the level height in tank	48 ft
T_{set}	Set point of the temperature in tank	537 °R

Notice that the model parameters in the above equations can be found in Table B.2.

B.3. Mathematical model used for simulating CSTR reactor system

$$\frac{dV}{dt} = F_0 - F \tag{B.13}$$

$$V = A_r h \tag{B.14}$$

$$r_A = k_0 e^{-\frac{E}{RT}} C_A \tag{B.15}$$

$$\frac{dC_A}{dt} = \frac{F_0}{V}(C_{A_0} - C_A) - r_A \tag{B.16}$$

$$\frac{dT}{dt} = \frac{F_0}{V}(T_0 - T) + \frac{r_A(-\Delta H)}{\rho C_p} - \frac{UA(T - T_c)}{V\rho C_p} \tag{B.17}$$

$$\frac{dT_c}{dt} = \frac{F_c}{V_j}(T_{c0} - T_c) + \frac{UA(T - T_c)}{V_j \rho_j C_j} \tag{B.18}$$

$$F_c = F_{cs} - K_c^T\left[(T_{set} - T) + \frac{1}{\tau_I^T}\int_0^t (T_{set} - T)\right] \tag{B.19}$$

$$F = F_s - K_c^H\left[(h_{set} - h) + \frac{1}{\tau_I^H}\int_0^t (h_{set} - h)\right] \tag{B.20}$$

Notice that the model parameters in the above equations can be found in Table B.3.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jprocont.2008.11.006.

References

[1] C. Chang, J. Chen, Implementation issues concerning the ekf-based fault diagnosis technique, Chemical Engineering Science 50 (1995) 2861.
[2] T. Petti, J. Klein, P. Dhurjati, Diagnostic model processor: using deep knowledge for process fault diagnosis, AIChE Journal 36 (1990) 565.
[3] J. Hoskins, K. Kalivur, D. Himmeblau, Fault diagnosis in complex chemical-plants using artificial neural networks, AIChE Journal 37 (1991) 137.
[4] M. Kramer, B. Polowitch, Rule-based approach to fault diagnosis using the signed directed graph, AIChE Journal 33 (7) (1987) 1067–1078.
[5] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics and Intelligent Laboratory Systems 2 (1) (1987) 37.
[6] J. Wang, H. Hu, Vibration-based fault diagnosis of pump using fuzzy technique, Measurement 39 (2006) 185.
[7] M. Catelani, A. Fort, C. Alippi, A fuzzy approach for soft fault detection in analog circuits, Measurement 32 (2002) 73.
[8] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S. Kavuri, A review of process fault detection and diagnosis, part I: Quantitative model based methods, Computers and Chemical Engineering 27 (2003) 293–311.
[9] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, A review of process fault detection and diagnosis, part II: Qualitative model and search strategies, Computers and Chemical Engineering 27 (2003) 313–326.
[10] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, K. Yin, A review of process fault detection and diagnosis, part III: Process history based methods, Computers and Chemical Engineering 27 (2003) 327.
[11] M. Maurya, R. Rengaseamy, V. Venkatasubramanian, A signed directed graph-based systematic framework for steady state malfunction diagnosis control loops, Chemical Engineering Science 61 (2006) 1790–1810.
[12] M. Maurya, R. Rengaswamy, V. Venkatasubramanian, Application of signed digraph-based analysis for fault diagnosis chemical process flowsheets, Engineering Applications of Artificial Intelligence 17 (2004) 501.
[13] S. Chang, C. Chang, A fuzzy-logic based fault diagnosis strategy for process control loops, Chemical Engineering Science 58 (2003) 3395.
[14] S. Chang, C. Lin, C. Chang, A fuzzy diagnosis approach using dynamic fault trees, Chemical Engineering Science 57 (2002) 2971.
[15] J. Chen, C. Chang, Fuzzy diagnosis method for control systems with coupled feed forward and feedback loops, Chemical Engineering Science 61 (2006) 3105–3128.
[16] J. Chen, C. Chang, Systematic enumeration of fuzzy diagnosis rules for identifying multiple faults in chemical processes, Industrial and Engineering Chemistry Research 46 (2007) 3635–3655.
[17] M. Iri, K. Aoki, E. O'Shima, H. Matsuyama, An algorithm for diagnosis of system failure in the chemical process, Computers and Chemical Engineering 3 (1979) 489.
[18] J. Shiozaki, H. Matsuyama, E. O'Shima, M. Iri, An improved algorithm for diagnosis of system failures in the chemical process, Computers and Chemical Engineering 9 (1985) 285.
[19] E. Tarifa, N. Scenna, Fault diagnosis for msf dynamic states using a sdg and fuzzy logic, Desalination 166 (2004) 93.
[20] Y. Tsunge, J. Shiozaki, H. Matuyama, E. O'Shima, Fault diagnosis algorithms based on the signed directed graph and its modifications, Institute of Chemical Engineering Symposium Series 92 (1985) 133.
[21] Z. Zhang, C. Wu, B. Zhang, Y. Xia, A. Li, Sdg multiple fault diagnosis by real-time inverse inference, Reliability Engineering and System Safety 87 (2005) 173–189.
[22] Mathworks, Simulink 6 using simulink, The Mathworks Inc., 2006.
[23] Mathworks, Fuzzy logic toolbox user guid 2, The Mathworks Inc., 2006.
[24] J. Chen, Y. Yea, C. Kong, Diagnosis of cascade control loop status using performance analysis based approach, Industrial and Engineering Chemistry Research 45 (2006) 7540–7551.
[25] S. Dash, R. Rengaswamy, V. Venkatasubramanian, Fuzzy-logic based trend classification for fault diagnosis of chemical process, Computers and Chemical Engineering 27 (2003) 347.