A Novel Dynamic Bayesian Network-Based Networked Process Monitoring Approach for Fault Detection, Propagation Identification, and Root Cause Diagnosis

Jie Yu and Mudassir M. Rashid
Dept. of Chemical Engineering, McMaster University, Hamilton, ON L8S 4L7, Canada

DOI 10.1002/aic.14013
Published online January 24, 2013 in Wiley Online Library (wileyonlinelibrary.com)

A novel networked process monitoring, fault propagation identification, and root cause diagnosis approach is developed in this study. First, process network structure is determined from prior process knowledge and analysis. The network model parameters including the conditional probability density functions of different nodes are then estimated from process operating data to characterize the causal relationships among the monitored variables. Subsequently, the Bayesian inference-based abnormality likelihood index is proposed to detect abnormal events in chemical processes. After the process fault is detected, the novel dynamic Bayesian probability and contribution indices are further developed from the transitional probabilities of monitored variables to identify the major faulty effect variables with significant upsets. With the dynamic Bayesian contribution index, the statistical inference rules are, thus, designed to search for the fault propagation pathways from the downstream backwards to the upstream process. In this way, the ending nodes in the identified propagation pathways can be captured as the root cause variables of process faults. Meanwhile, the identified fault propagation sequence provides an in-depth understanding as to the interactive effects of faults throughout the processes. The proposed approach is demonstrated using the illustrative continuous stirred tank reactor system and the Tennessee Eastman chemical process with the fault propagation identification results compared against those of the transfer entropy-based monitoring method. The results show that the novel networked process monitoring and diagnosis approach can accurately detect abnormal events, identify the fault propagation pathways, and diagnose the root cause variables. © 2013 American Institute of Chemical Engineers *AIChE J*, 59: 2348–2365, 2013
Keywords: dynamic Bayesian network, networked process monitoring, fault detection, fault propagation identification, root cause diagnosis, probabilistic inference

Introduction

Process monitoring and fault diagnosis are very important in modern industrial plants due to the increased complexity of process operations along with the advanced automation, control, and optimization systems. Monitoring and assessing plant performance can assist the detection and diagnosis of abnormal events and improve process productivity, energy efficiency, environmental sustainability, product quality, and economic profit margins.[1–3] Traditional process monitoring methods are based on mechanistic models that use state estimation, parameter identification, and parity relation techniques.[4] However, the challenges of these analytical model-based approaches lie in the facts that (1) in-depth process knowledge is required to develop the first-principle models that can accurately characterize the complex physical, chemical, and biological phenomena of different kinds of processes; (2) extensive time and effort are needed to develop those differential equation-based mechanistic models; (3) reliable estimation of model parameters from process data is necessary to mitigate plant-model mismatch and ensure the fidelity of monitoring methods. Although knowledge-based expert systems are proposed as alternatives to mechanistic model-based methods for process monitoring, the heuristic rules used for inferential decisions may be incoherent and incomplete.[5] To overcome the drawbacks of both first-principle model-based and expert system-based monitoring approaches, data-driven multivariate statistical process monitoring (MSPM) methods are developed and have attracted growing attention in the field.[6]

MSPM techniques such as principal component analysis (PCA) and partial least squares (PLS) can be used to project the highly correlated data onto a low-dimensional subspace, and the well-known multivariate statistical control charts are adopted to detect abnormal operating events.[7–13] Moreover, multivariate contribution plots can be generated without prior process knowledge for faulty variable diagnosis.[14] The control limits of the PCA or PLS-based monitoring indices, however, depend on the assumption that the process data follow an approximate multivariate Gaussian distribution, which may not be valid in industrial practices.[15,16] In addition, the variable contribution methods may not explicitly identify the root causes of abnormal events because the faults can propagate throughout the processes due to the

Correspondence concerning this article should be addressed to J. Yu at jieyu@mcmaster.ca.

intricate variable interactions, process dynamics, and closed-loop control systems.[17] Although statistical process monitoring techniques can be integrated with knowledge-based systems for interpreting the diagnosis results of the contribution plots, the comprehensive knowledge base is required for effective process monitoring.[18,19] Machine learning techniques such as artificial neural networks (ANN) and support vector machines (SVM) are applied to fault detection and diagnosis with some successes.[20–23] These black-box models are suitable for the situations where in-depth process knowledge is lacking or difficult to acquire. Nevertheless, the ANN approach cannot readily diagnose process faults because it is difficult to compute the variable contributions while SVM is a supervised monitoring method requiring preclassified training data with both normal and faulty sample labels. A Bayesian classification-based PCA approach is also proposed for fault detection and diagnosis, in which a new fault identification index is developed from the means and covariances of Bayesian clusters within the principal component subspace.[24,25] However, this method is unable to identify the propagation pathways of process faults. Furthermore, a Bayesian inference-based Gaussian mixture model (GMM) method is proposed to handle the multi-Gaussianity of processes, though the data from each individual operating mode is assumed to follow a multivariate Gaussian distribution approximately.[26–28] Aimed at dynamic batch process monitoring, discrete hidden Markov model approach is developed for fault detection and classification by estimating the probabilistic state transitions from observed measurements.[29] To improve the fault diagnosis ability, a semisupervised approach is developed by integrating independent component analysis for dimensionality reduction, GMM, and support vector machine for unsupervised and supervised classifications.[30] Similar to MSPM techniques, the above machine learning-based monitoring methods are usually data driven and do not make use of the preliminary process knowledge. Thus, they may not have the root-cause diagnosis capability, especially in terms of identifying the fault propagation pathways throughout the processes given complex multivariable interactions.

Alternately, an incipient fault diagnosis approach incorporating a two-level hierarchy with signed directed graphs and qualitative trend analysis is proposed to determine candidate faults from a database of all potential faults.[31,32] Nevertheless, diagnosing abnormal operating events that are not within the prior fault database would be challenging and the fault propagation pathway diagnosis is not considered in this type of approach. Meanwhile, causal map-based qualitative models have been developed to identify the cause and effect relationships among process measurement variables.[33,34] The Kullback–Leibner distance is used to measure the similarity of the monitored variables against historical operating conditions for fault detection and then the multivariate T^2 statistic is used to determine the causal dependence between different measurement variables. In this way, the test statistic value between pairs of variables can be evaluated against a threshold to determine if the causal dependency is significant or not.[33] However, the casual dependence index is essentially a second-order statistic that may not capture the non-Gaussian dependency, time-varying dynamics, or process uncertainty. Moreover, the searching for fault propagation pathways begins from the root-cause variable to the downstream effect variables, which requires the root-cause variable to be preidentified. More recently, the data-driven method using transfer entropy is proposed to characterize the causality between two time series and consequently determine the direction of fault propagation. However, the estimation of probability density functions is not a trivial task and can be computationally intensive.[35] Similarly, the nearest neighbors of process measurements are also compared to historical process data for the purpose of identifying the propagation direction of disturbances. The disturbance propagation pathway can be traced throughout the plant from the root cause by detecting interdependency between the time series of measurements, though the implementation relies on the appropriate selection of clustering parameters.[36] A method for the source identification and propagation analysis of plant-wide faults is also designed mainly through the time delay estimation based on nearest neighbor imputation.[37] However, the heavy computational load of this procedure may complicate the online implementation if the sampling frequency is fast. Some research effort has been attempted to extract qualitative information from process schematic descriptions to form connectivity matrix for cause-and-effect diagnosis of the plant-wide disturbances. In this technique, the searching for propagation paths requires the reliable XML descriptions of the process schematics.[38] Another strategy to diagnose the root cause of plant-wide oscillations using the adjacency matrix that incorporates process knowledge in terms of process flow sheet or topology, control configuration, and instrument information is also investigated, though the study is mainly focused on loop oscillation diagnosis.[39]

Dynamic Bayesian networks (DBN) are a type of graphical models for conducting time-varying probabilistic inference and causal analysis under system uncertainty. The network consists of nodes representing random variables along with connecting arcs that quantify the causal relationships among variable nodes.[40,41] Bayesian networks have been applied to different areas including medical diagnosis, gene modeling, cancer classification, sensor validation, risk management, and reliability analysis. However, there are limited literature reports on the use of Bayesian networks for process fault detection and diagnosis.[42–44] In this study, a dynamic Bayesian network-based networked process monitoring and diagnosis approach is proposed to detect the abnormal events in chemical processes, identify the propagation pathways of faults throughout processes, and diagnose the root-cause variables of faulty operation. First, the process network structure is determined from prior process knowledge and process flow diagrams. Then, the network model parameters including the conditional probability density functions of different variable nodes are estimated from plant historical data. Further, a new abnormality likelihood index (ALI) is proposed to quantify the probabilistic likelihood of the entire process to be under abnormal operating conditions. Such likelihood index can be used to monitor chemical processes with stochastic uncertainty and detect different kinds of operational faults. To identify the affected variables along the fault propagation pathways within the process network, a dynamic Bayesian probability index (DBPI) along with its upper and lower control limits is developed for characterizing the faulty behavior or response of each process variable from probabilistic perspective. Subsequently, a novel dynamic Bayesian contribution index (DBCI) is derived and the corresponding statistical inference rules are developed so as to capture the major effect variables of faulty operation

from the downstream back to the upstream process. In this way, the fault propagation pathways throughout the processes can be identified, and the root-cause variables leading to the process upsets can be diagnosed.

The remainder of the article is organized as follows. Section Preliminaries briefly reviews the DBN. Then the novel networked process monitoring approach for fault detection, fault propagation pathway identification, and root cause diagnosis is developed in section Networked Process Monitoring Approach for Fault Detection, Propagation Identification, and Root Cause Diagnosis. Section Case Studies demonstrates the application of the novel networked process monitoring and diagnosis approach to an illustrative continuous stirred tank reactor and the Tennessee Eastman chemical process. Moreover, the fault propagation diagnosis results of the proposed approach in the Tennessee Eastman chemical process are compared against those of the transfer entropy-based monitoring method. The conclusions of this work are summarized in section conclusions.

Preliminaries

DBN can effectively characterize the complex causal relationships among variables for dynamic systems with stochastic uncertainty. A DBN is a directed acyclic graph consisting of hidden and observed nodes with statistical dependencies represented by the arcs connecting the various nodes. The nodes with arcs directed into them are termed as child nodes, whereas the nodes from which the arcs depart are parent nodes. The nodes without any parents are called root nodes. A simple illustrative example of a three-node DBN with interdependency and intradependency is shown in Figure 1.

In addition to the qualitative causal reasoning within the graphical structure, the conditional probability density functions, denoted as $\theta_i \in \Theta$, give the quantitative causal relationships among variables and need to be specified for each node $n_i \in N$.[40] An arbitrary node X_i obeys the conditional probability density function $p\left(X_i|X_{\mathrm{pa}(i)}^{(k)}=y_k\right)$ given the value

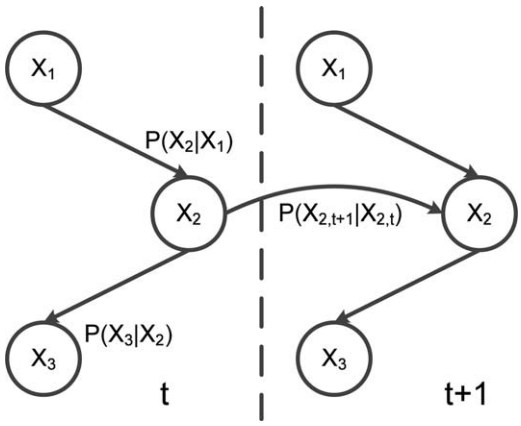

Figure 1. An illustrative example of dynamic Bayesian network.

of its k-th parent node $X_{\mathrm{pa}(i)}^{(k)}$. The probabilistic transition model from the previous state to the current state for all variables is given by

$$p(Z_t|Z_{t-1})=\prod_{i=1}^{N} p\left(X_{i,t}|X_{\mathrm{pa}(i),t}\right) \tag{1}$$

where $Z_t=\left[X_{1,t},X_{2,t},\ldots,X_{N,t}\right]$ and $X_{\mathrm{pa}(i),t}$ are the parent nodes of $X_{i,t}$ at time t. Thus, the joint probability density function of a DBN from sampling time $t=1$ to T is expressed as

$$p(Z_{1:T})=\prod_{t=1}^{T}\prod_{i=1}^{N} p\left(X_{i,t}|X_{\mathrm{pa}(i),t}\right) \tag{2}$$

After the network structure is determined, the set of model parameters Θ corresponding to the conditional probability density functions of all nodes need to be identified. The goal of network learning is to estimate each conditional probability density function corresponding to every node that maximizes the likelihood of the training data. It can be readily observed that each node is conditionally independent of its

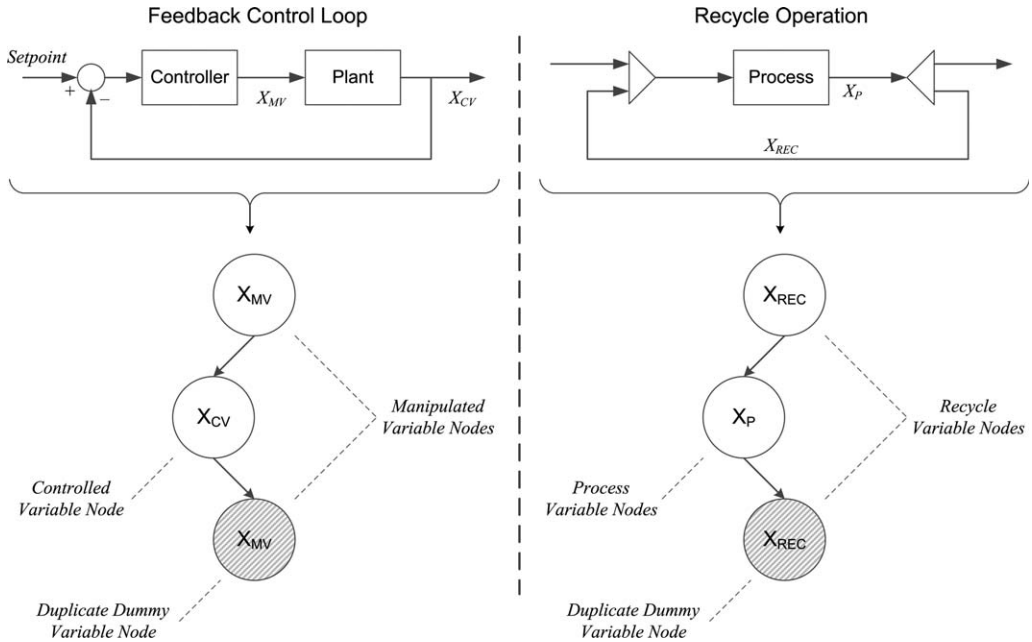

Figure 2. Illustrative diagram on the treatment of feedback controller and recycle operation using duplicate dummy nodes.

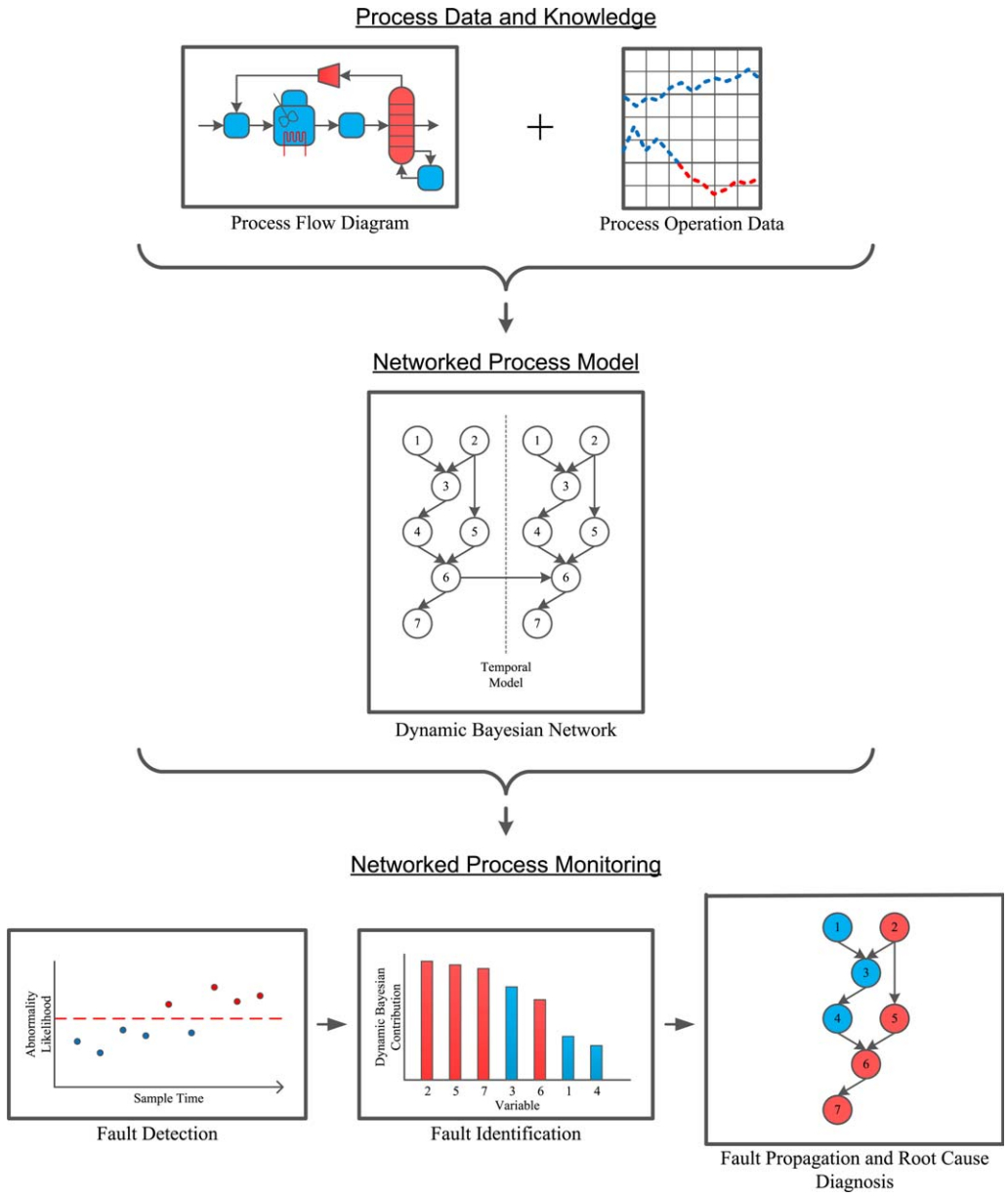

Figure 3. Illustrative diagram of the networked process monitoring framework.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

nondescendants given the parent nodes and the joint probability density function of all the network nodes is used to develop the log-likelihood function for parameter estimations. The log-likelihood-based objective function J of the model parameters Θ is expressed as

$$J(\Theta)=\{\log P(Y|\Theta)\}=\left\{\sum_{i=1}^{N}\log P\left(y_{i,1:T}|\Theta\right)\right\} \quad (3)$$

where $y_{i,1:T}$ is the sequence of measurement values of the node X_i from sampling time $t=1$ until T and $P\left(y_{i,1:T}|\Theta\right)$ is the likelihood function of process observations. The above log-likelihood function can be decomposed according to the identified network structure and then the contribution to the log-likelihood of each network node can be maximized independently based on the process measurement data. Thus, the optimal DBN model parameters are estimated from the sequence of process observations by maximizing the log-likelihood-based objective function below

$$\hat{\Theta}=\arg\max_{\Theta} J(\Theta)=\arg\max_{\Theta}\left\{\sum_{i=1}^{N}\log P\left(y_{i,1:T}|\Theta\right)\right\} \quad (4)$$

The detailed DBN model learning procedure can be found in literature.[41] The dynamic time-series data are used to estimate the time-varying conditional probabilities of all network nodes across different sampling instants. After the dynamic Bayesian network is constructed and the network model parameters are estimated from training data, it can be used for probabilistic inference with the new test data.

Networked Process Monitoring Approach for Fault Detection, Propagation Identification, and Root Cause Diagnosis

DBN model is an excellent tool to characterize nonsteady-state processes with stochastic uncertainty using conditional probability-based dynamic state transitions. Hence, it can be

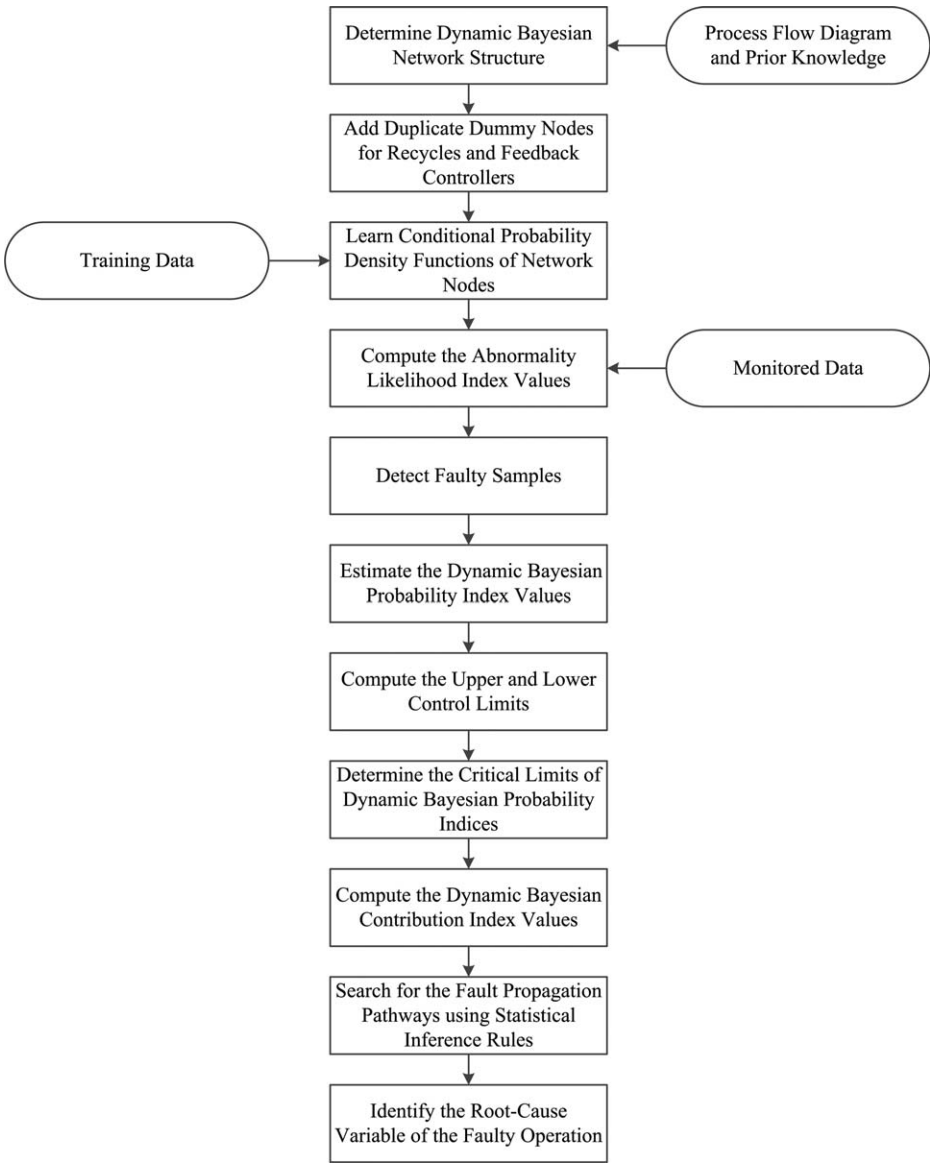

Figure 4. Schematic diagram of the networked process monitoring, fault propagation identification, and root cause diagnosis approach.

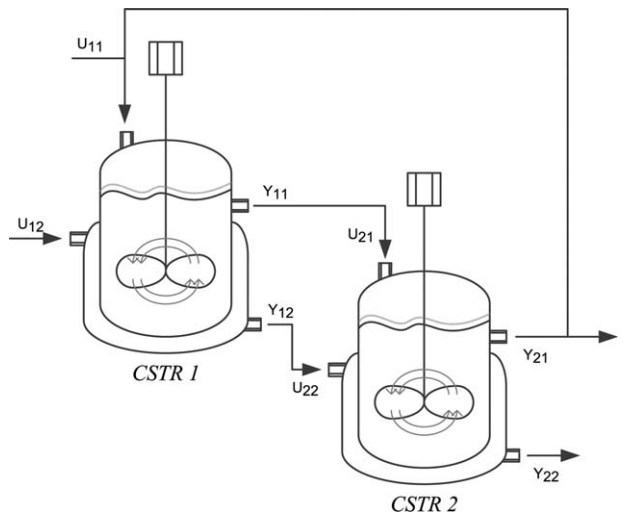

Figure 5. Process flow diagram of the cascaded continuous stirred tank reactor system.

adopted to identify the propagating probabilities among different measurement variables so as to determine the operating status of processes and diagnose the root causes of abnormal events. In this section, a novel networked process monitoring approach is developed to detect abnormal operation events, identify fault propagation pathways throughout the process, and diagnose the root-cause variables of process abnormalities.

The first task in the networked monitoring approach is to determine the qualitative network structure from the prior process knowledge. The process flow diagrams along with the preliminary process understanding are typically used for sorting out the monitored variables in process flow order from upstream to downstream process units. The process flow diagrams are analyzed to determine how all the measurement variables interact with each other in the unit operations. Such prior process knowledge is used to determine if any physical or chemical interactions exist between the adjoining variables. In this way, the causal relationships between process variables can be identified so that the arcs

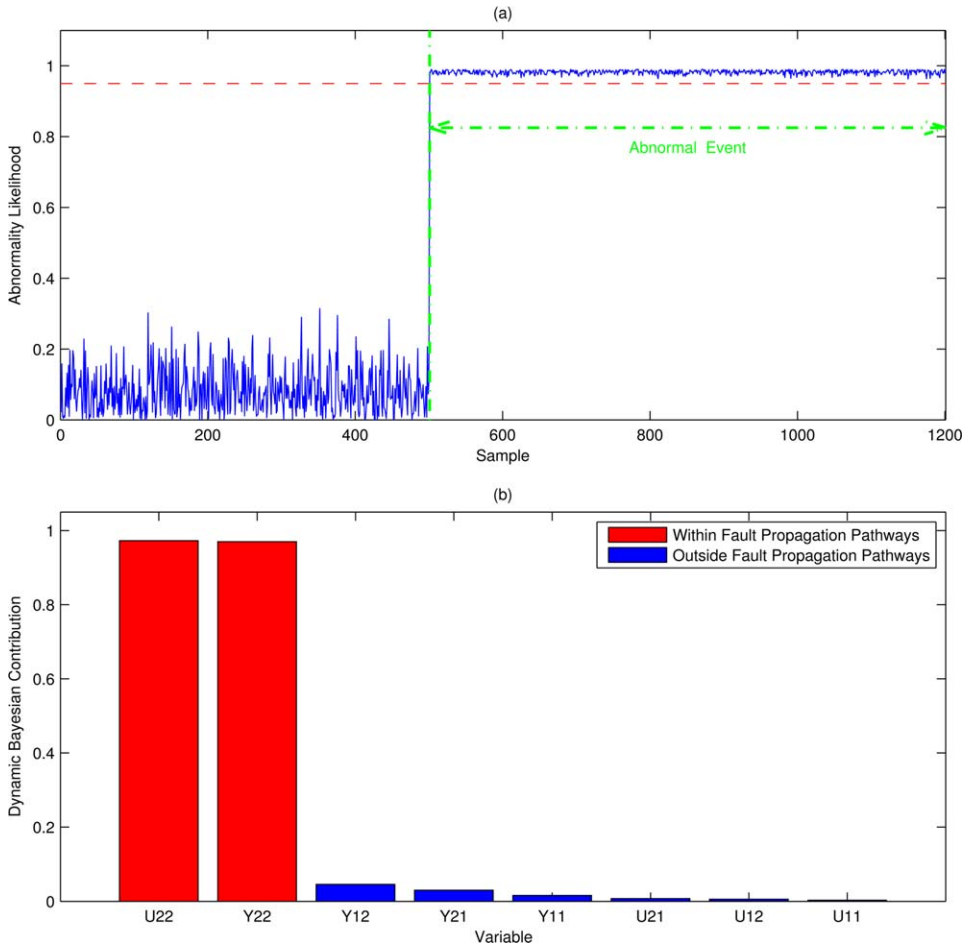

Figure 6. (a) Fault detection and (b) fault identification results of the proposed method in the cascaded continuous stirred tank reactor system.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

are used to connect the parent and child nodes based on such causal dependencies. For chemical processes with recycles and feedback controllers, a special treatment is needed as the Bayesian network is essentially acyclic graph. In this work, duplicate dummy variable nodes are designed within the network structure to denote the recycled variables or manipulated variables in control loops so that the feedback effects of the recycles and controllers can be accounted for. As illustrated in Figure 2, the manipulated and controlled variables of a control loop, X_{MV} and X_{CV}, may have the causal relationship from X_{MV} to X_{CV} due to the open-loop process interaction. Meanwhile, the close-loop feedback effect of the controller can be back-propagated from X_{CV} to X_{MV}. Thus, a duplicate dummy node X_{MV} is added to the network structure right after the node X_{CV} to capture such feedback influence. Likewise, a duplicate dummy variable node X_{REC} can be inserted into the network right after the process measurement variables along the recycle to form an inner loop that represents such entire recycle operation. It should be noted that the duplicate dummy nodes are needed in the network only if the particular recycled or manipulated variables are also monitored variables. Meanwhile, the correctly identified network structure is necessary to ensure that the fault propagation and root cause diagnosis results are unbiased and reliable.

With the network structure determined, the optimal DBN model parameters can then be estimated from the sequence of process observations. After the DBN model is obtained, the external evidence from new operating data can be used to update the transition probabilities of network nodes for process monitoring and fault propagation diagnosis. The likelihood index for the observations of the network nodes Z_t at sampling time t can be defined as follows

$$P(Z_t)=\prod_{i=1}^{N} P\left(X_{i,t}|X_{\mathrm{pa}(i),t}\right) \tag{5}$$

which represents the overall probability of all the network nodes operating under normal conditions. Further, the following ALI is proposed to quantify the probability of the monitored sample to be faulty

$$\zeta(t)=1-P(Z_t)=1-\prod_{i=1}^{N} P\left(X_{i,t}|X_{\mathrm{pa}(i),t}\right) \tag{6}$$

The control limit of the ALI ζ can be set to the prespecified confidence level $(1-\alpha)$ 100%, and the process fault is detected if the above index value exceeds the control limit.

To further identify the fault propagation pathways and diagnose root-cause variables, a DBPI is developed from the conditional probability of the i-th node X_i at sampling time t as follows

$$\gamma_i(t)=1-P\left(X_{i,t}|X_{\mathrm{pa}(i),t}\right) \tag{7}$$

The above time-varying transition probability index is of dynamic randomness due to the influence of the parent nodes

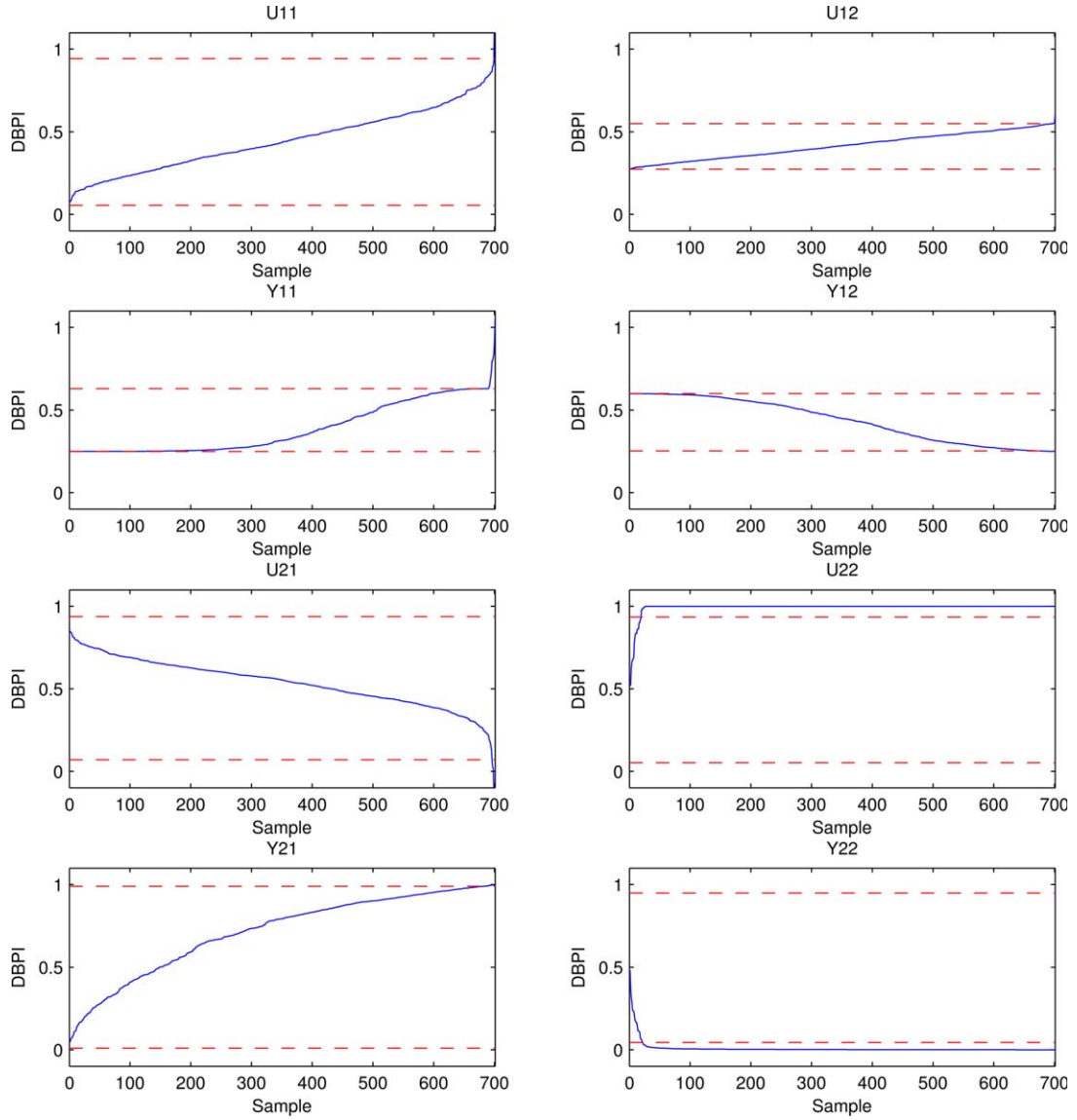

Figure 7. Trend plots of the DBPI values of the monitored variables for the cascaded continuous stirred tank reactor system.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

on the child node and can be used to quantify the effect of each monitored variable by its directly connected upstream variables in a probabilistic fashion. Then, the upper and lower control limits of DBPI, denoted as γ_i^{UCL} and γ_i^{LCL}, can be estimated through kernel density estimation under the confidence levels $(1-\alpha) \times 100\%$ and $\alpha \times 100\%$, respectively. The kernel density estimator is defined as

$$\hat{f}_h(\gamma_i) = \frac{1}{T \cdot h} \sum_{t=1}^{T} K\left(\frac{\gamma_i - \gamma_i(t)}{h}\right) \tag{8}$$

where h is the kernel bandwidth and K is the Gaussian kernel function expressed as

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}u^2\right)} \tag{9}$$

with u representing an arbitrary data point.[45] A diffusion-based plug-in method is used to obtain the optimal bandwidth by minimizing the mean integrated squared error of the estimated density.[46] With the zero boundary conditions on the first-order derivative of density function, the estimated control limits γ_i^{UCL} and γ_i^{LCL} are ensured to be within the range of [0, 1]. Further, either the upper or the lower control limit is selected as the critical limit γ_i^C depending on the following rules

$$\gamma_i^C = \begin{cases} \gamma_i^{\mathrm{LCL}} & \text{if} \quad \gamma_i(t) < \gamma_i^{\mathrm{LCL}} \\ \gamma_i^{\mathrm{UCL}} & \text{if} \quad \gamma_i(t) > \gamma_i^{\mathrm{UCL}} \end{cases} \tag{10}$$

The number of faulty samples that exceed the critical limit γ_i^C is counted and normalized as the proposed DBCI Γ_i for each variable as

$$\Gamma_i = \frac{1}{T} \sum_{t=1}^{T} \xi_{i,t} \tag{11}$$

where $\xi_{i,t}$ is a binary value corresponding to the node X_i at sampling time t given by

$$\xi_{i,t} = \begin{cases} 1 & \text{if} \quad |\gamma_i(t)| \geq |\gamma_i^C| \\ 0 & \text{if} \quad |\gamma_i(t)| < |\gamma_i^C| \end{cases} \tag{12}$$

The above DBCI represents the likelihood of a process variable with significantly abnormal behavior and can serve

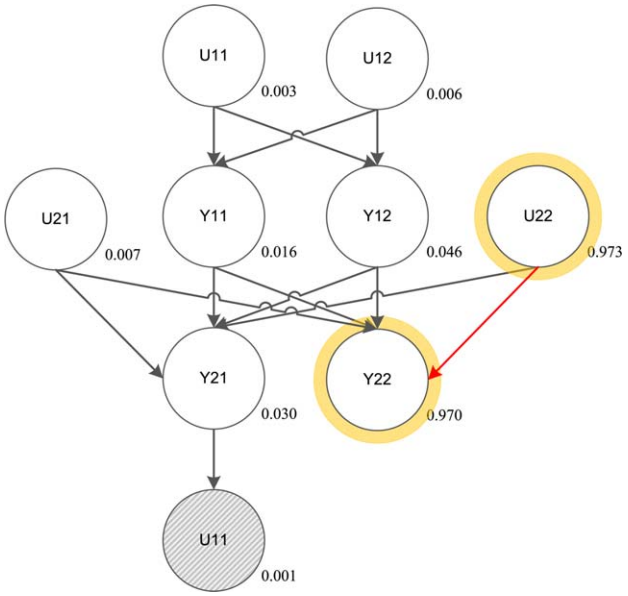

Figure 8. Networked process model and fault propagation diagnosis results for the cascaded continuous stirred tank reactor system.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

as an indicator pointing to the major cause or effect variables due to process faults.

As the faults can often propagate throughout the process due to complicated variable interactions in a dynamic manner, it would be useful to identify the fault propagation pathways within the monitored process network and then the root causes of abnormal process operation can also be diagnosed. Following the above networked process monitoring framework, a probabilistic inference strategy is developed to determine the fault propagation pathways.

The proposed fault propagation pathway search starts from the bottom leaf nodes of the network that do not have any child nodes. First, a threshold value ε is set at the confidence level $(1-\alpha)$ 100%. Among all the leaf nodes of the network, the ones satisfying the following criterion are then selected as the starting nodes in the fault propagation pathway $S=\{\eta_1, \eta_2, \ldots, \eta_M\}$

$$\eta_1^* = \arg_{\eta_1 \in Z_L} \{\Gamma_{\eta_1} \geq \epsilon\} \tag{13}$$

where η_1^* denotes the selected node at which the fault propagation pathway starts, Z_L represents the complete set of leaf nodes in the network, Γ_{η_1} is the DBCI value of the leaf node η_1 and M is the total number of identified nodes within the fault propagation pathway. Then, the next node η_j in the propagation pathway sequence is determined by inferring from the previous node η_{j-1} along the reversed direction of the network. The statistical decision rules are designed such that the parent nodes of η_{j-1} with the DBCI values no less than the threshold value are first selected in the fault propagation pathway

$$\eta_j^* = \arg_{\eta_j \in X_{\mathrm{pa}(j-1)}} \left\{\Gamma_{\eta_j} \geq \epsilon\right\} \tag{14}$$

where η_j^* corresponds to the j-th node in the propagation pathway sequence, $X_{\mathrm{pa}(j-1)}$ represent all the parent nodes of the previous node η_{j-1} in the pathway, and Γ_{η_j} is the DBCI value of the node η_j. If the DBCI values of all the parent nodes are below the threshold, then one of the parent nodes with the largest DBCI value should be included in the pathway as follows

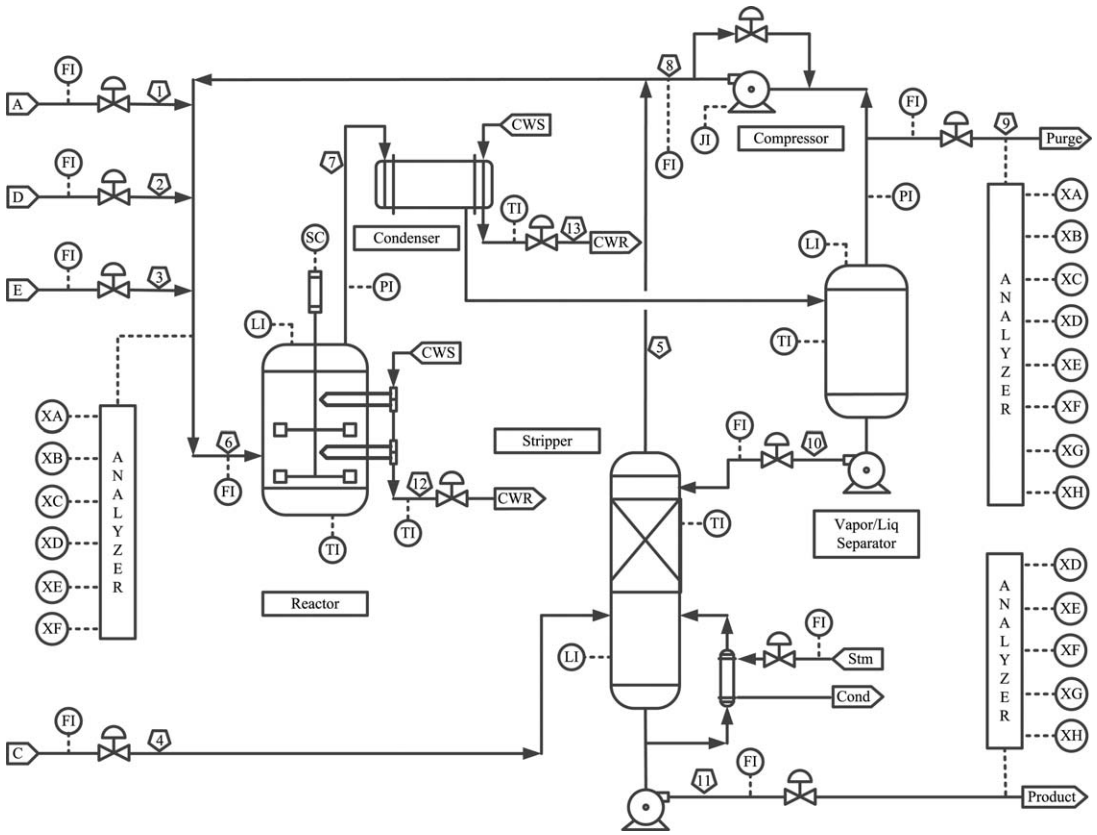

Figure 9. Process flow diagram of the Tennessee Eastman chemical process.

Table 1. Monitored Variables of Tennessee Eastman Chemical Process

Variable No.	Variable Symbol	Variable Description
1	F1	A feed
2	F2	D feed
3	F3	E feed
4	F4	A and C feed
5	F5	Recycle flow
6	F6	Reactor feed rate
7	P7	Reactor pressure
8	L8	Reactor level
9	T9	Reactor temperature
10	F10	Purge rate
11	T11	Separator temperature
12	L12	Separator level
13	P13	Separator pressure
14	F14	Separator underflow
15	L15	Stripper level
16	P16	Stripper pressure
17	F17	Stripper underflow
18	T18	Stripper temperature
19	F19	Stripper steam flow
20	J20	Compressor work
21	T21	Reactor cooling water outlet temperature
22	T22	Condenser cooling water outlet temperature

Table 2. Test Scenarios of Tennessee Eastman Chemical Process

Case No.	Test Scenario	Duration
Case 1	Normal operation	1st–300th samples
	A feed loss. Switch pressure controller to purge stream and reduce production rate by 23.8%.	301st–550th samples
Case 2	Normal operation	1st–300th samples
	Increased random variations in reactor pressure	301st–540th samples

$$\eta_j^* = \underset{\eta_j \in X_{\mathrm{pa}(j-1)}}{\arg\max}\ \Gamma_{\eta_j} \tag{15}$$

The search of fault propagation pathway continues until all the network nodes whose DBCI values are not less than the threshold have been included in the pathway as

$$\{X_i | \Gamma_{X_i} \geq \epsilon, 1 \leq i \leq N\} \in S = \{\eta_1, \eta_2, \ldots, \eta_M\} \tag{16}$$

Otherwise, the pathway search will not terminate until it reaches one of the root nodes that does not have any more parent nodes. Within the identified fault propagation pathway, the ending node represents the root-cause variable resulting in the abnormal operation event.

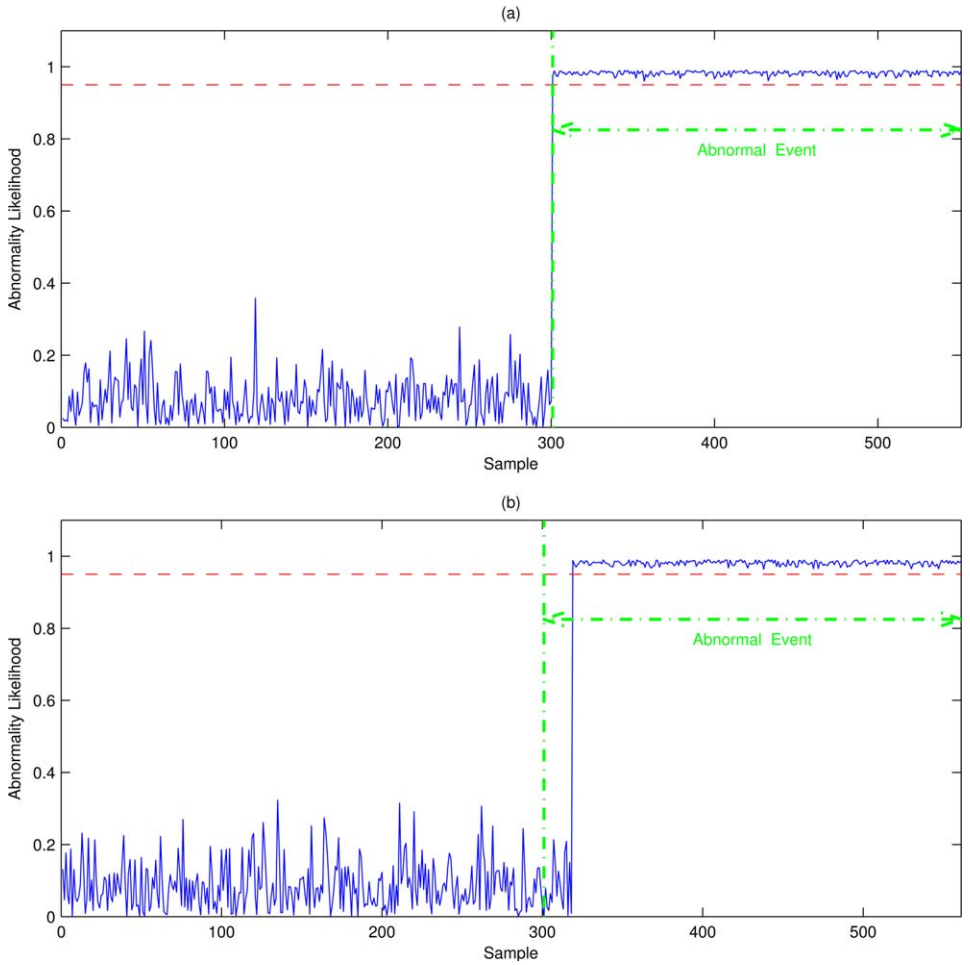

Figure 10. Fault detection results of the proposed method in (a) the first and (b) the second test cases of the Tennessee Eastman chemical process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

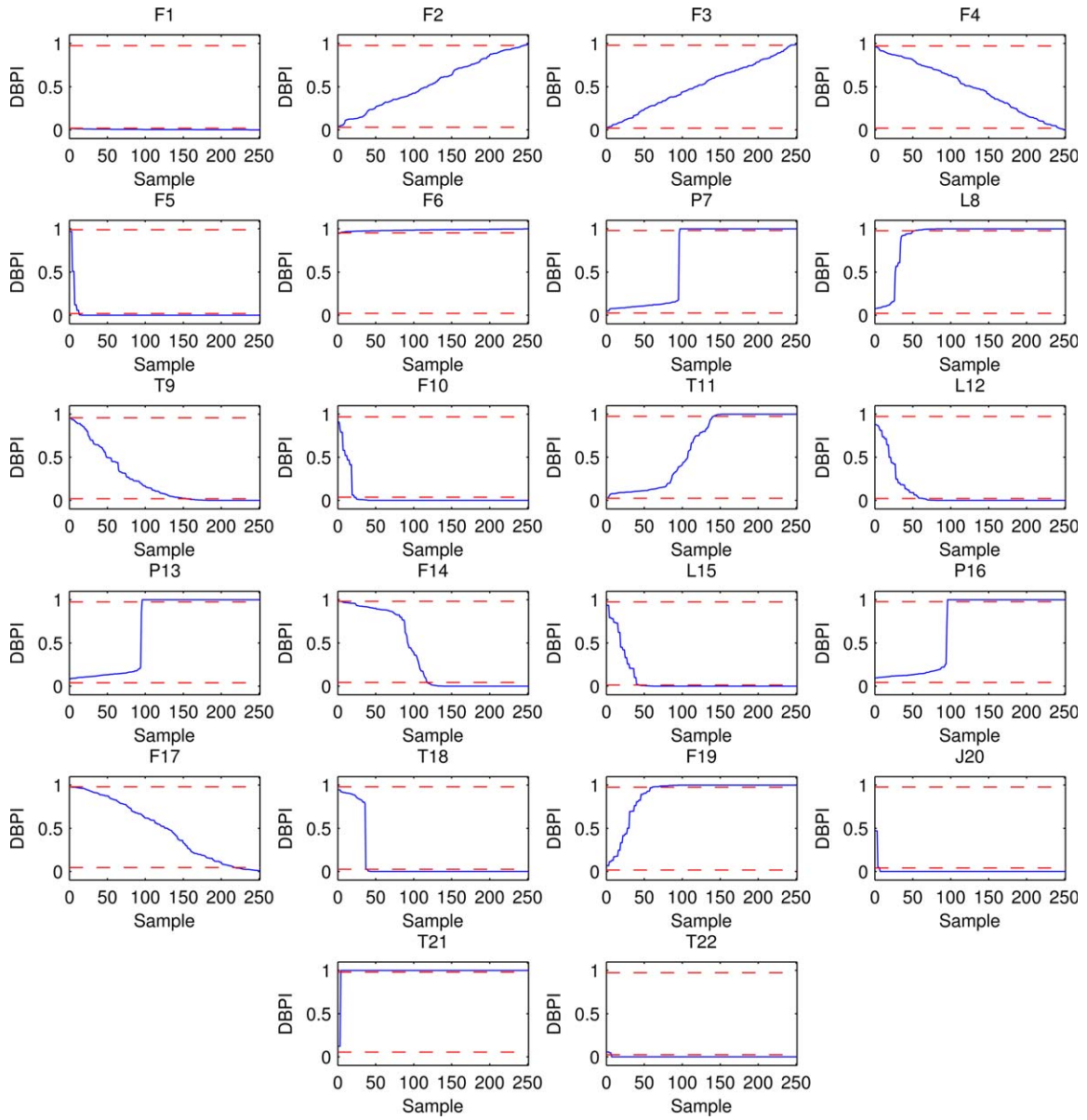

Figure 11. Trend plots of the DBPI values of the monitored variables in the first test case of the Tennessee Eastman chemical process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

The entire networked process monitoring framework is illustrated in Figure 3. Moreover, the step-by-step procedure of the proposed monitoring approach is given below and the corresponding flow diagram is shown in Figure 4.

1. Determine the dynamic Bayesian network structure based on prior process knowledge and process flow diagram with the network nodes denoting the monitored process variables and the arcs representing the variable interactions;
2. Add duplicate dummy nodes for the recycled or manipulated variables in the recycle operation or feedback controllers if those variables are monitored variables;
3. Learn the conditional probability density functions of variable nodes in the network model from process historical data;
4. Compute the ALI values of the monitored data and determine abnormal operation for the samples whose ALI values are above the prespecified confidence level;
5. Estimate the DBPI values of the detected faulty samples for all the network nodes;
6. Compute the upper and lower control limits of the dynamic Bayesian probability indices for different nodes using kernel density estimation;
7. Determine the critical limits of the dynamic Bayesian probability indices of various nodes;
8. Further compute the DBCI values of different variable nodes in the network and produce the corresponding contribution plots;
9. Search for the fault propagation pathways according to the developed statistical inference rules;
10. Terminate the pathway search based on the stopping criteria and identify the ending node in the pathway as the root-cause variable for the abnormal operation.

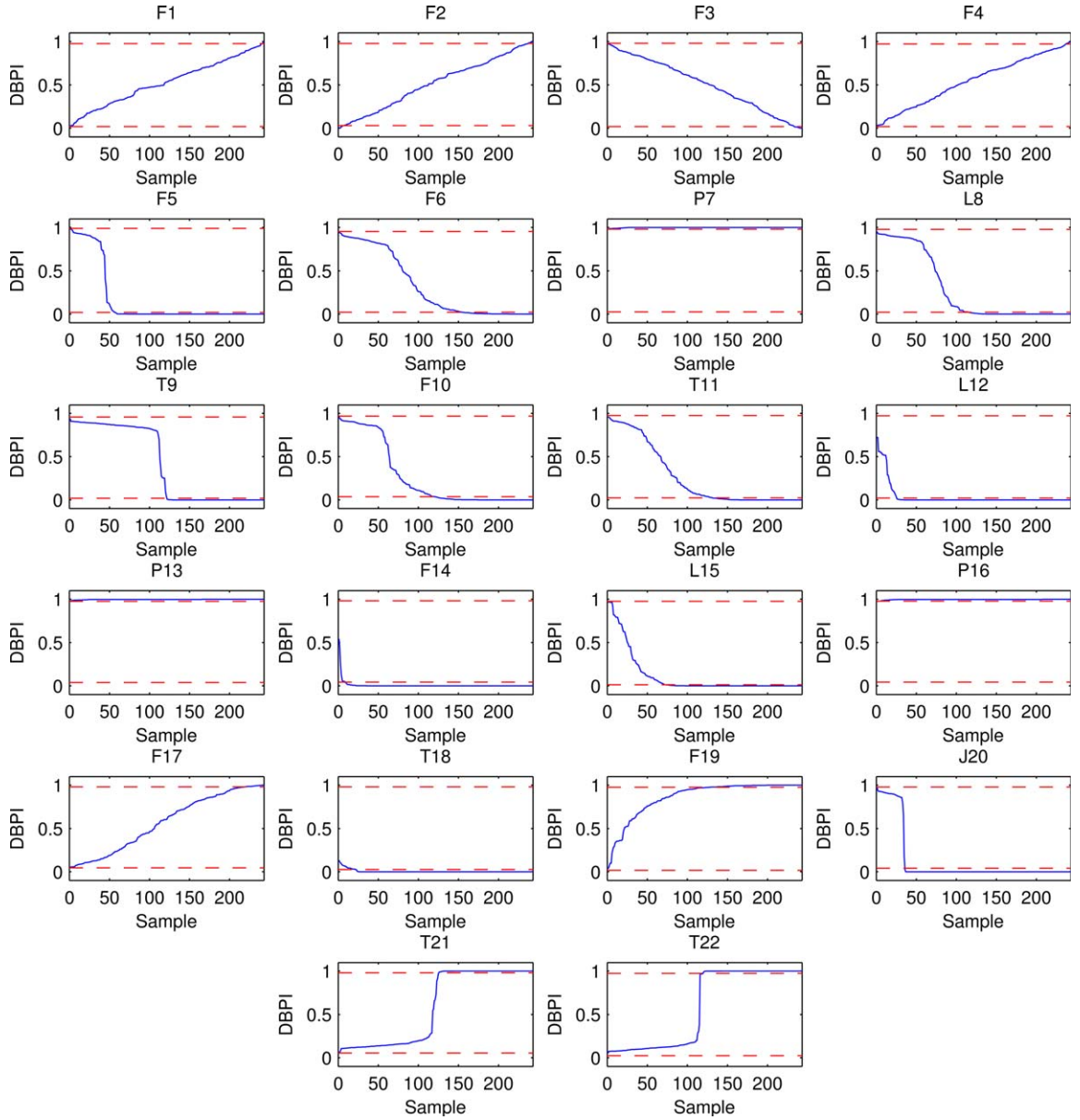

Figure 12. Trend plots of the DBPI values of the monitored variables in the second test case of the Tennessee Eastman chemical process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

Case Studies

An illustrative example

A simulated process of two continuous stirred-tank reactors (CSTR) connected in series is used to illustrate the usage of the proposed networked process monitoring and fault propagation diagnosis approach. A schematic illustrating the process setup is shown in Figure 5. A first-order exothermic reaction

$$A \xrightarrow{k_A} B \tag{17}$$

takes place within the reactors where k_A is the reaction rate with an Arrhenius temperature dependence. The reactors are fed with continuous inlet streams at the flow rates of U_{11} and U_{21}, respectively. In addition, the constant coolant flow rates of U_{12} and U_{22} are used to maintain the desired temperature profile for the reaction. A recycle stream carries part of the product of the second CSTR back to the first one. The monitored variables include the input flow rates (U_{11} and U_{21}), the output concentrations (Y_{11} and Y_{21}), the inlet coolant flow rates (U_{12} and U_{22}), and the outlet coolant temperatures (Y_{12} and Y_{22}). A test scenario is designed whereby the cooling jacket input to the second reactor is affected by increased flow variations starting from 501st sample until the 1200th sample.

The monitored variables in terms of process flow order can be first sorted according to the process flow diagram to form the network hierarchy. Then, the prior process knowledge regarding the potential physical/chemical interactions along those monitored variables can be used to determine the node connections and network structure. In this example, it is evident that the variables U_{11} and U_{12} that represent the inputs to the first reactor are the parent variables for the child variables Y_{11} and Y_{12} that represent the output variables of the first reactor. Similarly, the inputs to the second reactor denoted by U_{21} and U_{22} and the outputs of the first reactor

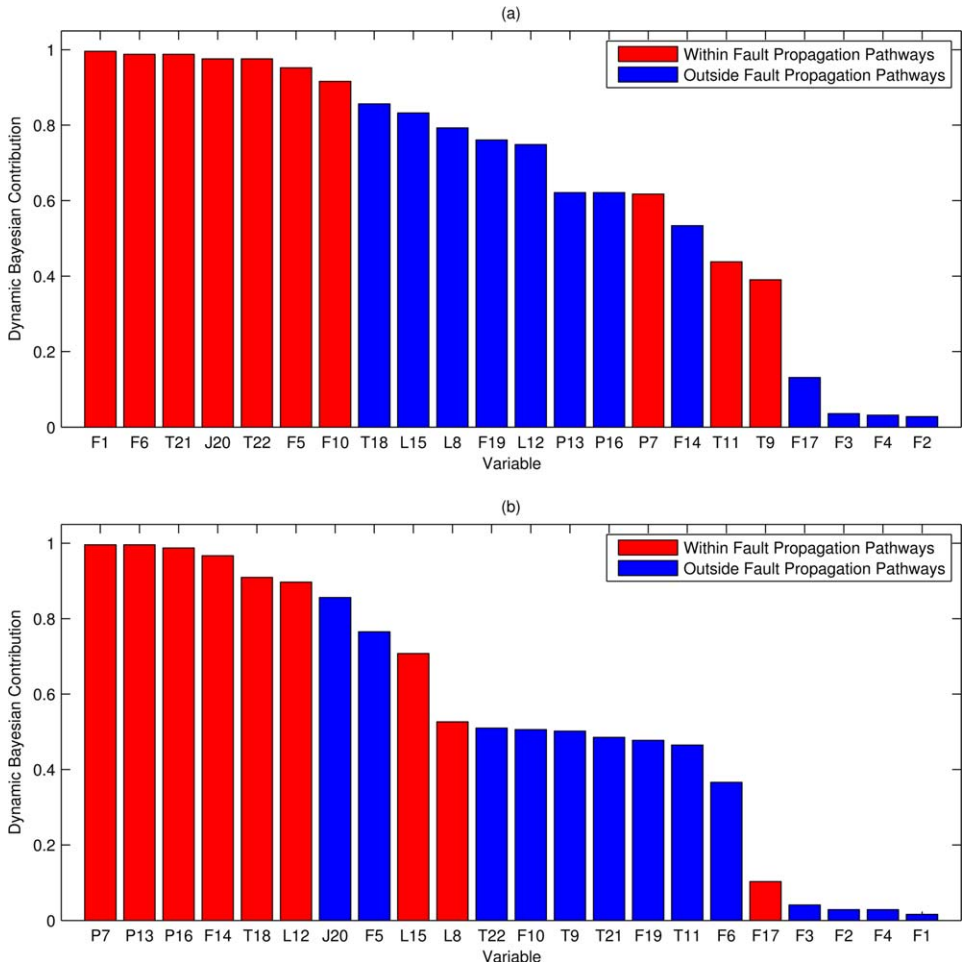

Figure 13. Dynamic Bayesian contribution plots in (a) the first and (b) the second test cases of the Tennessee Eastman chemical process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

Y_{11} and Y_{12} affect the outputs of the second reactor represented by Y_{21} and Y_{22}. Due to the existing recycle stream, the duplicate dummy variable node to represent the recycled variable of inlet flow of the first CSTR should be added to the network, as shown in Figure 8. The nodes in the network structure represent the monitored variable, whereas the conditional probabilities comprise the model parameters with the inference of casual dependency among the variables. Here, the conditional density functions of all nodes can be estimated from the normal training data. Then, the ALI is computed from Eq. 6 for process fault detection and the results are shown in Figure 6a. The confidence level used in this research is 95%. The DBPI values along with the corresponding upper and lower limits are further estimated for the detected faulty samples, as shown in Figure 7. With the DBPI values obtained, the DBCI values can, thus, be derived from Eqs. 10 and 11. The DBCI-based contribution plot is shown in Figure 6b, and it can be combined with the network structure and statistical inference rules for fault propagation diagnosis. It is noted that the numerical value marked next to each node in the process network model (Figure 8) corresponds to the DBCI value of the monitored variable. The fault propagation search starts from the leaf nodes, and the variable Y_{22} is selected as the starting node in the pathway because its corresponding DBCI value is above the threshold of 0.95. Then among the four parent nodes of Y_{22}, only the variable U_{22} has its corresponding DBCI value above 0.95 and, thus, identified as the second node in the fault propagation pathway. Because all the network nodes whose DBCI values are not less than the threshold have been included in the pathway, the search is terminated and the root node U_{22} is diagnosed as the root-cause variable of faulty operation. Because the process abnormality occurs on the cooling water flow into the second CSTR, the root-cause variable is identified accurately. Meanwhile, the abnormal variations on the coolant flow can directly affect the outlet coolant temperature of the second reactor and this explains why the process fault has effect on the variable Y_{22} so that it is also included in the fault propagation pathway.

Tennessee Eastman chemical process example

In this subsection, the Tennessee Eastman chemical process is further used to evaluate the effectiveness of the networked process monitoring, fault propagation identification, and root cause diagnosis approach. The process consists of five major units, which include an exothermic two-phase reactor, a flash separator, a recycle compressor, a reboiled stripper, and a product condenser. The process flow diagram is shown in Figure 9. The process has total 41 measurement variables and 12 manipulated variables.[47] A decentralized control strategy is adopted so that the process is closed-loop stable.[48] There are 22 continuous process variables among

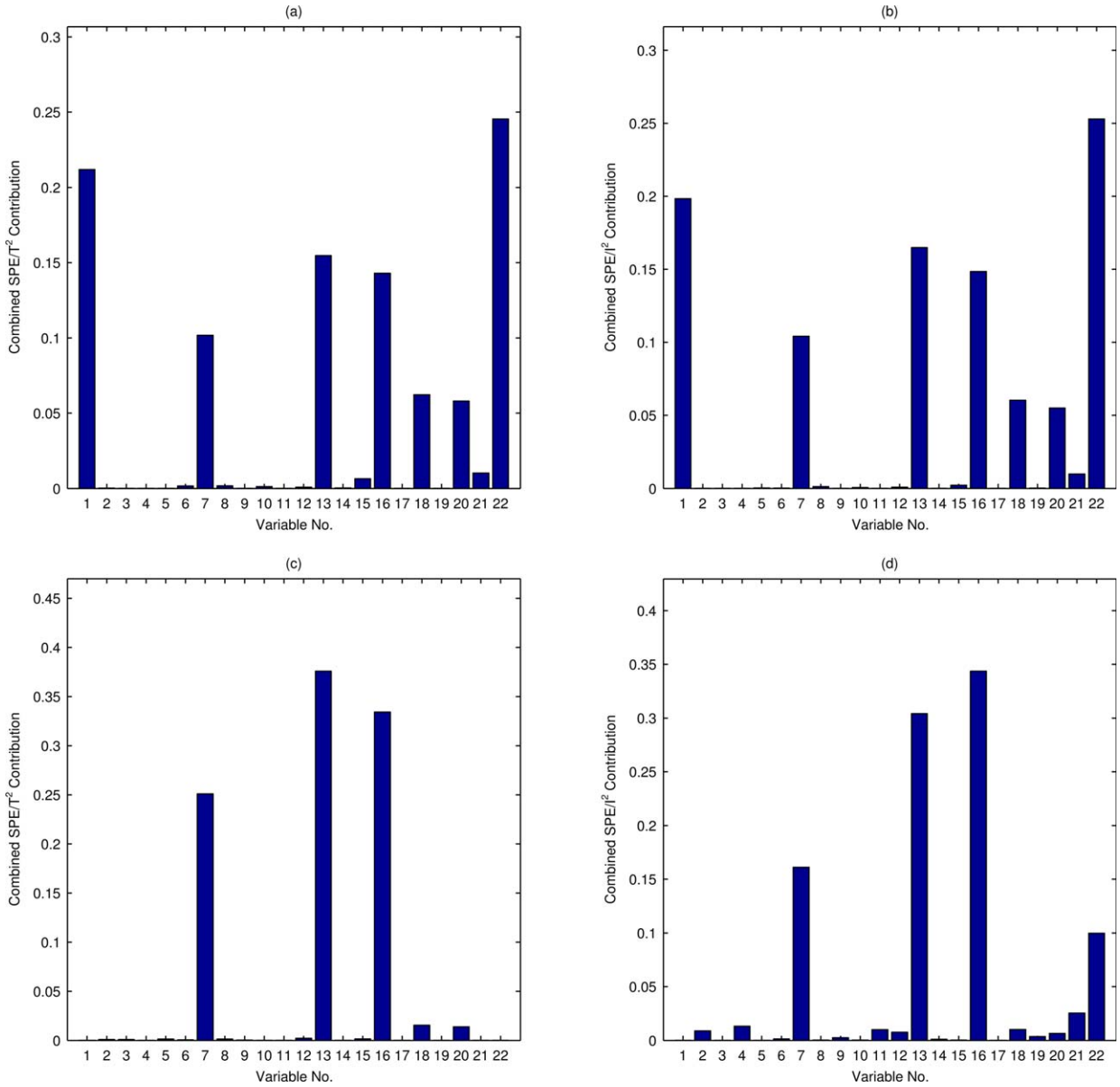

Figure 14. Variable contribution plots of the Tennessee Eastman chemical process (a) PCA-based contribution in the first test case, (b) ICA-based contribution in the second test case, (c) PCA-based contribution in the second test case, and (d) ICA-based contribution in the second test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

the 41 measurement variables as listed in Table 1 that are used to monitor the process operation. The variable symbols as shown in the table are used in this subsection. Moreover, process measurements are sampled with an interval of 3 min. A training data set consisting of 500 samples collected under normal operating conditions is used to develop the DBN model. In addition, two test cases as summarized in Table 2 are designed to examine the performance of the proposed monitoring and diagnosis method. The first test case begins with normal operation for the initial 15 h and then is followed by the process fault of a feed loss in the A stream for the remaining 12.5 h. In the second test scenario, the process starts at normal operating conditions and lasts for 15 h before an abnormal event of increased random variations in reactor pressure occurs with the duration of 12 h.

The first step is to design the network structure based on the prior process knowledge and process flow diagram. Basically, the monitored variables can be sorted in process flow order and then formulated into network hierarchy without node connections. Then, the process knowledge is used to analyze the potential interactions among variables along the network hierarchy and determine the node connections. The designed network consist of 22 nodes corresponding to all the monitored variables and one duplicate dummy node for the recycle flow. It should be noted that no duplicate dummy nodes are added for feedback controllers because those manipulated variables are not included as monitored variables. The network model parameters representing the conditional probability density functions of all the nodes are estimated from the training data. Thus, the ALI values of the test samples are computed to detect the occurrence of process faults. Within the identified faulty periods, the fault propagation pathway identification and root cause diagnosis are further carried out.

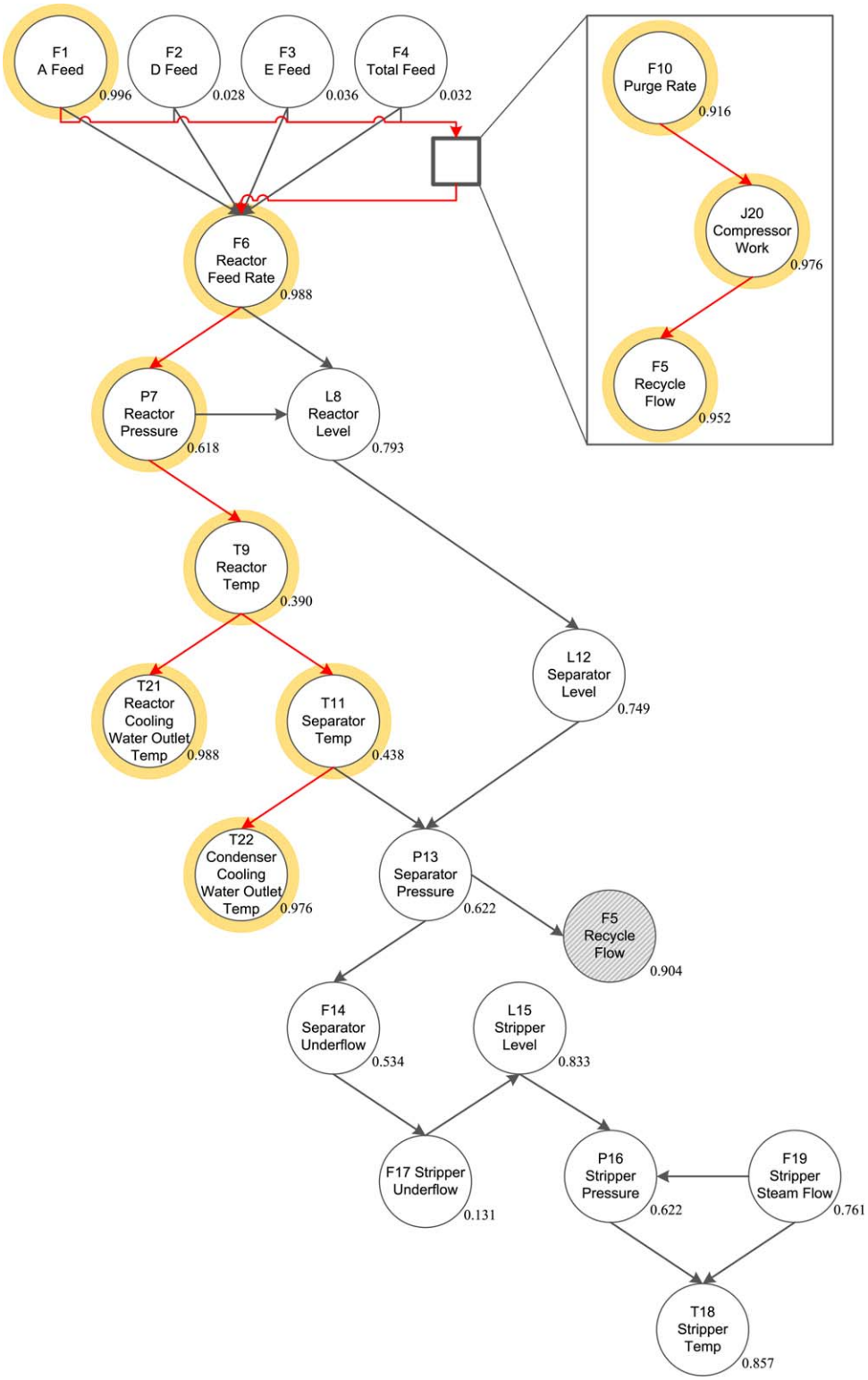

Figure 15. Networked process model and fault propagation diagnosis results in the first test case of the Tennessee Eastman chemical process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

For the first test case, the trend plot of the ALI is shown in Figure 10a. It is readily seen that the ALI value remains below the control limit line for the first 300 samples from normal operating conditions and then jumps above the limit once the process fault of A feed loss takes place. The ALI gives accurate alarm on process fault with very minimal delay. After the abnormal event is captured, the conditional probability density functions of the network nodes are used to estimate the DBPI values of different nodes during the identified faulty period, which along with the estimated upper and lower control limits are shown in Figure 11. Based on the DBPI values, the DBCI values are further

Table 3. Transfer Entropy-Based Causality Matrix in the First Test Case of Tennessee Eastman Chemical Process

		Variable No.																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Variable No.	1	0	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1
	6	0	0	0	0	0	0	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1
	7	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1
	8	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1
	9	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	0	1	1	1	1	1
	10	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	1
	11	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1
	12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	1	1
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

computed for all the variable nodes and the corresponding DBCI-based contribution plot is depicted in Figure 13a. It is obvious that the three leading variables with the largest contribution values are A feed rate (F1), reactor feed rate (F6), and reactor cooling water temperature (T21). Because the process abnormality occurs in A feed rate, the reactor feed rate and the reactor coolant temperature are consequently affected due to the variable interactions. The dynamic Bayesian network-based contribution plot can precisely identify the major variables with the significant abnormal behavior. As a comparison, the PCA and independent component analysis (ICA)-based variable contribution plots are shown in Figures 14a, b, respectively. Here, the squared prediction error (SPE)/T^2 combined and SPE/I^2 combined contribution indices are used in the PCA and ICA-based fault diagnosis methods, respectively. One can easily see that the conventional contribution plots are unable to correctly point to the faulty variables. As shown in Figure 15, the fault propagation pathway is identified with the DBCI values of different nodes and statistical inference rules. The fault propagation sequence includes A feed rate (F1), purge rate (F10), compressor work (J20), recycle flow (F5), reactor feed rate (F6), reactor pressure (P7), reactor temperature (T9), reactor cooling water temperature (T21), separator temperature (T11), and condenser cooling water temperature (T22). As a result of the loss in A feed rate, the reactor inlet stream is affected so that the abnormal behavior can be observed in reactor feed rate. Subsequently, the reactor pressure and cooling water temperature deviate from the normal values. Furthermore, the fault can cause the upset in the recycle stream and lead to abnormalities on purge rate, compressor work, and recycle flow. Meanwhile, the downstream condenser and separator temperatures are affected by the reactor temperature and recycle stream. Therefore, the identified pathway correctly reflects the actual fault propagations throughout the process and the ending node of A feed rate in the identified pathway is accurately captured as the root-cause variable of process fault. It needs to be noted that some variables with small dynamic Bayesian contributions may also be included in the propagation pathway because they are the intermediate variables along the process flow even though they may not exhibit significant abnormal responses. For industrial applications, it should be pointed out that the red-highlighted fault propagation pathway in Figure 15 can be extracted and presented separately in the user interface of monitoring system to operators and engineers so that they can easily see the root-cause variable as well as the faulty effects propagated throughout the process.

As a comparison, the transfer entropy method is used to identify the directions, along which the fault propagates from upstream to downstream variables.[35,36] The monitoring results of the transfer entropy method are shown in Table 3. This method quantifies the transfer entropies between each pair of variables to identify all the potential propagation pathways based on measurement data. The causal relationships are expressed in a matrix with the rows representing the cause variables while the columns the effect variables. A unit entry in the table means that a causal relationship exists, whereas a zero indicates that there is no causal relationship between the two variables. It can be observed that the faulty effects are propagated to downstream operations and many process variables are affected by the abnormal operation in the process. Nevertheless, this method is data driven and does not take into account the process knowledge to determine process network structure. The most significant propagation pathway along with the root-cause variable is unable to be isolated. In contrast, the proposed networked process monitoring approach uses prior process knowledge to determine the process network structure and further identify the major propagation pathway and root-cause variable.

For the second test scenario, the ALI-based fault detection result is shown in Figure 10b. Though the alarm is triggered with a short delay, 93.1% of all the abnormal samples are correctly captured without misclassification. Moreover, there is no false alarm triggered on the normal samples. Within the faulty period, the DBPI values and the corresponding upper and lower control limits for all the monitored variables are shown in Figure 12. Using DBPI, the DBCI values of all

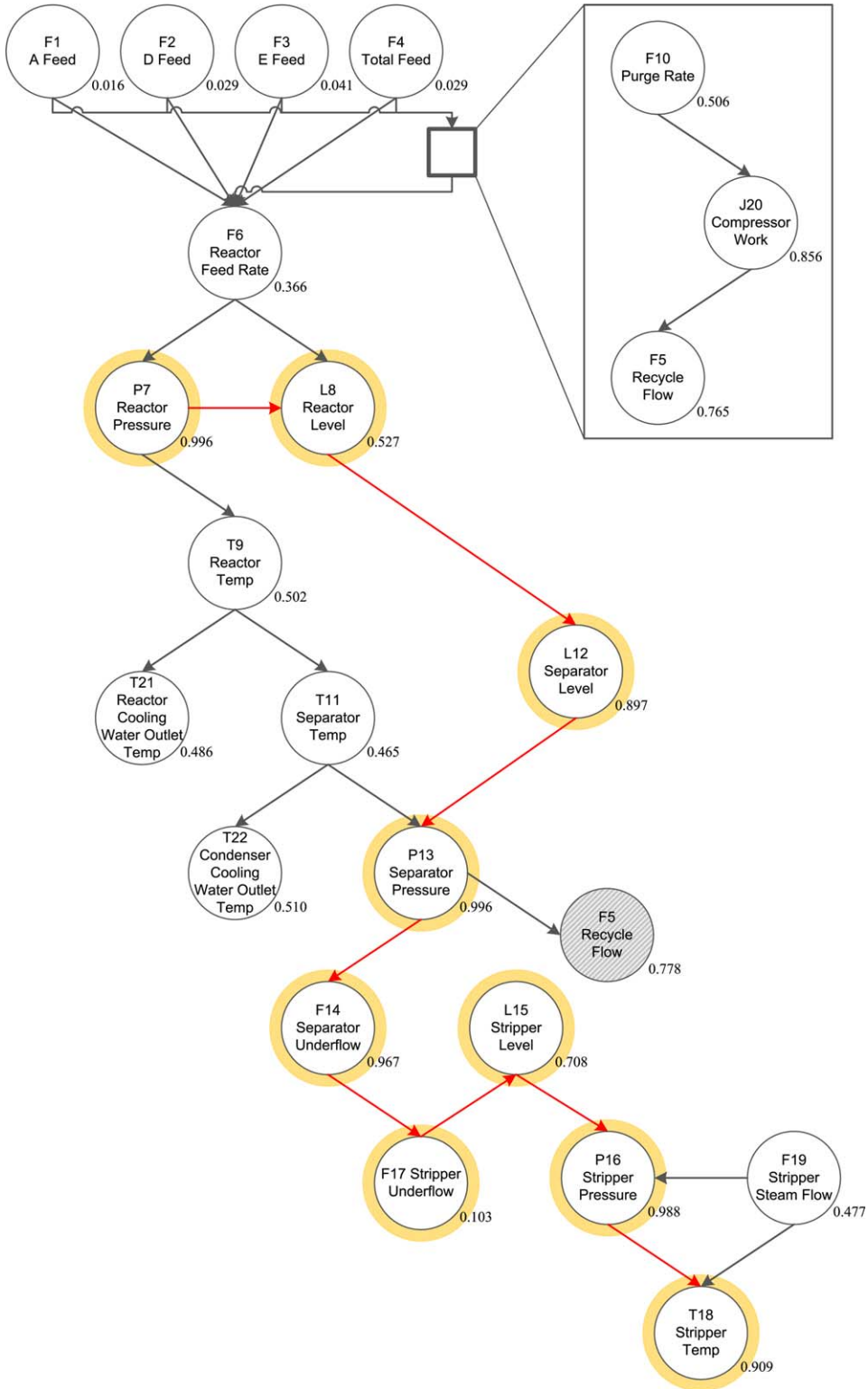

Figure 16. Networked process model and fault propagation diagnosis results in the second test case of the Tennessee Eastman chemical process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

the variables are further computed, as shown in Figure 13b. The leading variables with the highest contributions include reactor pressure (P7), separator pressure (P13), and stripper pressure (P16). As the fault occurs in the reactor pressure, the separator and stripper pressures in the downstream units are subsequently affected along the process flow. In contrast, the PCA and ICA-based contribution plots in Figures 14c, d do not well-isolate the major variables with the most significant upsets. For instance, the reactor pressure variable has smaller contribution than the separator and stripper pressure in the PCA-based contribution plot. Likewise, the stripper pressure is also the largest contributor from the ICA-based

Table 4. Transfer Entropy-Based Causality Matrix in the Second Test Case of Tennessee Eastman Chemical Process

		Variable No.																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Variable No.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	1	1	1	0	1	1	0	1	1	0	1	0	1	1	1
	8	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	1	0	1	0	1	1	1
	9	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	1	0	1	1	1
	10	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	0	1	1	1
	11	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	0	1	1	1
	12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	1	1	1
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	1
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	1
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

method. The fault propagation diagnosis results from the proposed approach are further shown in Figure 16. The abnormality in reactor pressure (P7) can affect the gas-phase reaction rate and result in increased variations of reactor level (L8). Then, the undesirable changes take place in the pressure (P13) and level (L12) of the downstream flash separator, which may cause the variations in separator underflow (F14). The abnormal behaviors may further propagate to the downstream stripper and have influence on its level (L15), pressure (P16), and temperature (T18). The identified fault propagation pathway coincide well the process analysis and the ending variable of reactor pressure is accurately located as the root cause of process fault.

Meanwhile, the transfer entropy method is also carried out to identify the fault propagations toward all the downstream variables. The results of the transfer entropy approach are shown in Table 4. It can be seen that the downstream variables related to the condenser, separator, and stripper are detected as being affected by the process fault in the reactor through different paths. However, the degree to which the faulty effect variables are affected by the abnormal event is not quantified as in the proposed networked process monitoring approach through the estimated DBCI values. Furthermore, it does not search for the fault propagation pathway within the constructed process networks. Therefore, the most significant fault propagation pathway and the root-cause variable are not identified by the transfer entropy method. Such comparison shows that the proposed networked process monitoring approach has the unique capability to trace the major fault propagation pathway and locate the root-cause variable.

Conclusions

In this study, a novel networked process monitoring, fault propagation identification, and root cause diagnosis approach is developed. Different from the conventional data-driven process monitoring techniques, the proposed method incorporates the prior process knowledge with plant historical data to design and estimate a temporal process network for monitoring and particularly fault propagation and root cause diagnosis. First, the new ALI is proposed from time-varying transition probabilities of network nodes for process fault detection. Once the fault is captured, statistical inference-based dynamic Bayesian probability and contribution indices are developed for fault propagation pathway identification and root-cause variable diagnosis. The proposed approach is applied to an illustrative continuous stirred tank reactor system and the Tennessee Eastman chemical process with the results compared against the transfer entropy-based monitoring method. The results demonstrate that the presented approach can effectively monitor process operation, identify fault propagation pathways, and locate root causes of process faults. It should be noted that the small unharmful disturbances will not trigger false alarms as they do not cause substantial upsets and propagations in the process. However, the harmful disturbances can be captured as abnormal operating event by the proposed method when the significant process upsets and fault propagations are caused.

The merits of the new networked monitoring approach lie in the facts that (1) it combines the advantages of both data-driven and knowledge-based monitoring techniques; (2) it monitors the plant operation in a networked fashion so that the variable interactions and causal relationships are well-captured; and (3) it can identify fault propagation pathway as well as diagnose root causes of process abnormality, which may not be easily achieved from routine monitoring methods. Future work may be focused on three aspects: (1) exploring the automated or semiautomated strategy to design the initial process network structure; (2) investigating the detection and diagnosis capability of the networked process monitoring approach on more subtle faults; and (3) extending the networked process monitoring and diagnosis approach to nonlinear batch/semibatch processes.

Literature Cited

1. Chiang LH, Russell EL, Braatz RD. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemom Intell Lab Syst*. 2000;50:243–252.

2. Hoo KA, Piovoso MJ, Schnelle PD, Rowan DA. Process and controller performance monitoring: overview with industrial applications. *Int J Adapt Control Signal Process*. 2003;17:635–662.
3. Yu J, Qin SJ. Variance component analysis based fault diagnosis of multilayer overlay lithography processes. *IIE Trans*. 2009;41: 764–775.
4. Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. A review of process fault detection and diagnosis: part I: quantitative model-based methods. *Comput Chem Eng*. 2003;27:293–311.
5. Isermann R. Fault diagnosis of machines via parameter estimation and knowledge processing–Tutorial Paper. *Automatica*. 1993;29: 815–835.
6. MacGregor JF, Kourti T. Statistical process control of multivariate processes. *Control Eng Pract*. 1995;3:403–414.
7. Piovoso MJ, Kosanovich KA, Yuk JP. Process data chemometrics. *IEEE Trans Instrum Meas*. 1992;41:262–268.
8. Raich A, Çinar A. Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. *AIChE J*. 1996;42: 995–1009.
9. Bakshi BR. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE J*. 1998;44:1596–1610.
10. Ündey C, Ertunç S, Tatara E, Teymour F, Çinar A. Batch process monitoring and its application to polymerization systems. *Macromol Symp*. 2004;206:121–134.
11. AlGhazzawi A, Lennox B. Monitoring a complex refining process using multivariate statistics. *Control Eng Pract*. 2008;16:294–307.
12. Yu J. Localized Fisher discriminant analysis based complex chemical process monitoring. *AIChE J*. 2011;57:1817–1828.
13. Yu J. Nonlinear bioprocess monitoring using multiway Kernel localized Fisher discriminant analysis. *Ind Eng Chem Res*. 2011;50: 3390–3402.
14. Qin SJ. Statistical process monitoring: basics and beyond. *J Chemom*. 2003;17:480–502.
15. Martin EB, Morris AJ. Non-parametric confidence bounds for process performance monitoring charts. *J Proc Cont*. 1996;6:349–358.
16. Yu J. A particle filter driven dynamic Gaussian mixture model approach for complex process monitoring and fault diagnosis. *J Proc Cont*. 2012;22:778–788.
17. Yu J, Qin SJ. Statistical MIMO controller performance monitoring. *Part II: performance diagnosis*. *J Proc Cont*. 2008;18:297–319.
18. Tatara E, Çinar A. An intelligent system for multivariate statistical process monitoring and diagnosis. *ISA Trans*. 2002;41:255–270.
19. Ündey C, Tatara E, Çinar A. Real-time batch process supervision by integrated knowledge-based systems and multivariate statistical methods. *Eng Appl Artif Intel*. 2003;16:555–566.
20. Sorsa T, Koivo HN. Application of artificial neural networks in process fault diagnosis. *Automatica*. 1993;29:843–849.
21. Chiang LH, Kotanchek ME, Kordon AK. Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Comput Chem Eng*. 2004;28:1389–1401.
22. Yélamos I, Escudero G, Graells M, Puigjaner L. Performance assessment of a novel fault diagnosis system based on support vector machines. *Comput Chem Eng*. 2009;33:244–255.
23. Yu J. A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. *Comput Chem Eng*. 2012;41:134–144.
24. Liu J. Process monitoring using Bayesian classification on PCA subspace. *Ind Eng Chem Res*. 2004;43:7815–7825.
25. Liu J, Chen DS. Fault detection and identification using modified Bayesian classification on PCA subspace. *Ind Eng Chem Res*. 2009;48:3059–3077.
26. Yu J, Qin SJ. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE J*. 2008;54:1811–1829.
27. Yu J, Qin SJ. Multiway Gaussian mixture model based multiphase batch process monitoring. *Ind Eng Chem Res*. 2009;48:8585–8594.
28. Yu J. A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chem Eng Sci*. 2012;68:506–519.
29. Yu J. Multiway discrete hidden Markov model-based approach for dynamic batch process monitoring and fault classification. *AIChE J*. 2012;58:2714–2725.
30. Monroy I, Benitez R, Escudero G, Graells M. A semisupervised approach to fault diagnosis for chemical processes. *Comput Chem Eng*. 2010;34:631–642.
31. Maurya MR, Rengaswamy R, Venkatasubramanian V. Application of signed digraphs-based analysis for fault diagnosis of chemical process flowsheets. *Eng Appl Artif Intel*. 2004;17:501–518.
32. Maurya MR, Rengaswamy R, Venkatasubramanian V. A signed directed graph and qualitative trend analysis-based framework for incipient fault diagnosis. *Chem Eng Res Des*. 2007;85:1407–1422.
33. Chiang LH, Braatz RD. Process monitoring using causal map and multivariate statistics: fault detection and identification. *Chemom Intell Lab Syst*. 2003;65:159–178.
34. Bauer M, Thornhill NF. A practical method for identifying the propagation path of plant-wide disturbances. *J Proc Cont*. 2008;18:707–719.
35. Bauer M, Cox JW, Caveness MH, Downs JJ, Thornhill NF. Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE Trans Control Syst Technol*. 2007;15:12–21.
36. Bauer M, Cox JW, Caveness MH, Downs JJ, Thornhill NF. Nearest neighbors methods for root cause analysis of plantwide disturbances. *Ind Eng Chem Res*. 2007;46:5977–5984.
37. Stockmann M, Haber R, Schmitz U. Source identification of plant-wide faults based on k nearest neighbor time delay estimation. *J Proc Cont*. 2012;22:583–598.
38. Thambirajah J, Benabbas L, Bauer M, Thornhill N. Cause-and-effect analysis in chemical processes utilizing XML, plant connectivity and quantitative process history. *Comput Chem Eng*. 2009;33:503–512.
39. Jiang H, Patwardhan R, Shah SL. Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix. *J Proc Cont*. 2009;19:1347–1354.
40. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco, CA: Morgan Kaufmann, 1988.
41. Murphy K. Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D. Thesis, University of California, Berkeley, 2002.
42. Mehranbod N, Soroush M, Piovoso M, Ogunnaike BA. Probabilistic model for sensor fault detection and identification. *AIChE J*. 2004;49:1787–1802.
43. Mehranbod N, Soroush M, Panjapornpon C. A method of sensor fault detection and identification. *J Proc Cont*. 2005;15:321–339.
44. Verron S, Li J, Tiplica T. Fault detection and isolation of faults in a multivariate process with Bayesian network. *J Proc Cont*. 2010;20:902–911.
45. Bishop CM. Neural Networks for Pattern Recognition. Oxford, UK: Oxford University Press, 1995.
46. Botev Z, Grotowski J, Kroese D. Kernel density estimation via diffusion. *Ann Stat*. 2010;38:2916–2957.
47. Downs JJ, Vogel EF. A plant-wide industrial process control problem. *Comput Chem Eng*. 1993;17:245–255.
48. Ricker NL. Decentralized control of the Tennessee Eastman challenge process. *J Proc Cont*. 1996;6:205–221.

Manuscript received Aug. 8, 2012, and revision received Nov. 12, 2012.