

# Multiscale Principal Component Analysis-Signed Directed Graph Based Process Monitoring and Fault Diagnosis

Husnain Ali, Abdulhalim Shah Maulud,\* Haslinda Zabiri, Muhammad Nawaz, Humbul Suleman, and Syed Ali Ammar Taqvi



Cite This: *ACS Omega* 2022, 7, 9496–9512

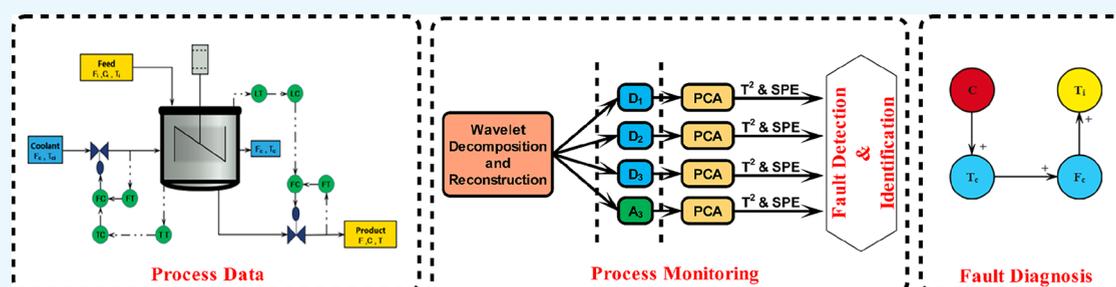


Read Online

ACCESS |

Metrics & More

Article Recommendations



**ABSTRACT:** The chemical process industry has become the backbone of the global economy. The complexities of chemical process systems have been increased in the last two decades due to online sensor technology, plant-wide automation, and computerized measurement devices. Principal component analysis (PCA) and signed directed graph (SDG) are some of the quantitative and qualitative monitoring techniques that have been widely applied for chemical fault detection and diagnosis (FDD). The conventional PCA-SDG algorithm is a single-scale FDD representation origin, which cannot effectively solve multiple FDD representation origins. The multiscale PCA-SDG wavelet-based monitoring technique has potential because it easily distinguishes between deterministic and stochastic characteristics. This study uses multiscale PCA-SDG to detect, diagnose the root cause and identify the fault propagation path. The proposed method is applied to a continuous stirred tank reactor system to validate its effectiveness. The propagation route of most process failures is detected, identified, and diagnosed, which is well-aligned with the fault description, demonstrating a satisfactory performance of the suggested technique for monitoring the process failures.

## 1. INTRODUCTION

With the advent of the fourth industrial revolution (IR 4.0), conventional industrial processes have been autonomous. Many modern industrial processes have well-developed sensors that collect process-related data to detect failures and process system monitoring. Careful monitoring, including process control and appropriate corrective measures, is required to ensure process efficiency. It improves and proceeds to the industrial environments with complete equipment and process automation.<sup>1–3</sup> Fault detection and diagnosis (FDD) systems have seen an increased demand for process safety, dependability, and product quality within process engineering systems. Fault detection and diagnostic methods have been proposed to characterize typical variations in a process plant and detect abnormal deviations. Early detection and diagnosis are always challenging yet needed to prevent process disruptions, shutdowns, or even process failures. It has become a crucial feature due to the advanced and complicated processes involving several variables.<sup>4</sup> For example, defective events, like disabling a catalyst, valve obstruction, or compressor failure, are unavoidably and regularly present.<sup>5</sup> Therefore, early detection and diagnosis increase process safety

and productivity and minimize the process defectivity.<sup>6,7</sup> The process industry might save up to millions of dollars if precise FDD systems are implemented correctly.<sup>8</sup>

Generally, FDD approaches are classified into model-based, knowledge-based, and process methods based on history. Exact and highly dependable process system models are used in model-based techniques. However, accurate primary model expansions are more challenging and almost unachievable for specific processes in modern process systems. The procedures for developing the knowledge models are based on the process conduct and plant operator experience. The accumulation of long-term development of the process knowledge base and expert experience is always in-demand. Instead, the process models and expert's knowledge are unnecessary for approaches

**Received:** December 3, 2021

**Accepted:** February 10, 2022

**Published:** March 9, 2022



that rely on the process history.<sup>9</sup> Furthermore, most process systems employ a distributed control system, and enormous amounts of data can be saved. The process data mining systems can be used for this information. This development in data extracting technology contributed to developing methods based on process history, which are also considered methods based on data. Therefore, most FDD applications are focused on data-driven methodologies in the process industry.<sup>10–13</sup>

The methodology of a signed directed graph (SDG)-based fault diagnostics has improved significantly over the last three decades. An SDG model identifies the flow of information and the effect direction (increase and decrease). The SDG for the modeling of chemical processes was initially introduced by Iri et al.<sup>14</sup> Recently, Maurya et al.<sup>15</sup> introduced the SDG algorithm and diagram for various system types to predict initial and stable system variables for deviations from their nominal fault diagnosis value on external variables and provided techniques for SDG analysis. In the 1980s, SDG-based fault diagnostics were only studied with qualitative system information, and their relationships with the cause/effects were retrieved by some researchers as expert rules.<sup>16,17</sup> In 1989, the concept of defect revealing time was initially presented, and the diagnosis of faults was made online with time data.<sup>18</sup> In the 1990s, different researchers modified the SDG model to address the multifault diagnostic problem.<sup>19</sup> The diagnostic method based on the simultaneously specific and quantifiable methodologies was coupled with the SDG.<sup>20,21</sup> Since 1999, there has been a strong interest in other techniques, including fuzzy-SDG,<sup>22</sup> PCA-SDG,<sup>23</sup> PLS-SDG,<sup>24</sup> and QTA-SDG.<sup>25</sup>

The PCA-based contribution plots are well-known approaches for diagnosing faults.<sup>26</sup> In general, contribution plots indicate the effect of each variable on the statistical index  $T^2$  or squared prediction error (SPE) from the PCA. The leading causes of the failure are the factors with the most contributions. However, the contribution from the root origin variable is extended to other variables, which may not be the root cause of the failure due to the quick interaction between process variables. Vedam and Venkatasubramanian proposed a PCA-SDG method, where the SPE statistic from PCA is used to determine node thresholds and find a consistent SDG model path.<sup>23</sup> However, the contribution can be propagated among variables with fault spreads, making the root cause of the failure hard to locate. These algorithms lead to inaccurate or no diagnosis in the presence of multiple failures. To select the measurable variables for SDG-based diagnosis, the number of thresholds required is decreased to one. Root causes of multiple faults are difficult to identify, reducing the operator number of control actions. The SDG-based fault diagnosis method has an insufficient diagnostic resolution. The standard SDG fault diagnostic methodology originates from a single-scale fault representation and cannot effectively address various roots of fault representation. It causes erroneous representations because of the qualitative nature of the SDG.

Bakshi introduced the MSPCA framework by combining wavelet transforms (WT) with PCA to address these challenges.<sup>27</sup> A specified wavelet family was used to break down each variable individually in this framework. The PCA model was applied to the coefficients at each scale to highlight pertinent events. Furthermore, various multiscale extensions have developed multiscale FDD frameworks in recent years.<sup>28–31</sup> A multiscale approach based on wavelet transform is a crucial technique. It offers several advantages over standard

single-scale approaches as deterministic and stochastic properties differ from the interim measurement of the process.

This study proposes a multiscale PCA-SDG approach to diagnose and identify process failure paths in a continuous stirred tank reactor (CSTR). Its vital contribution is introducing a process monitoring method based on multiscale PCA-SDG. Fault diagnostic techniques based on multiscale PCA-SDG are effective for convenient accuracy and easy diagnosis, and it allows operators to respond to strange events at an early stage.<sup>32</sup> The multiscale PCA-SDG diagnostics framework leads to early identification and diagnosis of unusual circumstances that may react to measured variable contribution changes in correlation information.

The rest of the paper is structured as follows. Section 2 thoroughly explains the theory, methods, and algorithms such as PCA, wavelet transforms, and the SDG. Section 3 focuses on the proposed MSPCA-SDG-based process monitoring and fault diagnosis framework. Section 3.4 describes the CSTR system used as a case study. The presented framework results are shown in Section 4 followed by the conclusion in Section 5.

## 2. METHODOLOGY

**2.1. Principal Component Analysis.** PCA is the multivariate statistical modeling methodology proposed by Pearson<sup>33</sup>

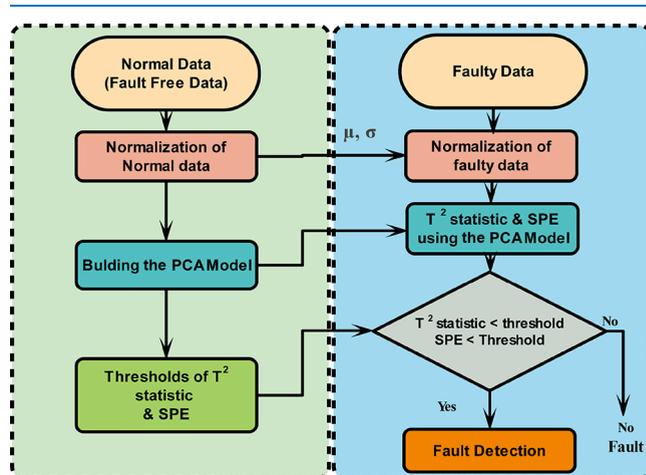


Figure 1. PCA-based monitoring framework.

and Hotelling<sup>34</sup> that identifies directions of considerable data variance by generating variable combinations. PCA has an excellent performance on feature extraction and dimension reduction.

Consider the matrix data set  $X$ , with  $n$  variables and the number of  $m$  observations.

The matrix of data is standardized with unit variance and zero means and interpreted through a single-value decomposition (SVD) into a new matrix<sup>35,36</sup>

$$X = UDV^T \quad (1)$$

where,  $U$  and  $V$  reflected the orthogonal matrixes and  $D$  expressed the diagonal matrix with positive real values.

$$X = TP^T \quad (2)$$

where  $T$  and  $P$  constitute the principal components (PCs) and loading vectors, respectively. These data sets are composed of the eigenvectors related to  $X$ 's covariance data matrix.

The covariance data matrix can be computed accordingly

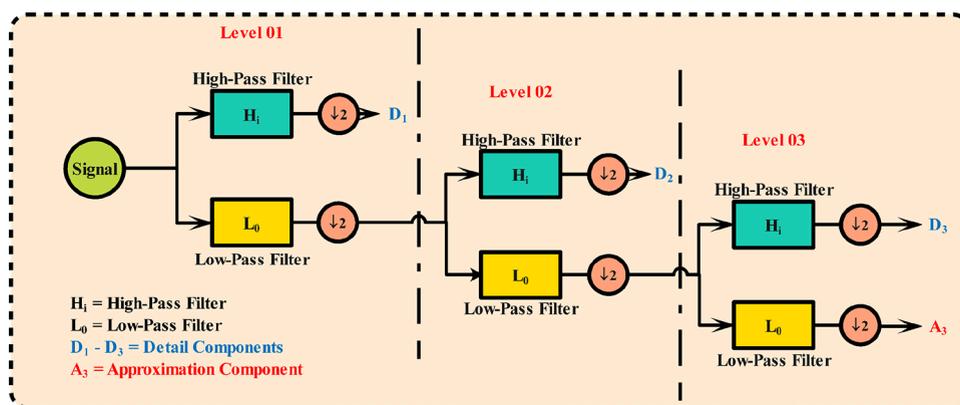


Figure 2. Level three wavelet decomposition.

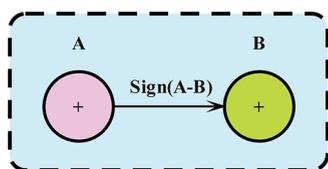


Figure 3. Typical SDG model.

$$\Sigma = \frac{1}{n-1} X^T X = P \Lambda P^T \text{ with } P P^T = P^T P = I_n \quad (3)$$

where  $\Lambda$  represents the eigenvalue diagonal matrix for  $m$  PCs and  $I_n$  represents the identity matrix.

The PCA model includes an optimal number of PCs as the model reliability depends on the number of PCs used. The numbers of suitable PCs for model development are determined using several techniques. These techniques include cross-validation using cumulative percent variance (CPV),<sup>37,38</sup> scree plots,<sup>39,40</sup> and profile likelihood.<sup>41</sup> This work uses the CPV-based methodology to select the dominating PCs, and it may be computed as follows.<sup>38</sup>

$$\text{CPV}(l) = \frac{\sum_{i=1}^l \lambda_i}{\sum_{i=1}^m \lambda_i} \times 100 \quad (4)$$

If  $l$  is the lowest range of PCs according to the specified percent of the total variance, then, in that case, the input data matrix can be displayed after determining the retained number of PCs.<sup>37</sup>

$$X = T P^T = [\hat{T} \tilde{T}] [\hat{P} \tilde{P}]^T \quad (5)$$

where the matrices  $\hat{T}$  and  $\tilde{T}$  comprise the number of principal components retained and ignored.

Meanwhile, the matrices  $\hat{P}$  and  $\tilde{P}$  contain both retained and eigenvectors, respectively. Now  $X$  refers to the given equation.<sup>37</sup>

$$X = \hat{T} \hat{P}^T + \tilde{P} \tilde{T}^T = X \hat{P} \hat{P}^T + X(I_m - \hat{P} \hat{P}^T) \quad (6)$$

In the above equation,  $X \hat{P} \hat{P}^T = \hat{X}$  and  $X(I_m - \hat{P} \hat{P}^T) = E$  are the matrices that represent modeled variation of  $X$  and variations that correspond to the process noise, respectively.

In establishing the PCA model, the two monitoring control charts,  $T^2$  and (SPE), are used to determine the model and the residual variation in space, respectively.

The  $T^2$  statistic can be calculated accordingly.<sup>42</sup>

$$T^2 = x^T P \Lambda^{-1} P^T x \quad (7)$$

The  $T^2$  threshold can be determined as follows:

$$T_{\text{lim}}^2 = \frac{l(m-1)}{m-l} F_{l, m-l, \alpha} \quad (8)$$

where Fisher distributions represented by  $F_{l, m-l, \alpha}$ ,  $l$ , and  $(m-l)$  are the degrees of freedom, whereas the significance level is shown by  $\alpha$

The SPE is determined as follows:<sup>42</sup>

$$\text{SPE} = r^T r = (I - P P^T) x \quad (9)$$

where the residual vector is denoted by  $r$ ; the SPE threshold is calculated as

$$\text{SPE}_{\text{lim}} = \varphi_1 \left[ \frac{h_0 c_\alpha \sqrt{2\varphi_2}}{\varphi_1} + 1 + \frac{\varphi_2 h_0 (h_0 - 1)}{\varphi_1^2} \right]^{1/h_0} \quad (10)$$

where  $\varphi_1 = \sum_{j=l+1}^m \lambda_j^1$ ,  $\varphi_2 = \sum_{j=l+1}^m \lambda_j^2$ ,  $\varphi_3 = \sum_{j=l+1}^m \lambda_j^3$ , and  $h_0 = 1 - \frac{2\varphi_1 \varphi_3}{3\varphi_2^2}$ . The value  $c_\alpha$  deduced from the significant

standard classification. An abnormality is indicated by  $T^2$  and SPE values being over the threshold. The PCA-based monitoring framework is illustrated in Figure 1. First, a PCA model is built using process data collected under normal operating conditions. Then, the  $T^2$  and SPE-based monitoring control charts are calculated on the faulty database. When the control chart of  $T^2$  and SPE surpasses the threshold boundary, as predicted during the model building phase, the processing system has a fault.

**2.2. Wavelet Transforms.** Multiscale process monitoring decomposes the original process data into multiscale components using wavelet transform (WT). According to the properties of the time-frequency locations, the tendency to accurately differentiate approximation function, reflecting deterministic characteristics, and the detailed function showing stochastic characteristics.<sup>43,44</sup> This enables the process phenomena to be more appropriately interpreted in their temporal frequency bands.<sup>45</sup> A mother wavelet  $\psi(\lambda)$  provides all essential functions  $\psi_{a,b}(\lambda)$  inside WT through dilating and by translating processes

$$\psi_{a,b}(\lambda) = a^{-1/2} \psi\left(\frac{\lambda - b}{a}\right) a, b \quad (11)$$

where  $a$  and  $b$  are the discrete parameters, respectively, of scale and orientation. Multiscale displays consist of a low pass and a high pass filter that transmit the signal in various scales. The

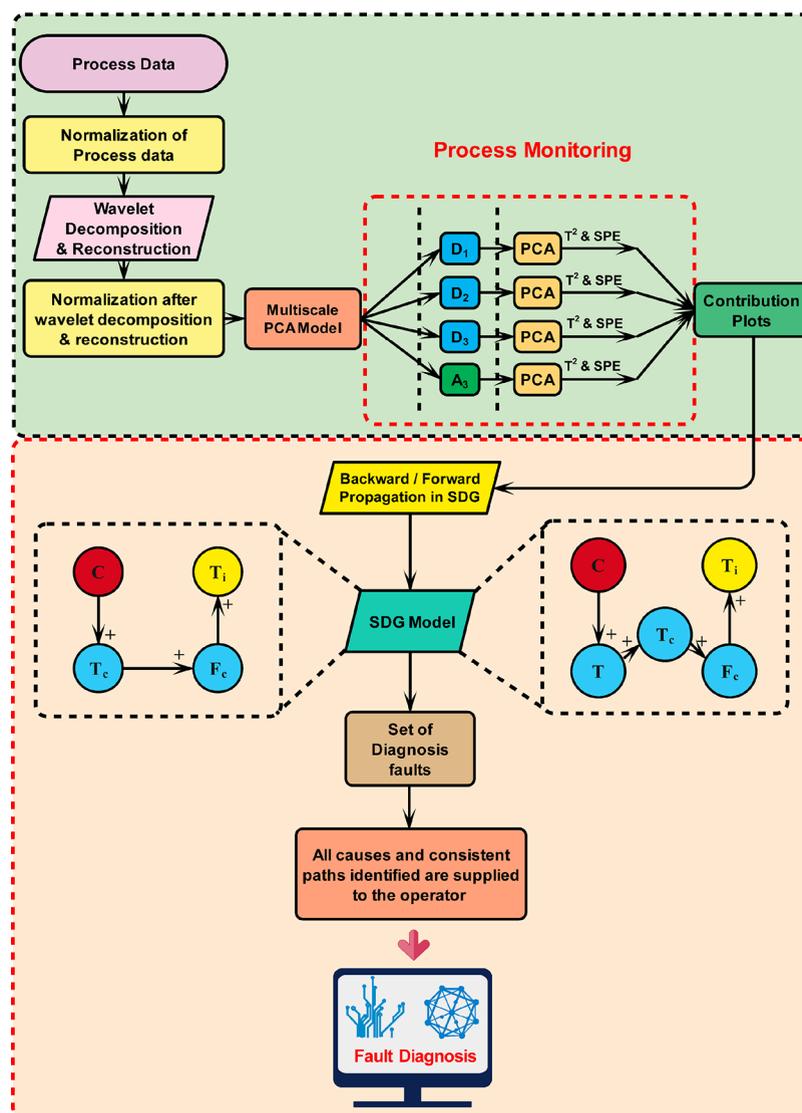


Figure 4. Multiscale PCA-SDG methodology.

original primary signal is projected orthonormal as the deployed scaled version

$$\varphi_{ij}(t) = 2^{-j/2} \varphi(2^{-j/2}t - s) \quad (12)$$

Details and approximation coefficients of the several wavelet functions represented by each level are generated via the signal projection as

$$\psi_{ij}(t) = 2^{-j/2} \psi(2^{-j/2}t - s) \quad (13)$$

The original primary signal is generated and indicated by merging the final deployed scaled signal with all the detailed signals

$$x(t) = \sum_{k=1}^{n2^j} a_{jk} \varphi_{jk} + \sum_{j=1}^J \sum_{k=1}^{n2^j} d_{jk} \psi_{jk}(t) \quad (14)$$

where  $J$  and  $n$  are, respectively, the decomposition level and length of the original signal.

Figure 2 shows that wavelet transformation decomposes the signal at several levels ( $L = 3$ ) followed by the PCA-SDG approach to develop a qualitative multiscale model.

**2.3. Signed Directed Graph.** An SDG represents the causal relationship of the processes, which represents process variables as graph nodes and causal relationships as directed arcs. The node states in the SDG are “0”, “+”, and “−”, which represent the typical values of a stable state. The arc pointing to a node of cause, which might indicate the effect in the same or opposite directions, is either a solid line as positive or a dotted line as negative. The positive arc sign is “+”, while the negative arc sign is “−”. As seen in Figure 3, the cause node “A” and the effect node “B” in the arc (solid line) are both “+”, showing that nodes “A” and “B” are in the same way.<sup>14</sup>

The circle nodes are measurable process variables in the SDG model for fault diagnostics. All the abnormal factors lead to a change in the next node. Various nodes of the reason might hold various sources of the fault. All nodes of reason are root nodes. At least one arc connects a root node with an effect node. An arc is said to be consistent when sign (arc)  $\times$  sign (effect node)  $\times$  sign (cause node) = “+”. The valid path is a cause node, effect nodes, and consistent arc.

The SDG-based approach of fault diagnosis was initially presented and characterized as follows for chemical processes:<sup>46</sup>

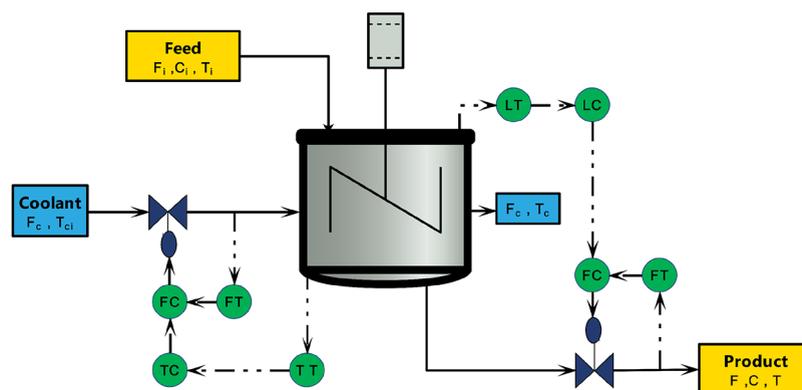


Figure 5. Jacketed CSTR system with cascade control.

$$G = (V, E, \varphi, \psi) \quad (15)$$

where  $V = \{\nu_1, \nu_2, \dots, \nu_n\}$  is a collection of node sets, representing the root cause faults, in which  $E = \{e_1, e_2, \dots, e_m\}$  is a set of branches, describing the causality of the relationship between various nodes, and  $\varphi: E \rightarrow \{+, -\}$ ,  $\varphi(e_k)$  ( $e_k \in E$ ) is the sign of the branch  $e_k$ ; “+” represents a positive impact, and on the contrary, “-” represents a negative impact.  $\psi: V \rightarrow \{+, 0, -\}$ ,  $\psi(\nu_j)$  ( $\nu_j \in V$ ) means the  $\nu_j$  node sign, expressing the node status, and a divergence of the value  $x_{\nu_j}$  projected node value  $\bar{x}_{\nu_j}$ , as indicated in eq 16, determines the status of that node<sup>47</sup>

$$\psi(\nu_j) = \begin{cases} 0, & \text{if } |x_{\nu_j} - \bar{x}_{\nu_j}| < \varepsilon_{\nu_j} \\ +, & \text{if } x_{\nu_j} - \bar{x}_{\nu_j} \geq \varepsilon_{\nu_j} \\ -, & \text{if } x_{\nu_j} - \bar{x}_{\nu_j} \leq -\varepsilon_{\nu_j} \end{cases} \quad (16)$$

### 3. MULTISCALE PCA-SDG-BASED PROCESS MONITORING AND FAULT DIAGNOSIS

#### 3.1. Multiscale PCA-SDG Methodology Framework. A

multiscale approach for the chemical process is the primary

Table 1. CSTR System Variables

Sr. no	system variables for process monitoring	
	process variables	variable description
1	$h$	height of CSTR
2	$C$	reactant A concentration in the reactor
3	$T$	the temperature in the reactor
4	$T_c$	the temperature of the coolant in the jacket
5	$F_i$	feed stream flow rate
6	$F$	the flow rate of the outlet stream
7	$C_i$	reactant A initial concentration in the feed
8	$T_i$	the temperature of the feed stream
9	$T_{ci}$	cooling liquid feed temperature in the cooling jacket
10	$F_c$	cooling liquid flowrate in the cooling jacket

target of the proposed framework. This work proposed and developed a multiscale process monitoring and fault diagnosis technique using the WT and PCA-SDG model characteristics, as shown in Figure 4. When an abnormal situation is monitored, the contribution of observed data to predicted model failure is posted and provided as information to the SDG model for process fault diagnosis.<sup>23</sup> After the fault is detected and the next step is to identify the correlated variables of the fault, the

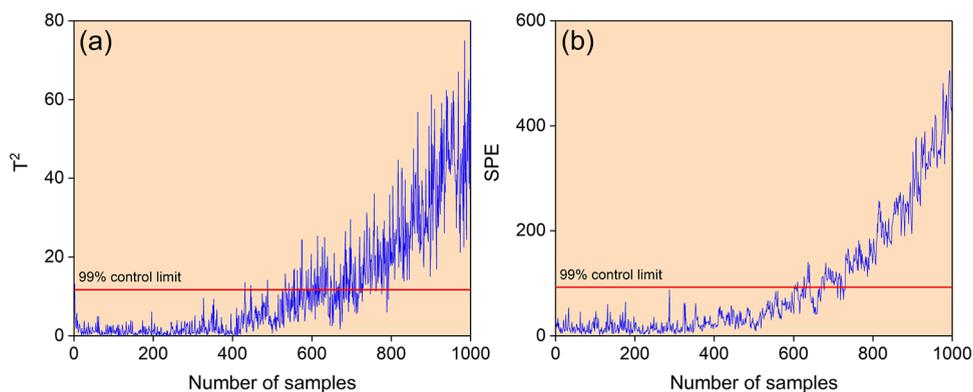
Table 2. Simulated Faults in the CSTR System

fault no.	description	variable	type of fault
1	ramp change	reactant A concentration in the reactor, $C$	sensor bias
2	step change	feed stream flowrate, $F_i$	process disturbance
3	ramp change	the temperature in the reactor, $T$	sensor bias

contribution plot approach is employed. Then, contribution plots are compared to identify the fault-correlated variables effectively and determine the fault propagation paths. When the correlated variables to the problem have been identified, they follow entry into the SDG model developed from the process knowledge to determine the root cause and the fault propagation path.

The significance of the proposed multiscale PCA-SDG-based process monitoring and fault diagnostic technique amplifies the accuracy and quick fault search efficiency. It provides operators with additional time to respond to abnormal occurrences. The frequency of thresholds necessary to detect measured variable variances for SDG-based diagnosis has been decreased to one, and the measurable variable contributions reflect the change in correlation values.

It can be indicated that the proposed developed multiscale PCA-SDG framework is more appropriate and accurate than the conventional PCA-SDG framework. In fault detection, when a fault occurs, the MSPCA model detects faults more immediately and effectively than conventional PCA. In fault identification, some variables show misleading and inaccurate representations due to this single FDD nature of the PCA. The WT-based contribution plots followed by detailed functions ( $D_1$ ,  $D_2$ , and  $D_3$ ) and approximation ( $A_3$ ) show the fault-correlated variables' correct identification. The conventional PCA-SDG fault diagnostic algorithm is a single-scale fault representation origin and cannot efficiently address the issue of actual fault representation roots. The conventional PCA-SDG creates inaccurate and misleading interpretations of fault nodes. The multiscale PCA-SDG framework-based results represent the actual propagation of the fault. The proposed technique shows more efficient fault detection, identification, and diagnosis performance than the conventional PCA-SDG technique. The expressed approach is relatively valuable and adaptable, and it may be applied to any chemical process system.



**Figure 6.** Fault detection monitoring charts of the fault in the reactant A concentration in the reactor based on PCA. (a) Monitoring chart of  $T^2$  and (b) monitoring chart of the SPE.

**3.2. Process Monitoring.** Process monitoring involves the model development phase and fault detection and identification phase.

**3.2.1. Model Development Phase.** The presented MSPCA model is classified into phases. First of all, the methodology uses the normal working conditions of fault-free training data.

**Step 1:** A fault-free normal data set is obtained. Then, the training data using zero mean and unit variance is normalized. After the normalization, every element in the database is decomposed individually using wavelet transformation into wavelet coefficients. After decomposition, the approximations and detail matrices are standardized with mean and standard deviation.

**Step 2:** Using the approximation ( $A_3$ ) and detail matrices ( $D_1$ ,  $D_2$ , and  $D_3$ ) obtained after normalization, the PCA model is developed. Then, the control limits of the monitoring charts are determined based on  $T^2$  and the SPE.

**3.2.2. Fault Detection and Identification Phase.** Detection and identification of faulty data follow model development. The faulty data set may include irregularities that may potentially contribute to the unexpected monitoring system operations.

**Step 1:** First, the faulty data sets are collected. Faulty database information with zero mean and unit variance is normalized. Every element in the faulty database is degraded into wavelet coefficients individually using wavelet transformation. After decomposition of the faulty data set, approximations and detail matrices are normalized with the mean and standard difference, determined in the development phase of the model in step 1.

**Step 2:** In this stage, the  $T^2$  and SPE-based monitoring control charts are calculated on the test database. When the control chart of  $T^2$  and the SPE surpasses the threshold boundary, as predicted during the model building phase, the processing system has a fault. The contribution plots are ultimately deployed to find the faulty variable in both control charts.

**3.3. Fault Diagnosis.** Many researchers discuss diagnostic algorithms based on the SDG. All algorithms use a backtracking search for overall potential routes explaining the impact.<sup>19,48–50</sup> Two alternative techniques of deducing faulty candidates are, as mentioned by Maurya et al.,<sup>15</sup> as follows:

1. The search for the backward method is based on a non-zero-qualitative value of a measured node to gain the possible quality of the subsequent nodes. When there is an additional measurement of any projected node value, the related signal derived from the backpropagation compares with the previous one. This search path is discontinued if

the two values disagree through the other pathways. When all nodes in the measured node are validated, the backward search is accomplished.

2. When reversing, propagation is happening as specified above combined with backward and forward search. Then, forward propagation is started from any potential fault candidate until all non-zero qualitative measures have been acknowledged. Again, the corresponding fault possibility is dismissed if any discrepancy is detected.

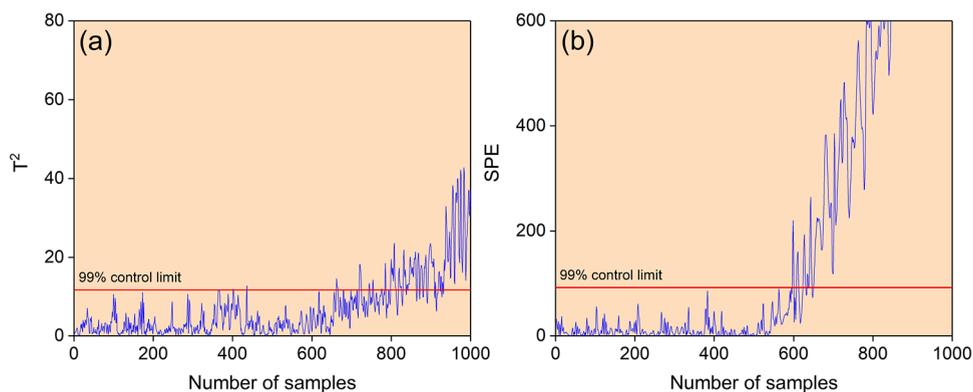
SDG uses a single-variable statistical method to establish the node thresholds, in which correlations between variables are not considered. Thresholds ( $2m$ ) shall be established where  $m$  indicates a variable number of procedures. Because of the performance of the PCA in the multivariable statistics, the threshold can be computed based on the variable contribution to the SPE, which is a remarkable residual space statistic. It requires only one threshold to be set, and the calculation can be significantly reduced.<sup>23</sup> It can be shown that the variables with significant contribution rates change with time. It determines a threshold of variable nodes in the SDG model to discover problematic propagation channels. Since the contribution from one variable transfers to all variables, the difference in the contribution plot for each sample grows considerably more significant as the fault spreads.<sup>51,52</sup> Furthermore, the previously indicated method of determining the threshold using the SPE statistic based on PCA yielded variable results across samples, perhaps failing to eliminate nodes that are not the root cause of the fault.

**3.4. Case Study: Continuous Stirred Tank Reactor (CSTR) System.** In this section, the multiscale PCA-SDG process monitoring and fault diagnosis approach are illustrated in the CSTR case study. An irreversible exothermic reaction of first order occurs in the CSTR system with cascade control shown in Figure 5. As mentioned below, reactant A enters, and product B comes out from the reactor.

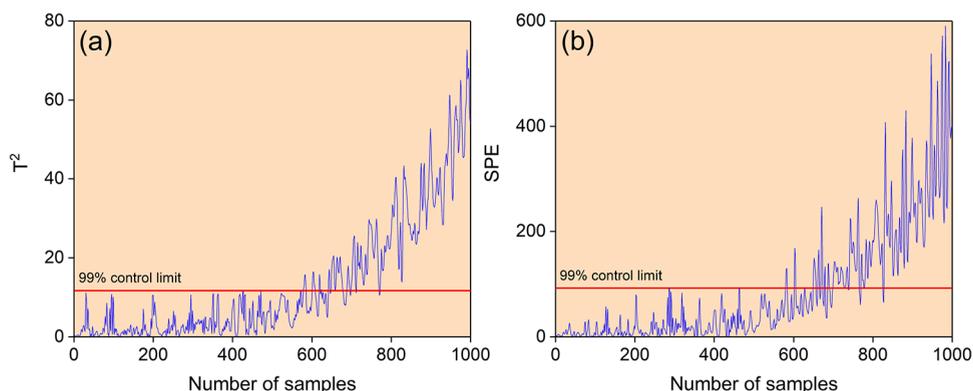


The jacketed refrigeration fluid dissipates heat from the exothermic impact. The reactor temperature and liquid level are regulated by the refreshing and outflow of the coolant correspondingly. The CSTR system models can be explained using the given equations.

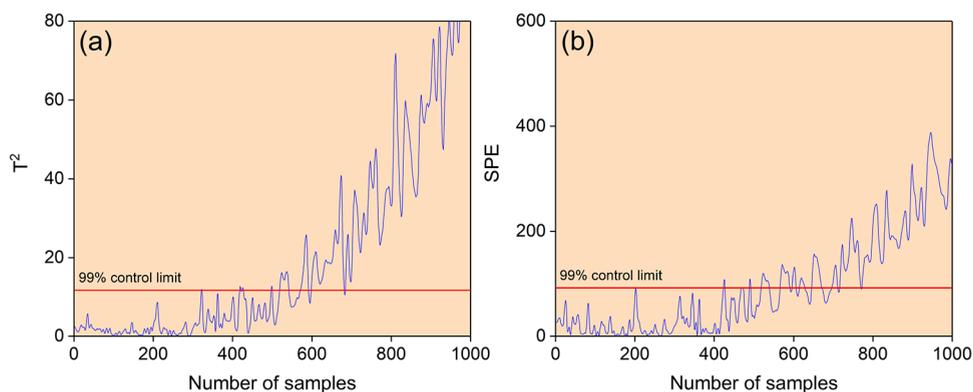
$$\frac{dh}{dt} = \frac{F_i - F}{A} \quad (18)$$



**Figure 7.** Fault detection monitoring charts of the fault in the reactant A concentration in the reactor based on MSPC. (a) Monitoring chart of  $T^2$  based on  $D_2$  and (b) monitoring chart of the SPE based on  $D_2$ .



**Figure 8.** Fault detection monitoring charts of the fault in the reactant A concentration in the reactor based on MSPCA. (a) Monitoring chart of  $T^2$  based on  $D_3$  and (b) monitoring chart of the SPE based on  $D_3$ .



**Figure 9.** Fault detection monitoring charts of the fault in the reactant A concentration in the reactor based on MSPCA. (a) Monitoring chart of  $T^2$  based on  $A_3$  and (b) monitoring chart of the SPE based on  $A_3$ .

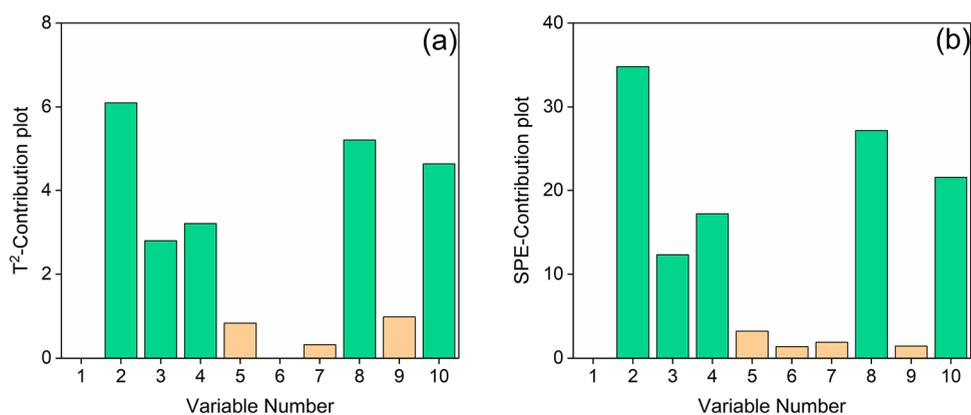
$$\frac{dC}{dt} = \frac{C_i F_i - CF}{Ah} - k_0 \exp\left(\frac{E_A}{RT}\right) C \quad (19)$$

$$\frac{dT}{dt} = \frac{F_i T_i - FT}{Ah} + \frac{k_0 \exp\left(\frac{E_A}{RT}\right) C (-\Delta H)}{\rho C_p} + \frac{UA_c(T_c - T)}{\rho C_p Ah} \quad (20)$$

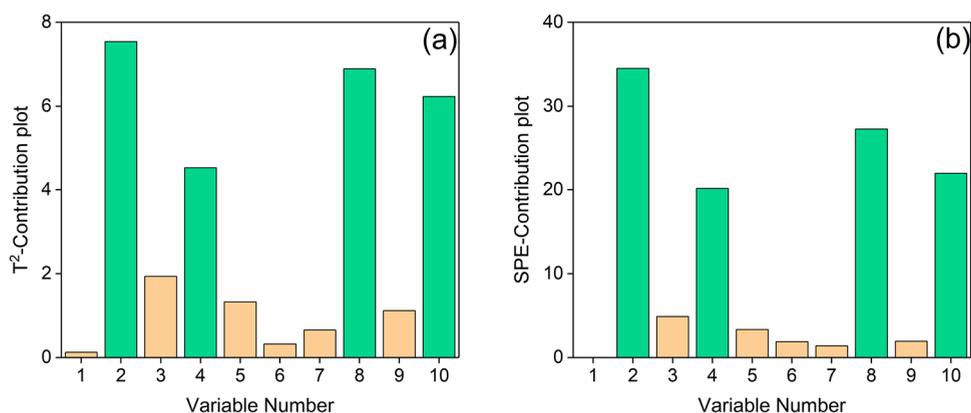
$$\frac{dT_c}{dt} = \frac{F_c(T_{ci} - T_c)}{V_c} + \frac{UA_c(T_c - T)}{\rho_c C_{pc} V_c} \quad (21)$$

The system variables of the CSTR system are displayed in Table 1. CSTR simulation system details are available from Kaisare's work.<sup>53</sup> All simulations and programming are performed using the appropriate toolboxes in the MATLAB/Simulink environment.

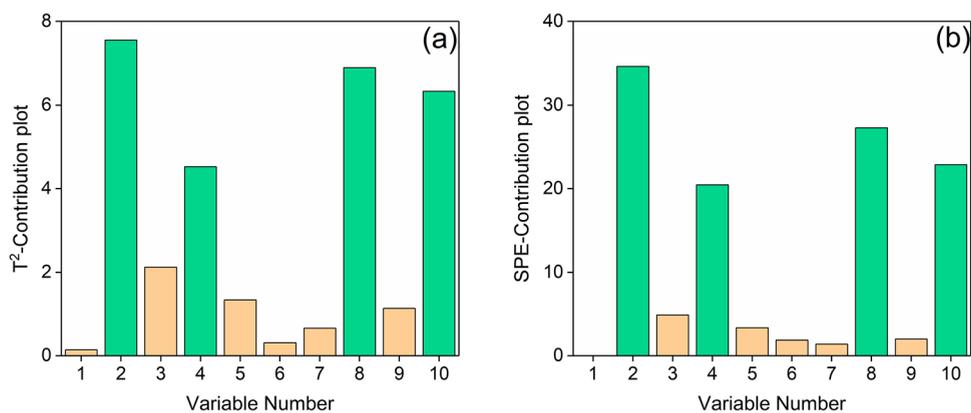
The process data has been generated using the CSTR Simulink model for fault-free and faulty operating circumstances. One thousand samples with normal disorders of the faultless process data have been recorded. In addition, three relevant fault situations were simulated, and 1000 fault pattern samples were recorded. The simulated fault situations, including the sensor biases and process faults, are described in Table 2.



**Figure 10.** PCA-based contribution plots of the fault in the reactant A concentration in the reactor. (a)  $T^2$  contribution and (b) SPE contribution.



**Figure 11.** MSPCA-based contribution plots of the fault in the reactant A concentration in the reactor. (a)  $T^2$  contribution based on  $D_2$  and (b) SPE contribution based on  $D_2$ .

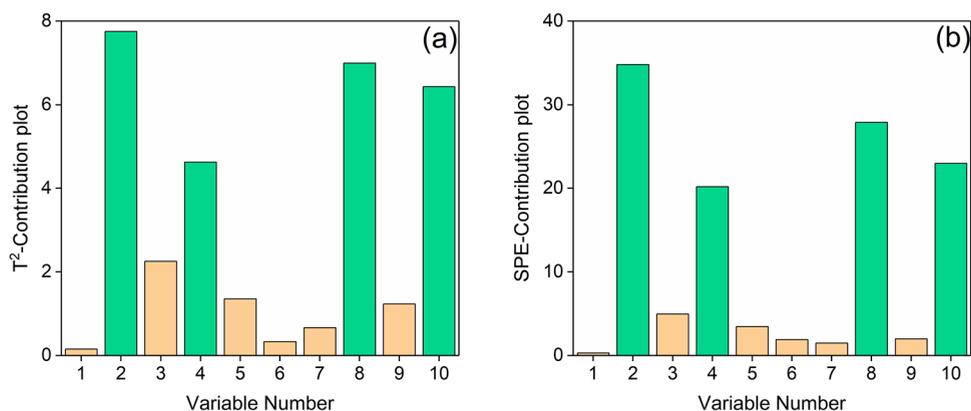


**Figure 12.** MSPCA-based contribution plots of the fault in the reactant A concentration in the reactor. (a)  $T^2$  contribution based on  $D_3$  and (b) SPE contribution based on  $D_3$ .

## 4. RESULTS AND DISCUSSION

This section discusses the conventional single-scale PCA-SDG and multiscale PCA-SDG framework's outcomes depending on approximation ( $A_3$ ) and detailed functions ( $D_1$ ,  $D_2$ , and  $D_3$ ). The detailed function ( $D_1$ ) in this study is not included because it solely contains noises. Therefore, using the proposed methodology and other comparative techniques does not observe fault detection in  $D_1$ . Monitoring charts based on  $T^2$  and the SPE with a confidence level of 99% are generated for all simulations. Confidence limits and statistical values are indicated as solid red and blue lines, respectively.

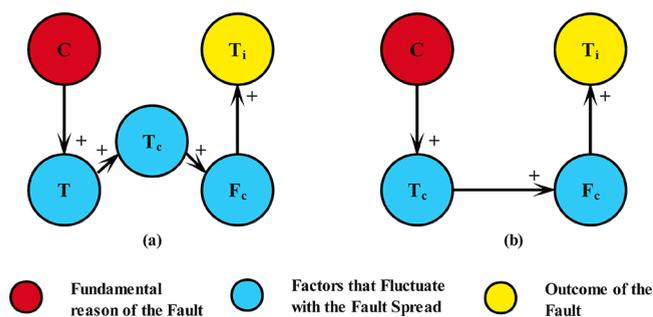
**4.1. Fault 1 - Ramp Change in the Reactant A Concentration in the Reactor.** In this case, a sensor-biased fault has been inducted at 500 sample points into the reactant concentration in the reactor. Figures 6–9 illustrate the monitoring results of PCA and multiscale PCA. The scenario of the fault detection technique based on PCA,  $T^2$ , and SPE monitoring charts detects confined faults, as illustrated in Figure 6a,b. In this scenario, the fault detection technique based on MSPCA followed by detailed functions ( $D_2$  and  $D_3$ ) shows that there is still limited detection of the fault presence on monitoring charts. In particular, the statistical monitoring data are still below the control boundary while the fault is persistent, as shown in



**Figure 13.** MSPCA-based contribution plots of the fault in the reactant A concentration in the reactor. (a)  $T^2$  contribution based on  $A_3$  and (b) SPE contribution based on  $A_3$ .

**Table 3.** Results of Variable Selection by the Contribution Plot Approach

No.	Process variables	Variable description	Fault in the reactant A concentration in the reactor (C)			
			Conventional PCA-SDG	Multi-Scale PCA-SDG		
				$D_2$	$D_3$	$A_3$
1	h	Height of CSTR				
2	C	Reactant A concentration in the reactor	✓	✓	✓	✓
3	T	The temperature in the reactor	✓			
4	$T_C$	The temperature of the Coolant in the jacket	✓	✓	✓	✓
5	$F_i$	Feed stream flow rate				
6	F	The flow rate of the outlet stream				
7	$C_i$	Reactant A initial concentration in the feed				
8	$T_i$	Feed stream temperature	✓	✓	✓	✓
9	$T_{ci}$	Feed temperature of the coolant in the jacket				
10	$F_c$	Coolant flowrate in the jacket	✓	✓	✓	✓



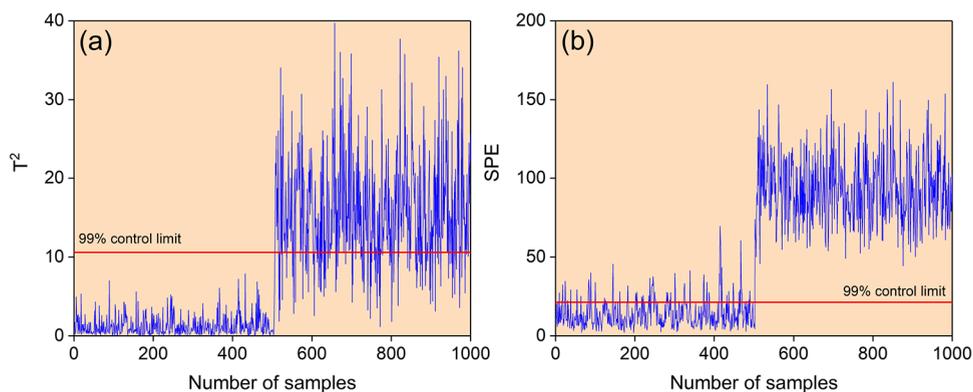
**Figure 14.** Fault propagation root of the fault in the reactant A concentration in the reactor. (a) Conventional PCA-SDG and (b) multiscale PCA-SDG based on the detailed functions ( $D_2$  and  $D_3$ ) and approximation function ( $A_3$ ).

Figures 7 and 8. Similarly, when a fault occurs at the 500th sample point, the MSPCA model followed by approximation ( $A_3$ ) detects the fault immediately, as shown in Figure 9a,b. The MSPCA fault detection results show that all the statistical monitoring data over the confidence limits and monitoring charts efficiently detect faults.

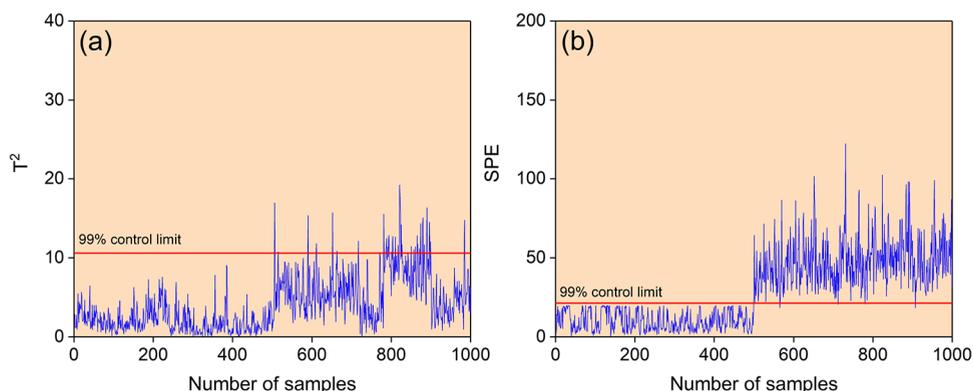
Figures 10–13 show the contribution plot results of PCA and MSPCA. First, to identify the correlated variables of the fault, the contribution plot approach is employed. Then, contribution

plots at the 500th sample are compared to identify the fault-correlated variables effectively. Figure 10a,b shows that the PCA-based contribution plots identify the five variables taking part in the fault. The conventional PCA algorithm is a single-scale FDD representation origin. It cannot effectively solve multiple FDD representation origins, which affect the variables in the same direction because of low FDD resolution. Due to this single-scale FDD nature of the PCA, variable three (temperature in the reactor,  $T$ ) shows misleading and inaccurate representations. The WT-based contribution plots followed by detailed functions ( $D_2$  and  $D_3$ ) and approximation ( $A_3$ ) are shown in Figures 11–13 at the 500th sample and compared to identify the fault-correlated variables, as shown in Table 3. MSPCA-based contribution plots identify the four variables that take part in the fault. They are then chosen and entered in the SDG model to help determine the fault propagation path. It is worth noting that the results are consistent in  $D_2$ ,  $D_3$ , and  $A_3$ . It suggests that selecting the findings at one of these plots to determine the fault propagation is reliable.

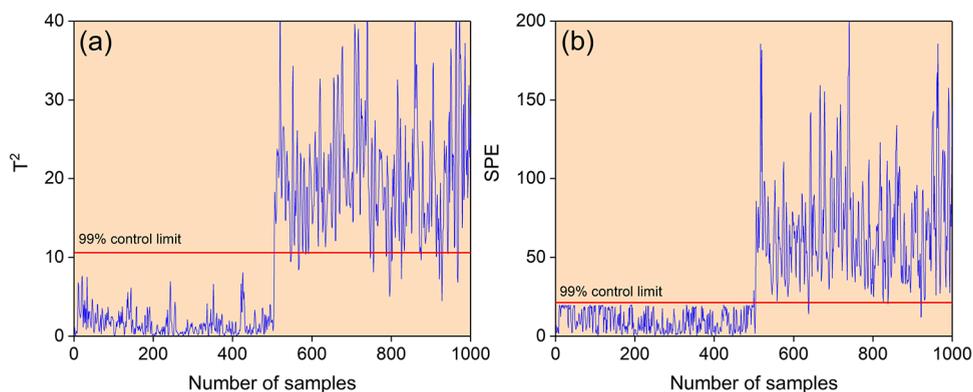
Since the correlated variables to the problem have been identified, they follow entry into the SDG model developed from the process knowledge to determine the root cause. As previously stated, the SDG model of the CSTR system is based on process knowledge. Figure 14a,b shows the fault diagnosis result. The fundamental reason for this failure is the



**Figure 15.** Fault detection monitoring charts of the fault in feed stream flowrate based on PCA. (a) Monitoring chart of  $T^2$  and (b) monitoring chart of the SPE.



**Figure 16.** Fault detection monitoring charts of the fault in feed stream flowrate based on MSPCA. (a) Monitoring chart of  $T^2$  based on  $D_2$  and (b) monitoring chart of the SPE based on  $D_2$ .

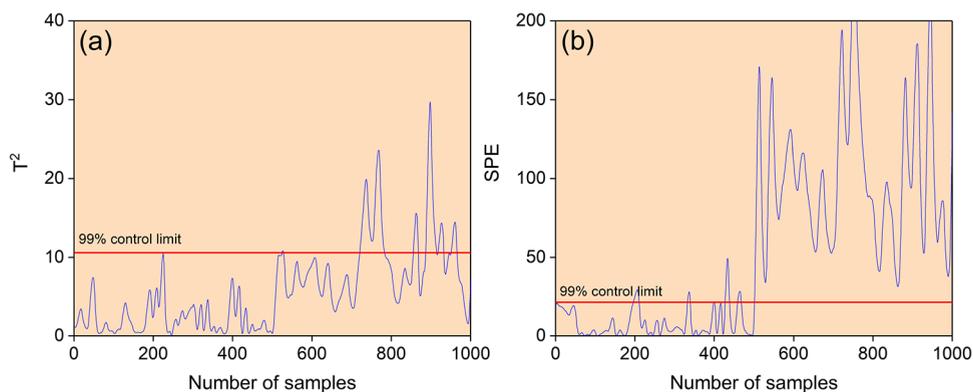


**Figure 17.** Fault detection monitoring charts of the fault in feed stream flowrate based on MSPCA. (a) Monitoring chart of  $T^2$  based on  $D_3$  and (b) monitoring chart of the SPE based on  $D_3$ .

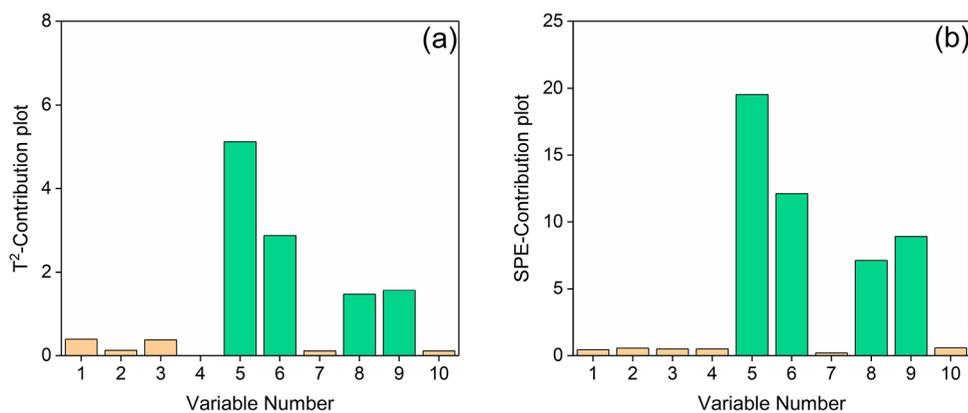
manipulated variable reactant concentration in the reactor ( $C$ ), indicated by a red node. The blue nodes show the factors that fluctuate with the fault spread, while the yellow node represents the outcome of the fault. Figure 14a shows that the conventional PCA-SDG fault diagnostic algorithm is a single-scale fault representation origin and cannot efficiently address the issue of actual fault representation roots. The conventional PCA-SDG creates inaccurate and misleading interpretations of fault node temperature in the reactor ( $T$ ), while the system component undergoes a non-single transformation. Figure 14b shows the multiscale PCA-SDG fault diagnostic results that represent the actual propagation of the fault. The fault (1) description causes a problem with the feed stream temperature ( $T_i$ ) to verify the

diagnosis's conclusion. It states that it may isolate the root cause and identify the channel of propagation of the CSTR system by the proposed fault monitoring and diagnostic approach.

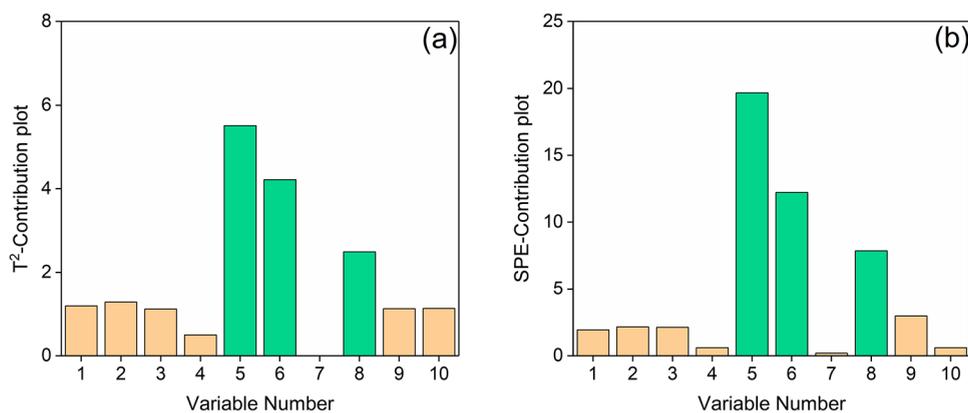
**4.2. Fault 2 - Step Change in the Feed Stream Flowrate.** In this case, a process-disturbed fault has been inducted at 500 sample points into the feed stream flowrate ( $F_i$ ). Figures 15–18 show the monitoring results of PCA and multiscale PCA. The scenario of the fault detection technique based on PCA,  $T^2$ , and SPE monitoring charts detects confined faults, as illustrated in Figure 15a,b. The scenario of the fault detection technique based on MSPCA followed by detailed functions ( $D_2$  and  $D_3$ ) shows that there is still limited detection of the fault presence on monitoring charts. In particular, the



**Figure 18.** Fault detection monitoring charts of the fault in feed stream flowrate based on MSPCA. (a) Monitoring chart of  $T^2$  based on  $A_3$  and (b) monitoring chart of the SPE based on  $A_3$ .



**Figure 19.** PCA-based contribution plots of the fault in feed stream flowrate. (a)  $T^2$  contribution and (b) SPE contribution.

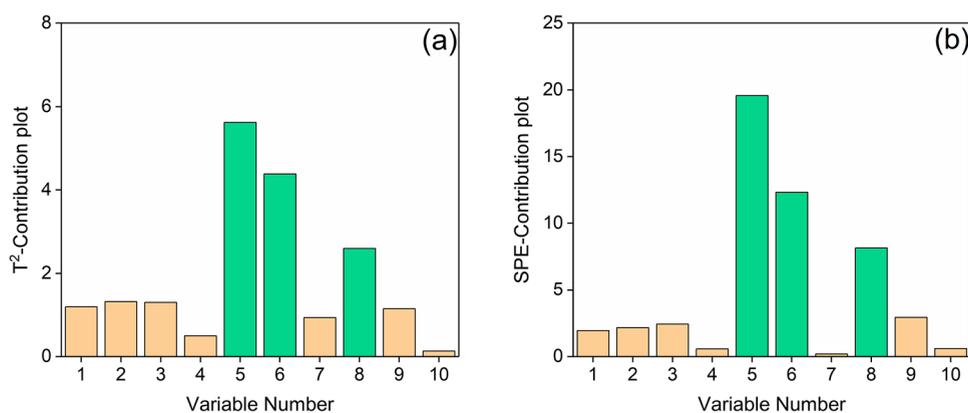


**Figure 20.** MSPCA-based contribution plots of the fault in feed stream flowrate. (a)  $T^2$  contribution based on  $D_2$  and (b) SPE contribution based on  $D_2$ .

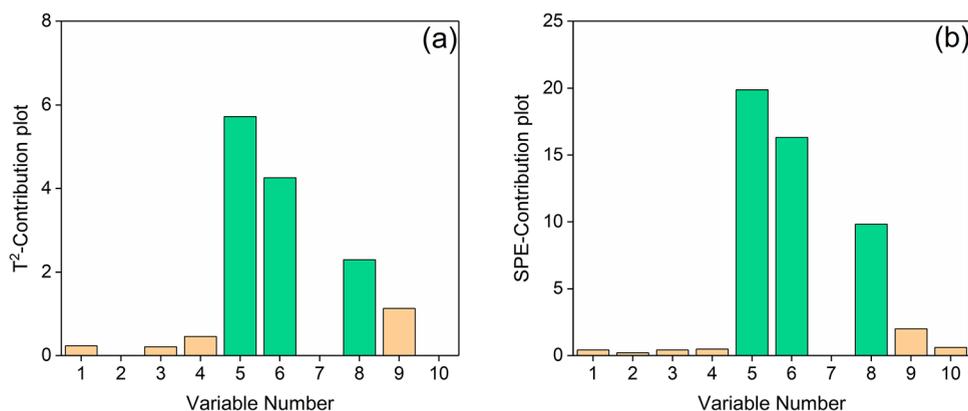
statistical monitoring data are still below the control boundary while the fault is persistent, as shown in Figures 16 and 17. Similarly, when a fault occurs at the 500th sample point, the MSPCA model followed by approximation ( $A_3$ ) detects the fault immediately, as shown in Figure 18a,b. The MSPCA fault detection results show that all the statistical monitoring data over the confidence limits and monitoring charts efficiently detect faults.

Figures 19–22 show the contribution plot results of PCA and MSPCA. First, to identify the correlated variables of the fault, the contribution plot approach is employed. Then, contribution plots at the 500th sample are compared to identify the fault-

correlated variables effectively. Figure 19a,b shows that the PCA-based contribution plots identify the four variables taking part in the fault. The conventional PCA algorithm is a single-scale FDD representation origin. It cannot effectively solve multiple FDD representation origins, which affect the variables in the same direction because of low FDD resolution. Due to this single-scale FDD nature of the PCA, variable nine (temperature of the coolant in the jacket,  $T_c$ ) shows misleading and inaccurate representations. The WT-based contribution plots followed by detailed functions ( $D_2$  and  $D_3$ ) and approximation ( $A_3$ ) are shown in Figures 20–22 at the 500th sample and compared to identify the fault-correlated variables shown in Table 4.



**Figure 21.** MSPCA-based contribution plots of the fault in feed stream flowrate. (a)  $T^2$  contribution based on  $D_3$  and (b) SPE contribution based on  $D_3$ .



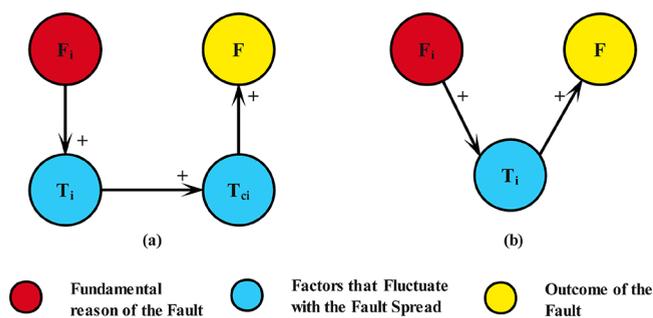
**Figure 22.** MSPCA-based contribution plots of the fault in feed stream flowrate. (a)  $T^2$  contribution based on  $A_3$  and (b) SPE contribution based on  $A_3$ .

**Table 4.** Results of Variable Selection by the Contribution Plot Approach

No.	Process variables	Variable description	Fault in feed stream flowrate ( $F_i$ )			
			Conventional PCA-SDG	Multi-Scale PCA-SDG		
				$D_2$	$D_3$	$A_3$
1	$h$	Height of CSTR				
2	$C$	Reactant A concentration in the reactor				
3	$T$	The temperature in the reactor				
4	$T_c$	The temperature of the Coolant in the jacket				
5	$F_i$	Feed stream flow rate	✓	✓	✓	✓
6	$F$	The flow rate of the outlet stream	✓	✓	✓	✓
7	$C_i$	Reactant A initial concentration in the feed				
8	$T_i$	Feed stream temperature	✓	✓	✓	✓
9	$T_{ci}$	Feed temperature of the coolant in the jacket	✓			
10	$F_c$	Coolant flowrate in the jacket				

MSPCA-based contribution plots identify the three variables that take part in the fault. They are then chosen and entered in the SDG model to help determine the fault propagation path. It is worth noting that the results are consistent in  $D_2$ ,  $D_3$ , and  $A_3$ . It suggests that selecting the findings at one of these plots to determine the fault propagation is reliable.

Since the correlated variables to the problem have been identified, they follow entry into the SDG model developed from the process knowledge to determine the root cause. As previously stated, the SDG model of the CSTR system is based on process knowledge. Figure 23a,b shows the fault diagnosis result. The fundamental reason for this failure is the manipulated variable feed flow rate in the reactor ( $F_i$ ), indicated



**Figure 23.** Fault propagation root of the fault in feed stream flowrate. (a) Conventional PCA-SDG and (b) multiscale PCA-SDG based on the detailed functions ( $D_2$  and  $D_3$ ) and approximation function ( $A_3$ ).

by a red node. The blue nodes show the factors that fluctuate with the fault spread, while the yellow node represents the outcome of the fault. Figure 23a shows that the conventional PCA-SDG fault diagnostic algorithm is a single-scale fault representation origin and cannot efficiently address the issue of actual fault representation roots. The conventional PCA-SDG creates inaccurate and misleading interpretations of the fault node feed temperature of the coolant in the jacket ( $T_{ci}$ ), while the system component undergoes a non-single transformation. Figure 23b shows the multiscale PCA-SDG fault diagnostic results that represent the actual propagation of the fault. The fault (2) description causes a problem with the flow rate of the outlet stream ( $F$ ) that can verify the diagnosis's conclusion. It states that it may isolate the root cause and identify the channel of propagation of the CSTR system by the proposed fault monitoring and diagnostic approach.

**4.3. Fault 3 - Ramp Change in Temperature in the Reactor.** In this case, a sensor-biased fault has been inducted at 500 sample points into the temperature in the reactor ( $T$ ) that explains the monitoring results of PCA and multiscale PCA. Although the case 3 detection monitoring results efficiently detect faults similarly to the previously mentioned cases 1 and 2, they are not illustrated in this scenario. Only the contribution plots based on PCA and multiscale PCA are discussed here.

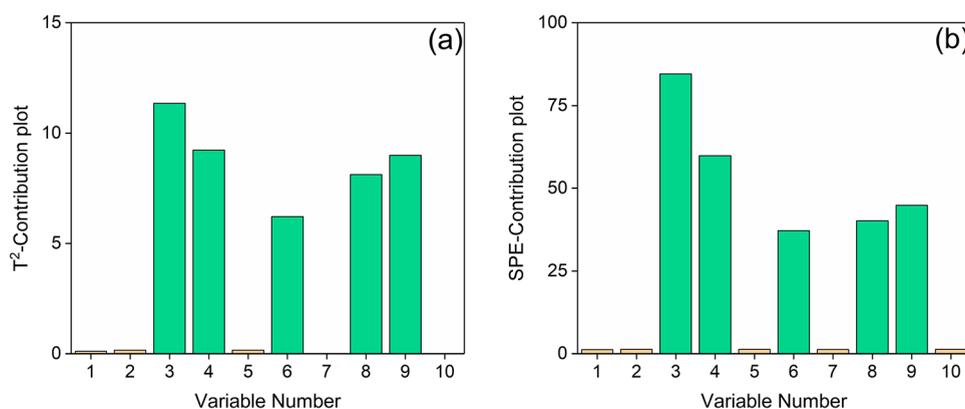
Figures 24–27 show the contribution plot results of PCA and MSPCA. First, to identify the correlated variables of the fault, the contribution plot approach is employed. Then, contribution plots at the 500th sample are compared to identify the fault-correlated variables effectively. Figure 24a,b shows that the PCA-based contribution plots identify the five variables taking part in the fault. The conventional PCA algorithm is a single-

scale FDD representation origin. It cannot effectively solve multiple FDD representation origins, which affect the variables in the same direction because of low FDD resolution. Due to this single-scale FDD nature of the PCA, variable nine (temperature of the coolant in the jacket,  $T_{ci}$ ) shows misleading and inaccurate representations. The WT-based contribution plots followed by detailed functions ( $D_2$  and  $D_3$ ) and approximation ( $A_3$ ) are shown in Figures 25–27 at the 500th sample and compared to identify the fault-correlated variables, as shown in Table 5. MSPCA-based contribution plots identify the four variables that take part in the fault. They are then chosen and entered in the SDG model to help determine the fault propagation path. It is worth noting that the results are consistent in  $D_2$ ,  $D_3$ , and  $A_3$ . It suggests that selecting the findings at one of these plots to determine the fault propagation is reliable.

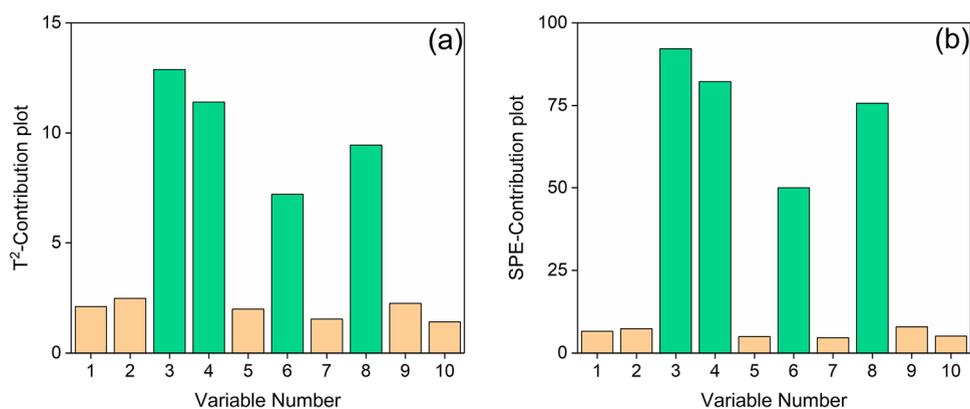
Since the correlated variables to the problem have been identified, they follow entry into the SDG model developed from the process knowledge to determine the root cause. As previously stated, the SDG model of the CSTR system is based on process knowledge. Figure 28a,b shows the fault diagnosis result. The fundamental reason for this failure is the manipulated variable temperature in the reactor ( $T$ ), indicated by a red node. The blue nodes show the factors that fluctuate with the fault spread, while the yellow node represents the outcome of the fault. Figure 28a shows that the conventional PCA-SDG fault diagnostic algorithm is a single-scale fault representation origin and cannot efficiently address the issue of actual fault representation roots. The conventional PCA-SDG creates inaccurate and misleading interpretations of the fault node feed temperature of the coolant in the jacket ( $T_{ci}$ ), while the system component undergoes a non-single transformation. Figure 28b shows the multiscale PCA-SDG fault diagnostic results that represent the actual propagation of the fault. The fault (3) description causes a problem with the temperature of the coolant in the jacket ( $T_{ci}$ ) that can verify the diagnosis's conclusion. It states that it may isolate the root cause and identify the channel of propagation of the CSTR system by the proposed fault monitoring and diagnostic approach.

## 5. CONCLUSIONS

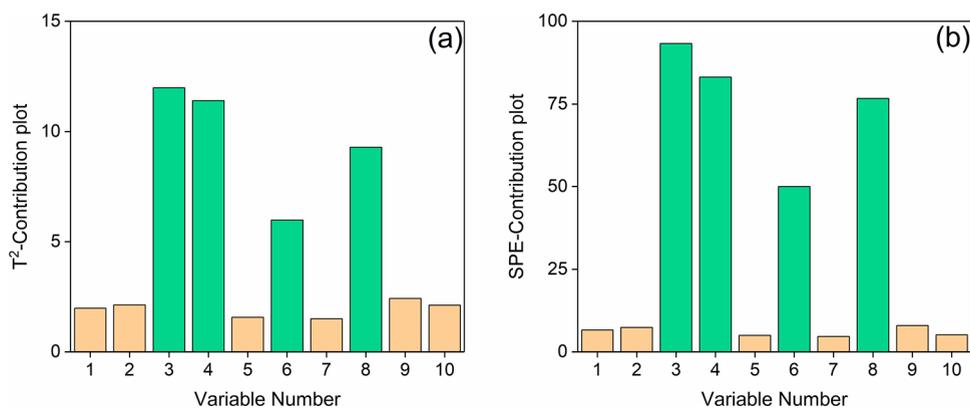
This study proposes a new and effective multiscale PCA-SDG-based process monitoring and fault diagnosis framework that improves fault search efficiency and diagnosis accuracy. This work implements, tests, and compares the process monitoring and fault diagnosis algorithm based on conventional single-scale PCA-SDG and multiscale PCA-SDG in the CSTR system and



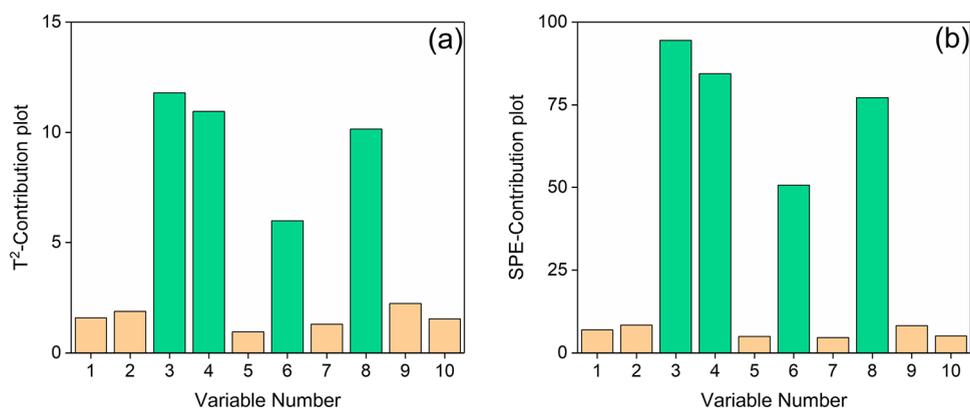
**Figure 24.** PCA-based contribution plots of the fault in temperature in the reactor. (a)  $T^2$  contribution and (b) SPE contribution.



**Figure 25.** MSPCA-based contribution plots of the fault in temperature in the reactor. (a) T<sup>2</sup> contribution based on D<sub>2</sub> and (b) SPE contribution based on D<sub>2</sub>.



**Figure 26.** MSPCA-based contribution plots of the fault in temperature in the reactor. (a) T<sup>2</sup> contribution based on D<sub>3</sub> and (b) SPE contribution based on D<sub>3</sub>.



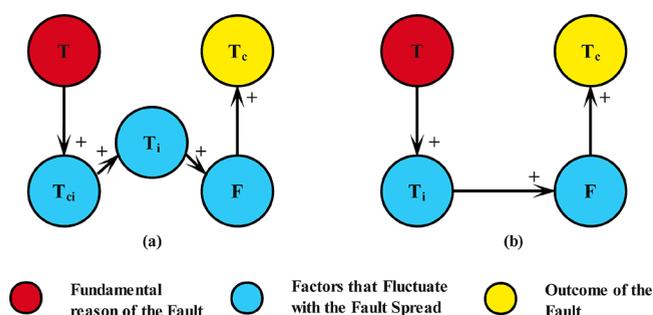
**Figure 27.** MSPCA-based contribution plots of the fault in temperature in the reactor. (a) T<sup>2</sup> contribution based on A<sub>3</sub> and (b) SPE contribution based on A<sub>3</sub>.

their intricate causal representation between process variables. The results show that the proposed multiscale PCA-SDG framework is more appropriate and accurate than the conventional single-scale PCA-SDG framework. The multiscale PCA fault detection results illustrate that the MSPCA model detects faults more immediately and effectively than conventional PCA when a fault occurs. All the monitoring statistics exceed the confidence threshold, and both monitoring charts efficiently detect faults and reduce noise and disturbance. In fault identification, some variables show misleading and inaccurate representations due to the single FDD nature of the PCA. The

contribution plot results based on correlated variable selection by multiscale PCA-SDG followed by WT-based detailed (D<sub>2</sub> and D<sub>3</sub>) and approximation (A<sub>3</sub>) functions consistently show the correct fault-correlated variable identification. However, the conventional PCA-SDG framework is a single-scale fault representation origin that illustrates the false and imprecise interpretations of the propagation path of the fault. It cannot efficiently address the issue of actual fault representation roots. The conventional PCA-SDG creates inaccurate and misleading descriptions of fault nodes. The multiscale PCA-SDG framework results represent the actual propagation of the fault. It can

Table 5. Results of Variable Selection by the Contribution Plot Approach

No.	Process variables	Variable Description	Fault in temperature in the reactor (T)			
			Conventional PCA-SDG	Multi-Scale PCA-SDG		
				D <sub>2</sub>	D <sub>3</sub>	A <sub>3</sub>
1	h	Height of CSTR				
2	C	Reactant A concentration in the reactor				
3	T	The temperature in the reactor	✓	✓	✓	✓
4	T <sub>C</sub>	The temperature of the Coolant in the jacket	✓	✓	✓	✓
5	F <sub>i</sub>	Feed stream flow rate				
6	F	The flow rate of the outlet stream	✓	✓	✓	✓
7	C <sub>i</sub>	Reactant A initial concentration in the feed				
8	T <sub>i</sub>	Feed stream temperature	✓	✓	✓	✓
9	T <sub>ci</sub>	Feed temperature of the coolant in the jacket	✓			
10	F <sub>c</sub>	Coolant flowrate in the jacket				



**Figure 28.** Fault propagation root of the fault in temperature in the reactor. (a) Conventional PCA-SDG and (b) multiscale PCA-SDG based on the detailed functions (D<sub>2</sub> and D<sub>3</sub>) and approximation function (A<sub>3</sub>).

correctly identify active process correction and efficiently specify deterministic and stochastic functions.

Hence, the proposed multiscale PCA-SDG technique shows more efficient fault detection, identification, and diagnosis performance than the conventional PCA-SDG technique. The expressed approach is relatively valuable and adaptable, and it can be applied to any chemical process system. It can provide powerful fault search efficiency, high detection and diagnostic resolution, and accurate fault propagation roots and offer a new approach to ensure process safety.

## AUTHOR INFORMATION

### Corresponding Author

**Abdulhalim Shah Maulud** – Chemical Engineering Department and Centre of Contaminant Control and Utilisation (CenCoU), Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak Darul Ridzuan 32610, Malaysia; [orcid.org/0000-0002-3590-7492](https://orcid.org/0000-0002-3590-7492); Phone: +6053687615; Email: [halims@utp.edu.my](mailto:halims@utp.edu.my)

### Authors

**Husnain Ali** – Chemical Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak Darul Ridzuan 32610, Malaysia; [orcid.org/0000-0002-4877-1045](https://orcid.org/0000-0002-4877-1045)

**Haslinda Zabiri** – Chemical Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak Darul Ridzuan 32610, Malaysia; [orcid.org/0000-0003-1821-1028](https://orcid.org/0000-0003-1821-1028)

**Muhammad Nawaz** – Chemical Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak Darul Ridzuan 32610, Malaysia

**Humbul Suleman** – School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough TS1 3BX, United Kingdom; [orcid.org/0000-0003-1917-6471](https://orcid.org/0000-0003-1917-6471)

**Syed Ali Ammar Taqvi** – Department of Chemical Engineering, NED University of Engineering & Technology, Karachi 75270, Pakistan

Complete contact information is available at: <https://pubs.acs.org/10.1021/acsomega.1c06839>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors appreciate the Universiti Teknologi PETRONAS for administrative assistance, the Chemical Engineering Department for technical encouragement, and Yayasan UTP (cost center: 015LC0-132) for funding support.

## REFERENCES

- Du, X. Fault detection using bispectral features and one-class classifiers. *J. Process Control* **2019**, *83*, 1–10.
- Isermann, R. *Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems*; Springer Science & Business Media, 2011.
- Ming, L.; Zhao, J. Review on chemical process fault detection and diagnosis. In *2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP)*; IEEE, 28–31 May 2017, 2017; pp. 457–462, DOI: [10.1109/ADCONIP.2017.7983824](https://doi.org/10.1109/ADCONIP.2017.7983824).
- Ayoubi, M.; Isermann, R. Neuro-fuzzy systems for diagnosis. *Fuzzy Sets Syst.* **1997**, *89*, 289–307.
- Severson, K.; Chaiwatanodom, P.; Braatz, R. D. Perspectives on process monitoring of industrial systems. *Annu. Rev. Control* **2016**, *42*, 190–200.
- Zhou, Y.; Gao, K.; Li, D.; Xu, Z.; Gao, F. Data-Efficient Constrained Learning for Optimal Tracking of Batch Processes. *Ind. Eng. Chem. Res.* **2021**, *60*, 15658–15668.

- (7) Park, D.; Na, J.; Lee, J. M. Clustered Manifold Approximation and Projection for Semisupervised Fault Diagnosis and Process Monitoring. *Ind. Eng. Chem. Res.* **2021**, *60*, 9521–9531.
- (8) Wang, H.; Chai, T.-Y.; Ding, J.-L.; Brown, M. Data Driven Fault Diagnosis and Fault Tolerant Control: Some Advances and Possible New Directions. *Acta Autom. Sin.* **2009**, *35*, 739–747.
- (9) Frank, P. M. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results. *Automatica* **1990**, *26*, 459–474.
- (10) Gao, K.; Lu, J.; Xu, Z.; Gao, F. Control-Oriented Two-Dimensional Online System Identification for Batch Processes. *Ind. Eng. Chem. Res.* **2021**, *60*, 7656–7666.
- (11) He, Y.-L.; Ma, Y.; Xu, Y.; Zhu, Q.-X. Fault Diagnosis Using Novel Class-Specific Distributed Monitoring Weighted Naïve Bayes: Applications to Process Industry. *Ind. Eng. Chem. Res.* **2020**, *59*, 9593–9603.
- (12) Wang, J.; Zhao, Z.; Liu, F. Robust Slow Feature Analysis for Statistical Process Monitoring. *Ind. Eng. Chem. Res.* **2020**, *59*, 12504–12513.
- (13) Chen, H.; Jiang, B.; Ding, S. X.; Huang, B. Data-Driven Fault Diagnosis for Traction Systems in High-Speed Trains: A Survey, Challenges, and Perspectives. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–17.
- (14) Iri, M.; Aoki, K.; O'Shima, E.; Matsuyama, H. An algorithm for diagnosis of system failures in the chemical process. *Comput. Chem. Eng.* **1979**, *3*, 489–493.
- (15) Maurya, M. R.; Rengaswamy, R.; Venkatasubramanian, V. A signed directed graph-based systematic framework for steady-state malfunction diagnosis inside control loops. *Chem. Eng. Sci.* **2006**, *61*, 1790–1810.
- (16) Kramer, M. A.; Palowitch, B. L., Jr. A rule-based approach to fault diagnosis using the signed directed graph. *AIChE J.* **1987**, *33*, 1067–1078.
- (17) Shiozaki, J.; Matsuyama, H.; O'Shima, E.; Iri, M. An improved algorithm for diagnosis of system failures in the chemical process. *Comput. Chem. Eng.* **1985**, *9*, 285–293.
- (18) Shiozaki, J.; Shibata, B.; Matsuyama, H.; Shima, E. O. Fault diagnosis of chemical processes utilizing signed directed graphs-improvement by using temporal information. *IEEE Trans. Ind. Electron.* **1989**, *36*, 469–474.
- (19) Vedam, H.; Venkatasubramanian, V. Signed digraph based multiple fault diagnosis. *Comput. Chem. Eng.* **1997**, *21*, S655–S660.
- (20) Tarifa, E. E.; Scenna, N. J. Fault diagnosis, direct graphs, and fuzzy logic. *Comput. Chem. Eng.* **1997**, *21*, S649–S654.
- (21) Vianna, R. F.; McGreavy, C. Qualitative modelling of chemical processes - a weighted digraph (WDG) approach. *Comput. Chem. Eng.* **1995**, *19*, 375–380.
- (22) Tarifa, E. E.; Scenna, N. J. Fault diagnosis for MSF dynamic states using a SDG and fuzzy logic. *Desalination* **2004**, *166*, 93–101.
- (23) Vedam, H.; Venkatasubramanian, V. PCA-SDG based process monitoring and fault diagnosis. *Control Eng. Prac.* **1999**, *7*, 903–917.
- (24) Lee, G.; Tosukhowong, T.; Lee, J. H.; Han, C. Fault Diagnosis Using the Hybrid Method of Signed Digraph and Partial Least Squares with Time Delay: The Pulp Mill Process. *Ind. Eng. Chem. Res.* **2006**, *45*, 9061–9074.
- (25) Maurya, M. R.; 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.
- (26) Mnassri, B.; Adel, E. M. E.; Ananou, B.; Ouladsine, M. Fault Detection and Diagnosis Based on PCA and a New Contribution Plot. *IFAC Proc. Vol.* **2009**, *42*, 834–839.
- (27) Bakshi, B. R. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE J.* **1998**, *44*, 1596–1610.
- (28) Maulud, A. H. S.; Wang, D.; Romagnoli, J. A. Wavelet-based nonlinear multivariate statistical process control. In *Computer Aided Chemical Engineering*; Puigjaner, L.; Espuña, A. Eds.; Vol. 20; Elsevier, 2005; pp. 1321–1326.
- (29) Maulud, A.; Wang, D.; Romagnoli, J. A. A multi-scale orthogonal nonlinear strategy for multi-variate statistical process monitoring. *J. Process Control* **2006**, *16*, 671–683.
- (30) Nawaz, M.; Maulud, A. S.; Zabiri, H.; Suleman, H.; Tufa, L. D. Multiscale Framework for Real-Time Process Monitoring of Nonlinear Chemical Process Systems. *Ind. Eng. Chem. Res.* **2020**, *59*, 18595–18606.
- (31) Nawaz, M.; Maulud, A. S.; Zabiri, H.; Taqvi, S. A. A.; Idris, A. Improved process monitoring using the CUSUM and EWMA-based multiscale PCA fault detection framework. *Chin. J. Chem. Eng.* **2021**, *29*, 253–265.
- (32) Taqvi, S. A. A.; Zabiri, H.; Tufa, L. D.; Uddin, F.; Fatima, S. A.; Maulud, A. S. A Review on Data-Driven Learning Approaches for Fault Detection and Diagnosis in Chemical Processes. *ChemBioEng Rev.* **2021**, *8*, 239–259. <https://doi.org/10.1002/cben.202000027> (accessed 2021/11/24)
- (33) Pearson, K. Principal components analysis. *London Edinburgh Dublin Philos. Mag. J. Sci.* **1901**, *6*, 559.
- (34) Hotelling, H. Analysis of a complex of statistical variables with principal components. *J. Educ. Psychol.* **1933**, *24*, 498–520.
- (35) Singh, R. P.; Likins, P. W. Singular Value Decomposition for Constrained Dynamical Systems. *J. Appl. Mech.* **1985**, *52*, 943–948.
- (36) Strang, G. *Linear algebra and its applications*; Harcourt: Brace, Jovanovich, Publishers, 1988.
- (37) Botre, C.; Mansouri, M.; Nounou, M.; Nounou, H.; Karim, M. N. Kernel PLS-based GLRT method for fault detection of chemical processes. *J. Loss Prev. Process Ind.* **2016**, *43*, 212–224.
- (38) Zhu, J.; Ge, Z.; Song, Z. Distributed Parallel PCA for Modeling and Monitoring of Large-Scale Plant-Wide Processes With Big Data. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1877–1885.
- (39) Josse, J.; Husson, F. Selecting the number of components in principal component analysis using cross-validation approximations. *Comput. Stat. Data Anal.* **2012**, *56*, 1869–1879.
- (40) Saccenti, E.; Camacho, J. Determining the number of components in principal components analysis: A comparison of statistical, crossvalidation and approximated methods. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 99–116.
- (41) Zhu, M.; Ghodsi, A. Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Comput. Stat. Data Anal.* **2006**, *51*, 918–930.
- (42) Jackson, J. E.; Mudholkar, G. S. Control Procedures for Residuals Associated With Principal Component Analysis. *Technometrics* **1979**, *21*, 341–349.
- (43) Mallat, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693.
- (44) Nounou, H. N.; Nounou, M. N. Multiscale fuzzy Kalman filtering. *Eng. Appl. Artif. Intell.* **2006**, *19*, 439–450.
- (45) Yoon, S.; MacGregor, J. F. Principal-component analysis of multiscale data for process monitoring and fault diagnosis. *AIChE J.* **2004**, *50*, 2891–2903.
- (46) Umeda, T.; Kuriyama, T.; O'Shima, E.; Matsuyama, H. A graphical approach to cause and effect analysis of chemical processing systems. *Chem. Eng. Sci.* **1980**, *35*, 2379–2388.
- (47) Yong-kuo, L.; Abiodun, A.; Zhi-bin, W.; Mao-pu, W.; Min-jun, P.; Wei-feng, Y. A cascade intelligent fault diagnostic technique for nuclear power plants. *J. Nucl. Sci. Technol.* **2018**, *55*, 254–266.
- (48) Chang, C. C.; Yu, C. C. On-line fault diagnosis using the signed directed graph. *Ind. Eng. Chem. Res.* **1990**, *29*, 1290–1299.
- (49) Nam, D. S.; Han, C.; Jeong, C. W.; Yoon, E. S. Automatic construction of extended symptom-fault associations from the signed digraph. *Comput. Chem. Eng.* **1996**, *20*, S605–S610.
- (50) Ould Bouamama, B.; Biswas, G.; Loureiro, R.; Merzouki, R. Graphical methods for diagnosis of dynamic systems: Review. *Annu. Rev. Control* **2014**, *38*, 199–219.
- (51) Ji, C.; Zhu, X.; Ma, F.; Wang, J.; Sun, W. Fault Diagnosis Algorithm of Chemical Process Based on Information Entropy. *Chem. Eng. Trans.* **2020**, *81*, 541–546.

- (52) Qin, S. J. Survey on data-driven industrial process monitoring and diagnosis. *Annu. Rev. Control* **2012**, *36*, 220–234.
- (53) Kaisare, N. S. *Computational techniques for process simulation and analysis using Matlab®*; CRC Press, 2017, DOI: 10.1201/9781315119519.