Kantorovich Distance Based Fault Detection Scheme for Non-Linear Processes

K. RAMAKRISHNA KINI [ID]1, MRUNMAYEE BAPAT2, AND MUDDU MADAKYARU [ID]2

1Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
2Department of Chemical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding author: Muddu Madakyaru (muddu.m@manipal.edu)

ABSTRACT Fault detection is necessary for safe operation in modern process plants. The kernel principal component analysis (KPCA) technique has been widely utilized for monitoring non-linear processes because it enhances dimension reduction and fault detection in non-linear space. In this paper, an improved non-linear fault detection strategy based on Kantorovich distance (KD) and kernel principal component analysis is proposed. The KD statistic is based on the optimal mass transport theory where the distance between two distributions is computed with respect to a cost function. The addressed fault detection problem models the data using the KPCA framework and utilizes the ability of the KD statistical indicator to detect faults. The detection stage involves comparing the residuals of training fault-free data and testing faulty data using the KD statistic. Additionally, the reference threshold for the KD statistic is computed using the kernel density estimation (KDE) approach as compared to the previously utilized three-sigma rule approach. The detection performance is illustrated with the help of three benchmark case studies: a continuous stirred tank reactor (CSTR) process, Tennessee Eastman (TE) process and an experimental distillation column process. The performance analysis suggests the superiority of the KPCA-KD fault detection scheme in monitoring various sensor faults. Moreover, comparison with traditional statistical indicators of PCA and KPCA schemes shows that the proposed scheme enhances fault detection and achieves an improved detection rate in monitoring different categories of faults.

INDEX TERMS Kantorovich distance, kernal principal component analysis, continuous stirred tank reactor process, Tennessee Eastman process, experimental distillation column process, fault detection.

I. INTRODUCTION

In modern process plants, the important requirement is to ensure process safety as well as consistent product quality and hence, a good fault detection (FD) scheme is required [1]. Advancements in the field of process automation have revolutionized the chemical engineering industry to a large extent with efficient conversion of raw materials such as oil, natural gas and minerals to final products of very good quality. The emergence of smart sensor networks and distributed control systems in chemical industries for catering to continuous needs has complicated the dynamics of chemical industries [2]. Owing to these added complexities, continuous hazards such as emission discharge and explosions occur regularly in process plants. Such hazardous activities pose serious threats to human health and are also the main cause of environmental pollution. Some of the main causes for such

The associate editor coordinating the review of this manuscript and approving it for publication was Mehrdad Saif [ID].

accidents are human error, poor maintenance of the plant and malfunctioning sensors and actuators in the process. Such mishaps can be kept in check if the process plants are continuously monitored. Fortunately, very good progress has been made in monitoring of automated processes in the last few decades by utilizing efficient fault detection schemes [3], [4].

The fault detection schemes can be divided broadly into two categories: model-based schemes and model-free schemes. The model-based FD scheme requires a precise mathematical model for its functioning and this is exigent vowing to the complex nature of chemical plants [5], [6]. The model-free FD schemes are further classified into knowledge-based and historical data-based schemes. The knowledge based schemes require prior expert knowledge of the system and have been frequently applied in recent years for fault detection problems [7]–[9]. In contrast, historical data-based schemes do not require prior expert knowledge and only large historical data are sufficient for monitoring [10]. The multi-variate statistical process

monitoring (MSPM) schemes belong to the family of historical data-driven FD schemes that emphasize monitoring multiple variables simultaneously [11]. Some commonly used MSPM methods include principal component analysis (PCA) [12], [13], partial least squares (PLS) [14], [15], canonical correlation analysis (CCA) [16], slow feature analysis (SFA) [17] and independent component analysis (ICA) [18], [19]. PCA is a linear orthogonal projection strategy that transforms the observations from larger dimensional space onto a smaller dimensional linear subspace by maximizing the variance of projections.

The PCA FD strategy has been established for applications in a variety of FD problems in the last few decades. Several extensions of PCA have been proposed in the literature including dynamic PCA [20], multi-scale PCA [21], multi-block PCA [22] and recursive PCA [23]. Though very popular in the literature, the PCA-based methods adhere to the linear approach which applies only to data sets that are linearly separable and do not consider non-linear variations in process data [24]. To overcome the linear limitation of PCA-based methods, a non-linear scheme based on kernel PCA (KPCA) has been utilized in practice because it has the ability to effectively capture non-linear features in process data by using kernel functions [25], [26]. The kernel PCA computes the principal components in a high-dimensional feature space by using non-linear mapping. In the last few years, considerable work has been carried out to have different extensions of conventional KPCA scheme [27]–[30]. In a recent work proposed by [31] an automatic artificial neural network-based approach with augmented hidden layer was proposed along with a correlation analysis-based approach to tackle the data-distribution mismatch issue. Results was found to improve the computational time and enhanced the monitoring of abnormalities.

The T^2 and squared prediction error (SPE) statistical indicators have been commonly integrated with PCA and KPCA-based modeling techniques. While the T^2 indicator captures variations in the latent subspace, the SPE indicator captures variations in the residual subspace of the model. The T^2 statistic is defined for samples that follow gaussian distribution and it has good ability to detect additive faults [32]. However, it fails in few cases since the latent subspace is sensitive to small changes in the process [33]. In contrast, the SPE statistics seems to be superior as compared to the T^2 statistics for faults other than the additive faults as observed in detection of traction systems in high-speed trains [34]. Also, the SPE statistics is sensitive to the presence of fault at each sampling instant, and this enables it to have improved detection of faults in comparison with the T^2 statistics. There is a good scope for the use of alternative statistical indicators to strengthen the fault detection task. In recent years, Kullback-Divergence (KLD) [32] and Hellinger Distance (HD) [35] indicators that are used to find distance between two probability distribution have been proposed to strengthen the task of fault detection. The performance of KLD and HD based fault indicators were found to be better than conventional T^2 and

SPE based fault indicators [36], [37]. In recent times, the Kantorovich distance (KD)-based indicator has been utilized in fault detection problems [38].

The KD metric indicates the minimum cost to shift a mass of data from source to destination distribution. The KD metric has been integrated with PCA modeling framework and the following advantages were observed: KD metric provided a smooth transition of faults, it detected the faults of small magnitude and provided good monitoring for data corrupted with noise [39]. The KD metric has been utilized for change point detection problem where it yielded good results with a minimum detection delay [40]. Later, the KD metric has been integrated with ICA modeling framework that enhanced monitoring results in comparison with the conventional indicators of PCA and ICA based strategies [41]. Though KD metric is focused on finding the distance between two distributions similar to KLD and HD based metrics, KD metric involves segment by segment comparison between the two distributions in a moving window of fixed length. This enables the KD metric to capture sensitive details in process data which enhances the detection of small magnitude faults. Since the KD statistic has demonstrated improved monitoring results, a fault detection scheme with KPCA as modeling framework and KD metric as statistical index is proposed in this study. The KD statistic is computed between the KPCA residuals generated from normally operating data and the residuals of new online data. In earlier studies on KD-based fault detection, the threshold for the KD statistic was computed using the three sigma rule. However, the three sigma rule may not be efficient way to compute the threshold. Hence, the kernel density approach (KDE) is utilized to compute the threshold in this study.

The reminder of the paper is organized as follows: In Section 2, an introduction to the PCA modeling scheme, the KPCA modeling scheme, along with two statistical indicators T^2 and SPE is provided. In Section 3, an overview of the optimal mass transport, KD statistic formulation, and threshold computation using KDE is discussed. Next, the proposed fault detection strategy that integrates the KPCA modeling technique and KD statistic, is presented in Section 4. In Section 5, the performance of the proposed KPCA-KD FD scheme is studied using three case studies: the CSTR process, the Tennessee Eastman process, and the experimental distillation column process. Comparisons were made between the PCA-T^2, PCA-SPE, PCA-KD, KPCA-T^2 and KPCA-SPE based FD schemes with the proposed KPCA-KD based FD scheme. Finally, the conclusions are presented in Section 6.

II. MODEL DEVELOPMENT
A. PRINCIPAL COMPONENT ANALYSIS
Principal component analysis is a dimensionality reduction technique that transforms variables from a large multi-dimensional space to a lower dimensional space through latent variables or principal components (PCs). For raw data, $\mathbf{X} \in R^{n \times d}$ having n observations and d variables,

$\mathbf{X} = [\mathbf{x_1}^T, \ldots, \mathbf{x_n}^T]$ where $\mathbf{x_i} \in \Re^d$. After normalization, singular value decomposition (SVD) is performed on $\mathbf{X_{sc}}$ to have [42]:

$$\mathbf{X_{sc}} = \mathbf{TV^T} \tag{1}$$

where $\mathbf{T} = [\mathbf{t_1}, \mathbf{t_2} \ldots \ldots \mathbf{t_d}]$ and $\mathbf{V} = [\mathbf{v_1}, \mathbf{v_2} \ldots \ldots \mathbf{v_d}]$ are the scores and loading vectors, associated with covariance of $\mathbf{X_{sc}}$. After computing the optimum PCs, the PCA model is represented as a sum of approximated matrix $\hat{\mathbf{X}}_{sc}$ and a residual matrix $\mathbf{F}$:

$$\mathbf{X_{sc}} = \hat{\mathbf{T}}\hat{\mathbf{V}}^T + \tilde{\mathbf{T}}\tilde{\mathbf{V}}^T = \hat{\mathbf{X}}_{sc} + \mathbf{F} \tag{2}$$

The approximated and residual matrices are evaluated through T^2 and SPE based fault indices. For a new data X_{new}, the T^2 fault indicator is described as [43]:

$$T^2 = \mathbf{X_{new}}^T \hat{\mathbf{V}} \hat{\Lambda}^{-1} \hat{\mathbf{V}}^T \mathbf{X_{new}} \tag{3}$$

For a new data $\mathbf{X_{new}}$, the SPE fault indicator is described as [43]:

$$SPE = \mathbf{X_{new}}^T (\mathbf{I} - \hat{\mathbf{V}}\hat{\mathbf{V}}^T)\mathbf{X_{new}} \tag{4}$$

When the value of T^2 and SPE fault indicators exceed the threshold limits T_α^2 and Q_α respectively, a fault is declared.

B. KERNEL PRINCIPAL COMPONENT ANALYSIS

The conventional PCA is a linear approach and applies only to the data sets that are linearly separable. If the data is non-linear and cannot be expressed in linear space, then the conventional PCA method fails to capture the non-linearities which eventually affects monitoring performance. The kernal PCA approach aims to project the dataset from a lower dimension to a higher dimension feature space where the data set is linearly separable. Consider a data $\mathbf{X} \in R^{n \times d}$ having n observations and d variables, $\mathbf{X} = [\mathbf{x_1}^T, \ldots, \mathbf{x_n}^T]$ with $\mathbf{x_i} \in \Re^d$. The data are transferred to linear feature space $\mathbf{F}^d$ and covariance matrix is computed as:

$$\Sigma^{\mathbf{F}} = \frac{1}{n} \sum_{j=1}^{n} \Psi(\mathbf{x_j})\Psi(\mathbf{x_j})^T \tag{5}$$

The $\Psi(.)$ is a mapping function. The kernel PCs (KPCs) are calculated by solving the eigenvalue problem $\lambda\mathbf{v} = \Sigma^{\mathbf{F}}\mathbf{v}$:

$$\lambda\mathbf{v} = \left(\frac{1}{n} \sum_{j=1}^{n} \Psi(\mathbf{x_j})\Psi(\mathbf{x_j})^T\right)\mathbf{v} = \frac{1}{n} \sum_{j=1}^{n} \langle\Psi(\mathbf{x_j}), \mathbf{v}\rangle \Psi(\mathbf{x_j}) \tag{6}$$

where λ, $\mathbf{v}$ and $\langle a, b\rangle$ denotes eigen values, eigen vectors and respective dot product between a and b. The $\lambda\mathbf{v} = \Sigma^{\mathbf{F}}\mathbf{v}$ is equal to:

$$\langle\Psi(\mathbf{x_k}), v\rangle = \left\langle\Psi(\mathbf{x_k}), \Sigma^F\right\rangle, \quad (k = 1, 2, \ldots \ldots n) \tag{7}$$

The coefficient of each sample is a_i ($i = 1, 2, \ldots \ldots$ n) such that:

$$\mathbf{v} = \sum_{i=1}^{n} a_i \Psi(\mathbf{x_i}) \tag{8}$$

Combining equations (7) and (8) yields the following:

$$\lambda \sum_{i=1}^{n} a_i \langle\Psi(\mathbf{x_k}), \Psi(\mathbf{x_i})\rangle$$
$$= \frac{1}{n} \sum_{i=1}^{n} a_i \left\langle\Psi(\mathbf{x_k}), \sum_{i=1}^{n} \Psi(\mathbf{x_i})\right\rangle \langle\Psi(\mathbf{x_j}), \Psi(\mathbf{x_i})\rangle \tag{9}$$

For $(k = 1, 2, \ldots \ldots n)$ and by introducing the kernel matrix $\mathbf{K}$, the equation (9) is reduced to [28], [29]:

$$\lambda\alpha = \frac{1}{n}K\alpha \tag{10}$$

where $\alpha = [a_1, a_2, \ldots a_d]^T$. Next, the Kernel PCs are computed using the following expression for new data:

$$t_k = \langle\mathbf{v_k}, \Psi(\mathbf{x})\rangle = \sum_{j=1}^{n} \alpha_j^k \langle\Psi(\mathbf{x_j})\Psi(\mathbf{x})\rangle \tag{11}$$

Once KPCA model is developed, it can be used to compute faults in new data using T^2 and SPE statistical indicators. The variations in model is computed as follows [25]:

$$T^2 = [t_1, t_2, \ldots, t_p]\Lambda^{-1}[t_1, t_2, \ldots, t_p]^T \tag{12}$$

where $\Lambda = diag(\lambda_1 > \lambda_2 > ..\lambda_p)$. The residual subspace of the KPCA model is monitored using the SPE statistical index and it is computed as follows:

$$SPE = ||\Psi(\mathbf{x}) - \Psi(\mathbf{p}(\mathbf{x}))||^2 \tag{13}$$

where, $\Psi(\mathbf{p}(\mathbf{x}))$ indicates p PCs in higher dimensional feature space. When the value of T^2 and SPE fault indicators exceed the threshold limits, a fault is declared [26].

III. KANTOROVICH DISTANCE

The Kantorovich distance is derived from the concept of the optimal mass transport theory. According to this theory, information can be efficiently transferred from source to destination distribution with reference to a cost function [44]. Because of the ability to compare signals and data from different distributions precisely, methods based on optimal mass transport theory are gaining prominence in various engineering domains. For any two distributions Q and R with supports X and Y, the optimal mass transport proposed by Monge is presented as:

$$M(Q, R) = \inf_{f \in MP} \int_X m(x, f(x))dQ(x) \tag{14}$$

where f(x) is mapping function and m(.,.) is a cost function and MP is measure-preserving mappings which is defined as:

$$MP = \left\{f : X \to Y \middle| \int_{f^{-1}(A)} dP(x) = \int_A dQ(x)\right\} \tag{15}$$

In the above formulation proposed by Monge, a mass could be relocated from one point in the source distribution to one point in the destination distribution. However, this formulation was improved by Kantorovich where multiple points from source distribution can be mapped with multiple points

---

Algorithm 1 Segmentation Process for KD Computation

1) The observations of Q are divided into l segments $Q_1, Q_2 \ldots Q_l$ and each segment is made of j data points, that is, $Q_1 = [q_{11}, q_{12} \ldots . q_{1j}]$, $Q_2 = [q_{21}, q_{22} \ldots . q_{2j}] \ldots Q_l = [q_{l1}, q_{l2} \ldots . q_{lj}]$.
2) The observations of R are divided into l segments $R_1, R_2 \ldots R_l$ and each segment is made of j data points, that is, $R_1 = [r_{11}, r_{12} \ldots . r_{1j}]$, $R_2 = [r_{21}, r_{22} \ldots . r_{2j}] \ldots R_l = [r_{l1}, r_{l2} \ldots . r_{lj}]$.
3) Each segment of Q is compared against each segment of R which implies that segment of source distribution R_1 is compared with all segments R_1, R_2 and R_3 until R_l of the destination distribution. The first sample of $Q_1(q_{11})$ is compared with the first sample of $R_1(r_{11}), R_2(r_{21}) \ldots R_l(r_{l1})$. The second sample of $Q_1(q_{12})$ is compared with the second sample of $R_1(r_{12}), R_2(r_{22}) \ldots R_l(r_{l2})$ and so on until all segments are covered.
4) The next segment Q_2 is compared with all segments R_1, R_2 and R_3 until R_l of distribution R. This continues until all the segments of Q undergo comparison with all the segments of R.
5) The distances are recorded between the two distributions, and evaluated using the KD statistic. If Q and R are similar, the KD statistic is small, and if they are dissimilar, the KD statistic is large.

---

in destination distribution and this method was named as Kantorovich Distance. This improved one-to-many mapping ultimately provided a more feasible solution than the earlier proposed one-to-one mapping [45].

For two distributions Q and R, the minimum shifting distance required for data from Q to R is termed the Kantorovich distance. The cost of transportation will be very low or close to zero for two similar distributions and large for dissimilar distributions, thus, indicating the degree of dissimilarity between the two [40]. Process monitoring problems measure the dissimilarity while comparing non-faulty data and faulty data, and hence, the KD indicator has scope for application in the detection of faults. The optimal mass transport with respect to cost function l for case $l = 1$ is termed as the Earth mover's distance and is expressed mathematically as follows:

$$W_r(Q, R) = \left(\inf_{\gamma \in \Gamma(Q,R)} \int ||q - r||^l d\gamma(q, r) \right)^{1/l} \quad (16)$$

$\Gamma(Q, R)$ represents the set of joint distributions and γ denotes the minimum optimal coupling [46]. For $l = 2$, the optimal mass transport is termed as Kantorovich Distance and expressed mathematically in the following manner:

$$W_2(Q, R)$$
$$= \left(||\mu_q - \mu_r||_2^2 + Tr\left(\Sigma_q + \Sigma_r - 2(\Sigma_q^{\frac{1}{2}} \Sigma_r \Sigma_q^{\frac{1}{2}})^{\frac{1}{2}} \right) \right)^{\frac{1}{2}} \quad (17)$$

where μ_q and μ_r represent the means while Σ_q and Σ_r represents the covariance matrix of two random variables q and r.

The computation of the KD metric is based purely on the available data and hence, it is not restricted to any particular distribution. This enables it to be applied to any type of distribution that includes Gaussian, non-Gaussian or exponential distributions. The KD metric between two distributions is calculated using a segmentation process. Since comparison of individual observations in source and destination distributions is a time-consuming process, the data in both distributions are divided into multiple segments. Next, each segment from source distribution undergoes comparison

with all the segments in the destination distribution. The division of data into different segments depends on the moving window parameter, which is selected based on a particular application. The computation of the KD metric between distribution Q and distribution R is presented in Algorithm 1.

A. KD STATISTIC THRESHOLD

In fault detection problems, statistical indicator is compared with reference threshold to determine the presence of fault. In earlier studies on KD-based fault detection, a simple three-sigma rule was used to compute the threshold for KD indicator [41]. The training data were initially split into two parts *Tr1* and *Tr2*. Next, a multi-variate model was developed for both data-sets and the residuals *Ru1* and *Ru2* were generated. The KD metric *KDA* was computed for *Ru1* and *Ru2*. Next, the threshold α was computed using the following expression [39]:

$$\alpha = \mu_{KDA} + 3\sigma_{KDA} \quad (18)$$

where μ_{KDA} and σ_{KDA} represent the mean and standard deviation of KD metric *KDA*.

However, the three-sigma threshold rule may not be an efficient option in all the scenarios. In few process applications, the data could be Gaussian or non-Gaussian in nature and in such cases, a better approach may be required for determining the threshold if the KD statistical indicator were to be applied. In this study, kernel density estimation (KDE) approach is proposed to be used as a threshold for KD-based fault indicator. KDE refers to the class of data-driven techniques for the non-parametric estimation of density functions. KDE is a powerful tool that can determine empirical distribution density function from the given samples from the population under consideration. The advantage of using the KDE approach in determining the threshold is that it follows the data more closely and gives very less prominence to the region of unknown operation [18]. The KDE approach has been used as a threshold for many fault detection problems in the literature [47], [48]. For a given random variable y, its PDF $f(\hat{y})$ can be estimated by n samples y_i,
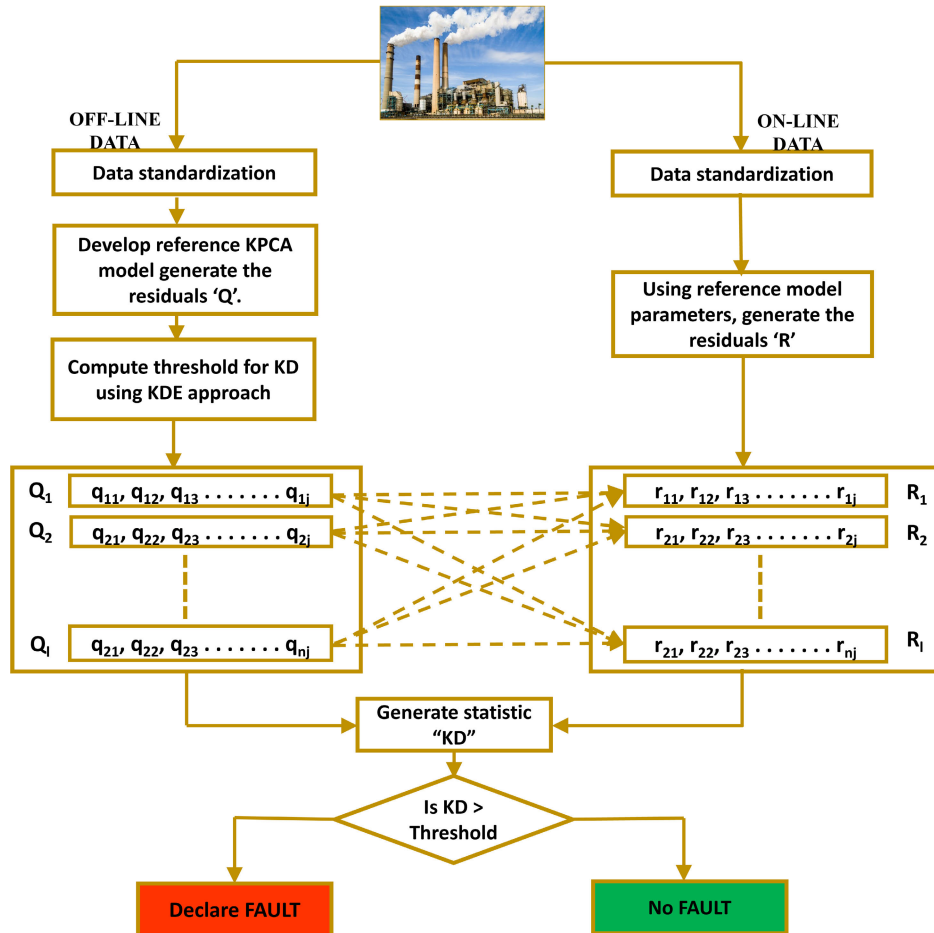
FIGURE 1. Block diagram representation of proposed KPCA-KD fault detection scheme.

$i = 1, 2, \ldots n$, as follows:

$$f(\hat{y}) = \frac{1}{nh} \sum_{i=1}^{n} K\{\frac{y - y_i}{h}\} \qquad (19)$$

where K is the univariate kernel estimator, n is the number of observations and h is the smoothing parameter. The kernel estimator is the sum of the bumps that are placed at different sample points. The smoothing parameter needs to be chosen carefully because small value makes the signal less smooth and a large value will oversmoothen the signal [47], [49]. The training data are initially split into two parts *Te1* and *Te2*. Next, the KPCA multi-variate model is developed for both data-sets, and the residuals *Re1* and *Re2* are then generated. The KD metric *KD* is computed between *Re1* and *Re2*. This is followed by the estimation of the density function of the KD statistic using uni-variate kernel density estimator presented in equation (19). The point occupying 99% of the total area in the density function plot is treated as the threshold limit for KD statistic.

IV. FAULT DETECTION USING KPCA-BASED KD STRATEGY

The primary objective of this work is to develop an enhanced fault detection scheme based on the KD metric to improve the

detection of sensor faults. The conventional fault indicators of PCA and KPCA based fault detection schemes have a few shortcomings, which makes the detection process less efficient. Hence, an alternative fault indicator option in the form of KD index is considered in this study. The residuals of any developed multi-variate model contain important information related to the process. For a new data **X**, the residuals from the reference KPCA model are computed using the following expression:

$$\mathbf{E} = \Psi(\mathbf{X}) - \Psi(\mathbf{X})\mathbf{v}_p\mathbf{v}_p^T \qquad (20)$$

where $\mathbf{v}_p$ denotes eigen vectors corresponding to p optimum PCs. If the process is having some abnormalities, information regarding abnormality will be reflected in the residuals. The proposed KD-based monitoring compares the residuals of training fault-free data and the testing data. The residuals of both datasets are evaluated using the KD metric to indicate the presence of abnormality. The proposed fault detection scheme is presented through block diagram format in Figure 1. The proposed scheme is divided into offline development and online monitoring stages, which are briefly described below:

   1) Offline development stage:

- Step-1: The data under normal operating conditions of the plant are recorded and scaled to zero mean and unit variance.
- Step-2: The data are projected on to higher dimensional feature space where kernel PCA model is developed, and dominant components are selected using the CPV approach.
- Step-3: The residuals *R1* are generated for the scaled data from the KPCA model parameters.
- Step-4: The data are split into two equal units and then, the KD statistic is computed between the two units and the estimation of the density function of KD statistic is used to determine the threshold α.

2) Online monitoring stage:
- Step-1: A new online data (possible faulty data) are scaled to zero mean and unit variance.
- Step-2: The residuals *R2* are generated for new online data from the reference KPCA model.
- Step-3: The KD statistic *KD* is developed between *R1* and *R2* using a moving window of fixed length as presented in Algorithm 1.
- Step-4: The online data is free of faults if $KD < \alpha$ and deemed to be faulty if $KD > \alpha$.

The performance of any FD strategy involving the KD statistic depends on the choice of the moving window parameter. The choice of this parameter depends solely on the type of process data and the amount of noise content in the data. More details regarding the selection of this parameter can be found in [40].

V. RESULTS AND DISCUSSION
To confirm the effectiveness of the proposed KPCA-KD fault detection scheme, three applications are considered: continuous stirred tank reactor process, Tennessee Eastman process and experimental distillation column process. This study compares PCA-T^2, PCA-SPE, PCA-KD, KPCA-T^2 and KPCA-SPE strategies with the proposed KPCA-KD based strategy. The fault detection rate (FDR) and false alarm rate (FAR) metrics are used for a fair comparison between different strategies [41].

A. CONTINUOUS STIRRED TANK REACTOR (CSTR)
A nonlinear continuous stirred tank reactor (CSTR) is considered to validate the performance of proposed KPCA-KD FD scheme. The CSTR problem has been used in many fault detection based problems over the last few years [29], [50], [51]. The CSTR process involves a non-isothermal, irreversible first order reaction of the form:

$$A \rightarrow B \tag{21}$$

where *A* is the reactant species, and *B* is the desired product. A schematic of the CSTR process considered in this study is shown in Figure 2. The reaction system is highly exothermic and therefore, the fluid in jacket of the reactor is used to cool the reactor. The feed flow rate and coolant flow rate is regulated to obtain the desired product concentration, that
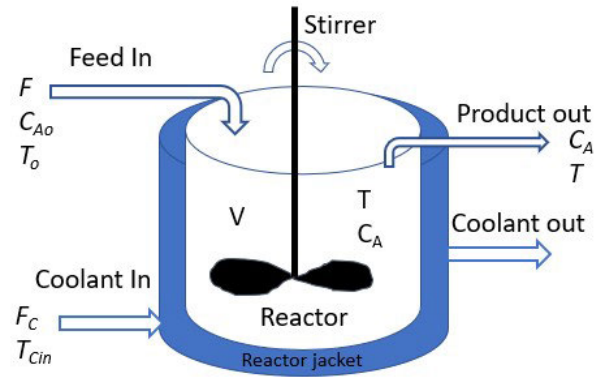
FIGURE 2. Continuous stirred tank reactor (CSTR).

is, *B*. The prediction ability of the model is used to predict the concentration, that is, *CA*. The reaction mechanism is presented as per the Equation (21). By writing the component balance and energy balance around the reactor system, the set of model ODE equations is shown in (22) to (24). The parameters of the model are presented in Table 1.

$$\frac{dC_A}{dt} = \frac{F}{V}(C_{Ao} - C_A) - k_o e^{-E/RT} C_A \tag{22}$$

$$\frac{dT}{dt} = \frac{F}{V}(T_o - T) + \frac{\Delta H_r}{\rho c_p} k_o e^{-E/RT} C_A - \frac{Q}{V \rho c_p} \tag{23}$$

$$Q = \frac{aF_c^{b+1}}{F_c + \frac{aF_c^b}{2\rho_c c_{pc}}}(T_c - T_{cin}) \tag{24}$$

1) DATA GENERATION
The data is generated for the above reactor example by perturbing the flow rate of the feed stream and coolant flow rate from the nominal steady state condition using the pseudo random binary signal (PRBS) in the frequency range of [0 0.05 w_N], where $w_N = \pi/T$ represents the Nyquist frequency. The measurement of the reactor concentration and reactor temperature are assumed to be corrupted with zero-mean Gaussian white noise sequence with a standard deviation of 0.02 and 0.5 respectively. The inlet reactant concentration of *A* (C_{Ao}), inlet temperature of the reactant (T_o) and inlet temperature of the coolant (T_{cin}) are treated as unmeasured disturbances and it is further assumed that their dynamics are governed by the following stochastic processes:

$$d_1(k) = \frac{0.1}{1 - 0.9z^{-1}} w_1(k)$$

$$d_2(k) = \frac{0.3}{1 - 0.7z^{-1}} w_2(k)$$

$$d_3(k) = \frac{0.3}{1 - 0.7z^{-1}} w_3(k) \tag{25}$$

where $\{w_1(k)\}$, $\{w_2(k)\}$ and $\{w_3(k)\}$ are white noise sequences with standard deviation of 0.5, 1 and 1, respectively. A total of 1390 observations with 7 variables were obtained which were later split equally to obtain 745 samples of training and testing data respectively. The PCA and KPCA models were

TABLE 1. Exothermic reaction in CSTR:Model parameters.

Notation	Parameter	Nominal value
k_o	First order reaction rate constant	$10^{10}\ (min^{-1})$
C_{Ao}	Inlet concentration reactant 'A'	$2.0\ (kmol/m^3)$
F	Steady state flow rate of 'A'	$1.0\ (m^3/min)$
ρ	Density of the reactant 'A'	$1.0^6\ (g/m^3)$
C_p	Specific heat capacity of reactant 'A'	$1.0\ (cal/g.^0C)$
ΔH_r	Heat of reaction	$-130 \times 10^6\ (cal/kmol)$
ρ_c	Density of the collant	$1.0^6\ (g/m^3)$
C_{pc}	Specific heat capacity of the coolant	$1.0\ (cal/g.^0C)$
V	Volume of the reactor	$1.0\ (m^3)$
F_c	Steady state coolant flow rate	$15\ (m^3/min)$
T	Reactor temperature	$393.54\ (k)$
C_A	Reactor concentration of 'A'	$0.265\ (kmol/m^3)$
T_{cin}	Inlet temperature of the coolant	$365\ (k)$
T_o	Inlet temperature of the reactant 'A'	$323\ (k)$
E/R		$8330\ (k^{-1})$
b		0.5
a		1.678×10^6

TABLE 2. Simulated faults in CSTR process.

Fault number	Description	Variable	Type of fault
1	Large step change	Reactor Temperature	Bias
2	Small step change	Reactor Temperature	Bias
3	Multiple large step change	Reactor concentration	Intermittent
4	Multiple small step change	Reactor concentration	Intermittent
5	Ramp change	Reactor Temperature	Drift
6	Ramp change	Reactor concentration	Drift

developed with four optimum PCs retained using the CPV approach. For the KD computation between two residuals, 50 was the size of the selected moving or sliding window. The monitoring of different sensor faults using the proposed KPCA-KD based approach is presented in next section.

2) MONITORING RESULTS

This section provides the confirmation of the proposed KPCA-KD fault detection approach in successfully monitoring the sensor faults in CSTR process. The table 2 describes different simulated fault scenarios for CSTR case study. The three categories of faults included in this study are bias, intermittent and drift faults. For bias and intermittent faults, two different magnitude of faults were simulated including faults with large step changes and small step changes. The three fault scenarios are presented in detail below:

3) CASE 1: FAULT 2

The fault 2 is a bias fault in which a step change of small magnitude is introduced in the reactor temperature variable between sampling time instant 350 and the end of the testing data. The monitoring results for PCA-T^2, PCA-SPE and PCA-KD based schemes for monitoring this fault are shown in Figure 3 while those of KPCA-T^2, KPCA-SPE

and KPCA-KD based schemes are shown in Figure 4. From the result plots, it can be observed that the PCA-T^2 and KPCA-T^2 based methods completely fail in detection of the fault since the T^2 statistical time evolution is below reference threshold in the region of fault. The PCA-SPE and KPCA-SPE based methods show improved detection but few missed detections may be observed in the faulty region. Even though the performance of the PCA-KD scheme is better than T^2 and SPE based schemes, it is unable to clearly detect the fault as per the expectations. In contrast, the KPCA-KD FD scheme demonstrates a clear detection of the bias fault without any missed detections or false alarms. The KPCA-KD scheme demonstrates the advantage of precisely detecting a small magnitude fault better than the other methods.

4) CASE 2: FAULT 4

This case presents the ability of the KPCA-KD scheme in monitoring an intermittent fault. This fault is introduced in the reactor concentration variable between sampling time instants [200,300] and [500,600] respectively. Figure 5 presents the result plots of PCA-based fault indicators and Figure 6 presents result plots of KPCA-based fault indicators in monitoring the intermittent fault. The PCA-T^2 as well as PCA-SPE based schemes have very
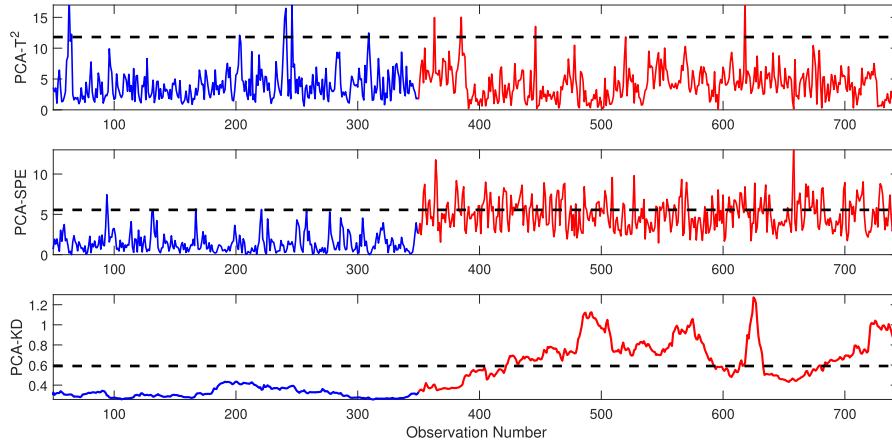
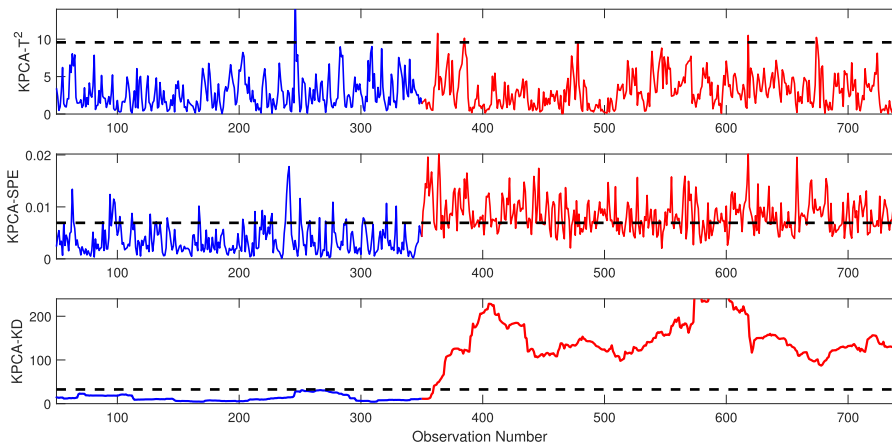**FIGURE 3.** PCA based monitoring plots using T^2, SPE and KD indicators in monitoring fault 2.

FIGURE 4. KPCA based monitoring plots using T^2, SPE and KD indicators in monitoring fault 2.
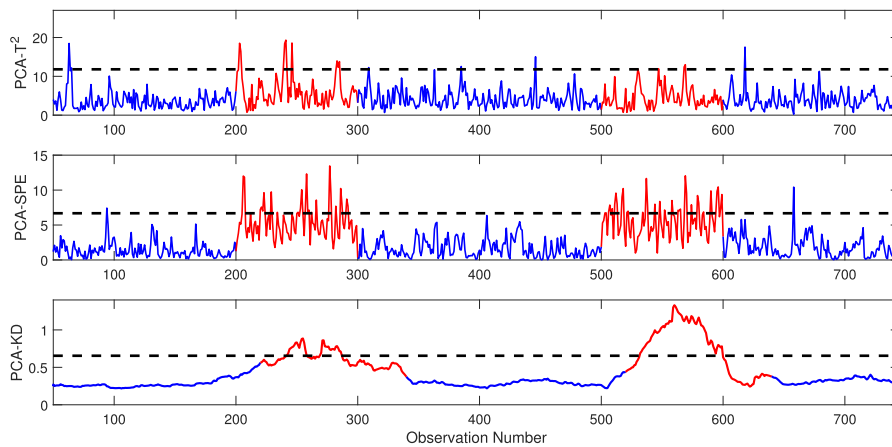

FIGURE 5. PCA based monitoring plots using T^2, SPE and KD indicators for fault 4.

poor detection performance because the fault statistics do not exceed the threshold clearly in the fault region. The PCA-KD scheme performs slightly better but few missed detections in sampling instants 100 and 200 reduce the fault

detection rate. From the result plot Figure 6, the KPCA-T^2 scheme completely fails in the detection of the fault. While the KPCA-SPE scheme fares slightly better in comparison, it is unable to clearly detect the fault in sampling time instants
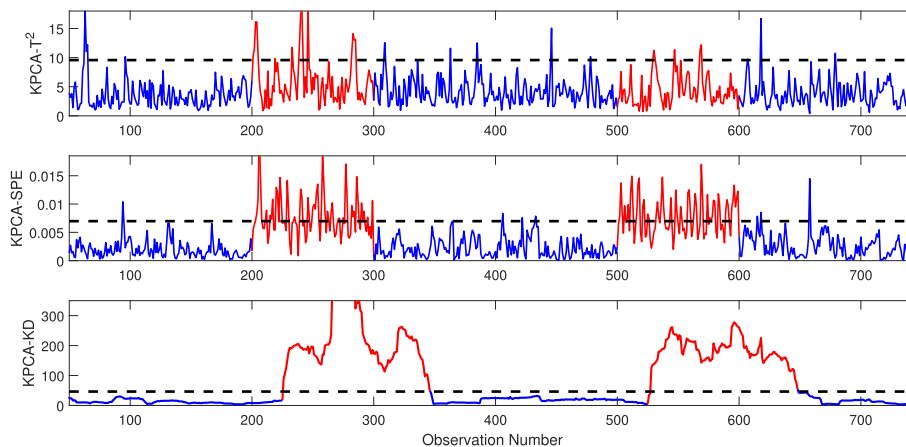
FIGURE 6. KPCA based monitoring plots using T^2, SPE and KD indicators for fault 4.
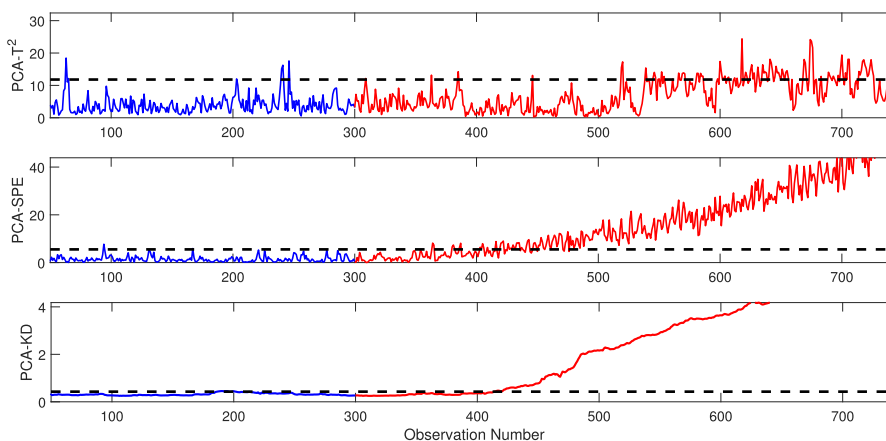

FIGURE 7. PCA based monitoring plots using T^2, SPE and KD indicators for fault 5.

[200,300] and [500,600] thereby reducing the detection rate. Among all the methods, the KPCA-KD FD scheme performs very well with precise detection in faulty region. The proposed FD scheme ensures that there are no missed detections in the region of the fault and no false alarms in the fault-free region.

5) CASE 3: FAULT 5

The considered fault is sensor drift or aging fault which is introduced in the reactor temperature variable between sampling time instant 300 and the end of the testing data. The monitoring results for PCA-T^2, PCA-SPE and PCA-KD based schemes in monitoring this fault are shown in Figure 7 while that of KPCA-T^2, KPCA-SPE and KPCA-KD based schemes are shown in Figure 8. As seen in the result plots, the T^2 indicator for both the PCA and KPCA schemes fails to recognize the presence of drift fault. This fault is introduced at sampling time instant 300 and the T^2 indicator detects it very late only after sampling time instant 550 and that too, with very less precision. The SPE indicator fares slightly better with improved detection of drift fault, however, with a small amount of delay. While PCA-SPE scheme detects the fault only after a sampling time of 450, the KPCA-SPE

scheme detects after sampling time of 400. The KD statistic detects the fault with a clear and smooth transition compared to the traditional statistical indicators. The PCA-KD scheme detects the fault at a sampling time instant of 410. In contrast, the KPCA-KD FD scheme detects the fault with less delay and at sampling time instant 375, thus providing a clear advantage.

The monitoring results for different fault scenarios evaluated using FDR and FAR can be observed in Table 3 and Table 4. It is observed that the faults 1 and 3 are of larger magnitude and hence, all monitoring methods can easily detect the fault with good FDR values. For faults with smaller magnitude, as observed in faults 2 and 4, the traditional statistical indicators fail to detect owing to the limitations that were discussed in section 1. In contrast, the proposed KPCA-KD FD scheme performs far better than the other methods, thus having a high FDR and lower FAR value as observed in the result table. Even for the faults 5 and 6, which are sensor drift faults, the performance of the KPCA-KD FD scheme is better than other methods. The KD statistical indicator is computed in a moving window involving segment-by-segment comparison, which enables it to capture sensitive details from the process data and enhance the detection of faults.
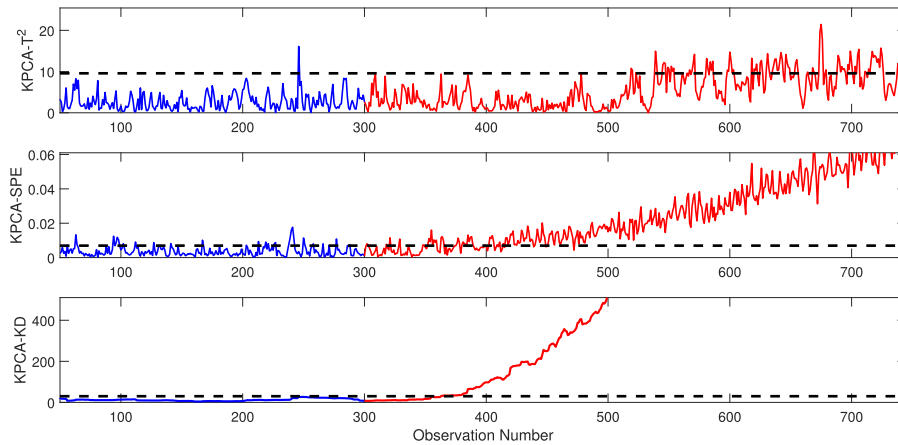
FIGURE 8. KPCA based monitoring plots using T^2, SPE and KD indicators for fault 5.

TABLE 3. Fault detection rates (in percentage) for CSTR process.

Faults	PCA-T^2	PCA-SPE	PCA-KD	KPCA-T^2	KPCA-SPE	KPCA-KD
F1	6.29	96.23	97.72	11.62	98.50	99.25
F2	4.27	36.47	65.56	5.89	48.35	96.72
F3	22.45	98.50	100.00	34.50	99.50	100.00
F4	7.00	25.38	57.50	11.64	49.50	100.00
F5	22.40	71.03	74.50	31.91	75.83	87.65
F6	30.45	70.88	79.20	40.25	75.45	85.36

TABLE 4. False alarm rates (in percentage) for CSTR process.

Faults	PCA-T^2	PCA-SPE	PCA-KD	KPCA-T^2	KPCA-SPE	KPCA-KD
F1	2.00	0.29	0.00	3.52	1.31	0.00
F2	2.29	1.88	0.00	4.57	1.75	0.00
F3	1.28	1.13	0.00	2.12	1.56	0.00
F4	1.47	1.38	0.00	2.94	2.47	0.00
F5	2.40	0.88	0.00	5.00	1.75	0.00
F6	2.40	0.75	0.00	4.20	1.45	0.00

B. TENNESSEE EASTMAN PROCESS

The Tennessee Eastman (TE) process problem has been regarded as a platform for multifarious process control and fault detection tasks in the last few decades. Most researchers in the multivariate process monitoring domain validated their newly proposed fault detection strategy on this benchmark process [53], [54] [55]. Flow diagram of the benchmark TE process is shown in Figure 9. In this process, two products and one by-product are generated by four gaseous reactants. It comprises 22 process measurements (XMEAS 1 to XMEAS 22), 19 composition measurements (XMEAS 23 to XMEAS 41) and 12 manipulated variables (XMV 42 to XMV 52) thus, comprising a total of 52 variables. The process flowsheet has 21 real fault scenarios (IDV(1) to IDV (21)) and it includes bias, drift, intermittent and valve stuck related abnormalities. Post normalization of training data, PCA and KPCA models were developed with 39 and 34 optimum PCs retained using the CPV approach. A moving window of 50 is selected for the PCA-KD and KPCA-KD based strategies. Three fault scenarios, namely, IDV(3), IDV(9) and IDV(15) have been excluded in this work since they give very poor FDR values [56]. The fault scenarios

comprising of IDV(3), IDV(9), and IDV(15) were excluded in the study because they yield small FDR values [56]. In all the fault scenarios presented in the TE process data-sheet, faults were introduced between sampling time instants 160 and 960 of the testing data.

To confirm effectiveness of the KPCA-KD based FD scheme, its performance was validated on different fault scenarios of the benchmark TE process. The performance of the proposed method and its contemporary methods in monitoring various faults through FDR and FAR indices is tabulated in Tables 5 and 6. The results table indicates the superiority of the KPCA-KD method since it has a good FDR and minimum FAR value for all fault scenarios in comparison with the contemporary methods. This is mainly because the KD metric can capture sensitive details in the process data. For large magnitude faults such as IDV(1), IDV(2), IDV(6), IDV(7), IDV(8), IDV(14), IDV(17), and IDV(18), all the FD methods can easily detect the faults. For fault scenarios IDV(10), IDV(19) and IDV(21), the performance of all FD strategies is reasonably good and even in these cases, the proposed method has a good advantage in terms of FDR and FAR. From the monitoring results, the proposed KPCA-KD
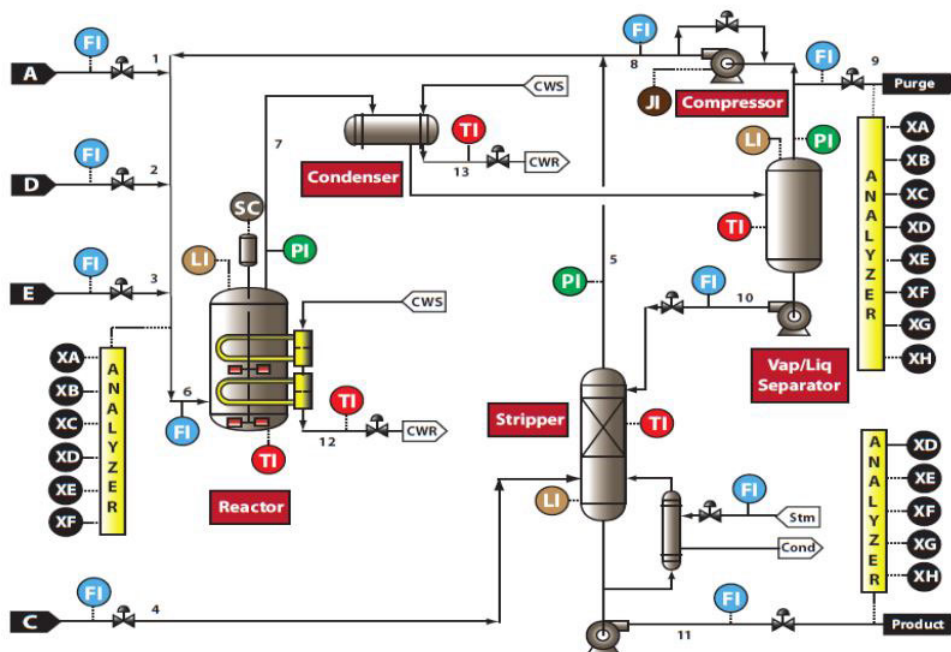
FIGURE 9. TE process flowsheet [52].

TABLE 5. TE Process: Fault detection rates (in percentage).

Faults	PCA-T^2	PCA-SPE	PCA-KD	KPCA-T^2	KPCA-SPE	KPCA-KD
IDV(1)	98.83	99.15	99.38	99.38	99.65	99.50
IDV(2)	98.50	99.00	98.50	98.62	99.15	99.50
IDV(4)	75.56	97.50	98.75	83.38	98.50	99.25
IDV(5)	28.80	48.50	38.43	34.38	49.63	98.12
IDV(6)	98.25	99.75	99.50	99.25	100.00	100.00
IDV(7)	99.50	98.88	99.50	100.00	100.00	100.00
IDV(8)	97.26	93.64	96.50	97.25	97.50	98.85
IDV(10)	54.36	65.83	67.85	55.64	67.75	69.25
IDV(11)	64.59	70.32	79.65	70.75	75.50	94.23
IDV(12)	98.75	95.00	98.75	97.85	99.00	99.75
IDV(13)	95.39	95.64	96.25	95.75	96.00	96.75
IDV(14)	100	90.40	100.00	99.50	100.00	100.00
IDV(16)	37.41	66.46	65.45	36.45	67.35	70.27
IDV(17)	91.15	95.63	96.63	88.57	96.52	97.57
IDV(18)	90.65	90.25	90.25	90.25	90.87	90.80
IDV(19)	15.84	61.15	50.12	33.75	55.72	58.32
IDV(20)	52.49	70.45	88.75	55.37	72.37	90.62
IDV(21)	49.43	57.61	44.13	47.00	56.75	57.85

scheme has superior advantages for three scenarios: IDV(5), IDV(11) and IDV(20) as compared to the other methods. For providing clarity to the reader, the monitoring performance of two fault scenarios is presented in detail. The fault scenarios considered are IDV(5) which is a step anomaly in the condenser cooling water inlet temperature and IDV(11) which is a random variation fault in the reactor cooling water inlet temperature.

The result plots of the PCA and KPCA based methods for monitoring IDV(5) fault are presented in Figure 10 and Figure 11 respectively. The PCA-T^2, PCA-SPE, PCA-KD,

KPCA-T^2 and KPCA-SPE based methods detects the fault between samples 160 and 380. After a sampling time instant 380, these methods are unable to detect the fault which eventually results in a poor detection rate. In contrast, the KPCA-KD scheme detects the fault clearly in sampling time instants 160 to 960, thus having a high FDR in comparison to other FD schemes. It can be observed that the KD indicator exceeds the reference threshold correctly in the faulty region. The performance of PCA and KPCA based methods in monitoring IDV(11) fault is presented in Figure 12 and Figure 13 respectively. The PCA-T^2, PCA-SPE, PCA-KD,

TABLE 6. TE Process: False alarm rates(in percentage).

| Faults	PCA-T^2	PCA-SPE	PCA-KD	KPCA-T^2	KPCA-SPE	KPCA-KD
IDV(1)	7.59	13.29	0.00	2.87	9.35	0.00
IDV(2)	6.96	6.33	0.00	1.85	9.35	0.00
IDV(4)	3.80	6.33	0.00	8.45	9.38	0.00
IDV(5)	3.68	6.01	1.25	9.37	9.87	0.00
IDV(6)	0.00	2.53	0.00	0.50	4.13	0.00
IDV(7)	0.00	1.90	0.00	1.87	1.87	0.00
IDV(8)	1.00	5.06	0.00	0.00	5.25	0.00
IDV(10)	2.53	5.06	0.00	1.45	5.32	0.00
IDV(11)	2.27	4.43	0.00	6.37	7.50	0.00
IDV(12)	3.16	3.80	0.00	2.37	6.50	0.00
IDV(13)	1.64	1.90	0.00	2.13	4.65	0.00
IDV(14)	2.53	5.70	0.00	1.25	3.13	0.00
IDV(16)	8.86	9.49	0.00	4.63	6.38	0.00
IDV(17)	2.53	5.70	0.00	0.63	7.50	0.00
IDV(18)	3.16	4.43	0.00	5.75	8.50	0.00
IDV(19)	0.00	6.33	0.00	2.25	4.63	0.00
IDV(20)	1.87	3.16	0.00	3.13	4.54	0.00
IDV(21)	6.33	7.59	5.13	2.23	11.35	2.50
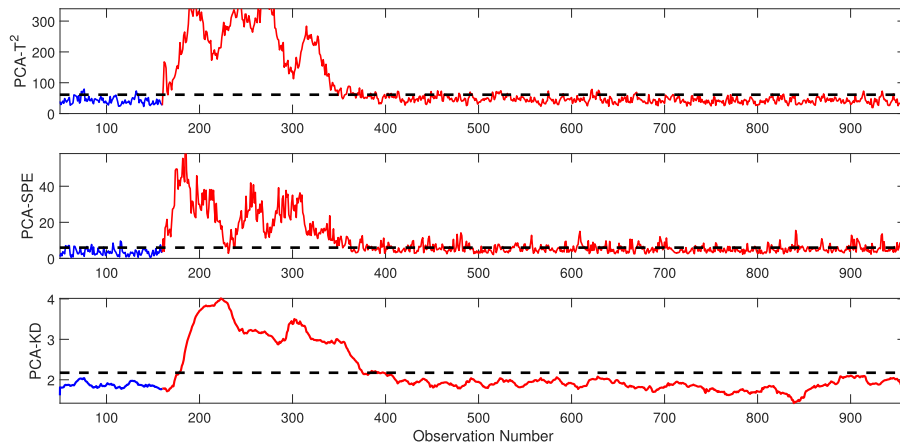
FIGURE 10. PCA based monitoring plots using T^2, SPE and KD indicators for IDV(5).
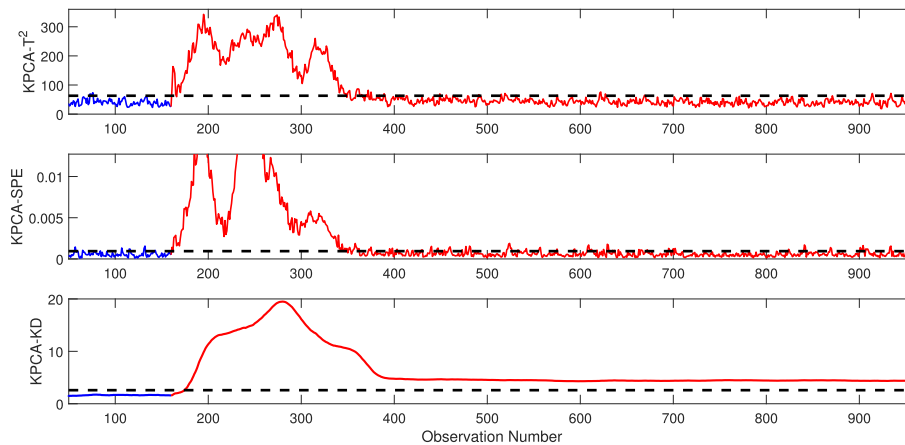

FIGURE 11. KPCA based monitoring plots using T^2, SPE and KD indicators for IDV(5).

KPCA-T^2 and KPCA-SPE based methods are able to partially detect the fault because the value of FDR is less vowing to the missed detections in the fault region. In contrast, the proposed KPCA-KD FD scheme detects the fault precisely with a high FDR value, thus, it has a superior advantage over other methods. Hence, it can be inferred that the KPCA-KD
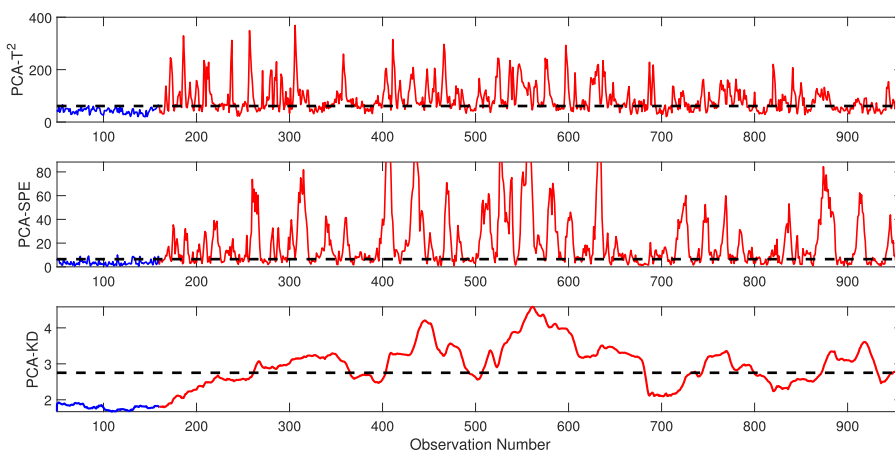
FIGURE 12. PCA based monitoring plots using T^2, SPE and KD indicators for IDV(11).
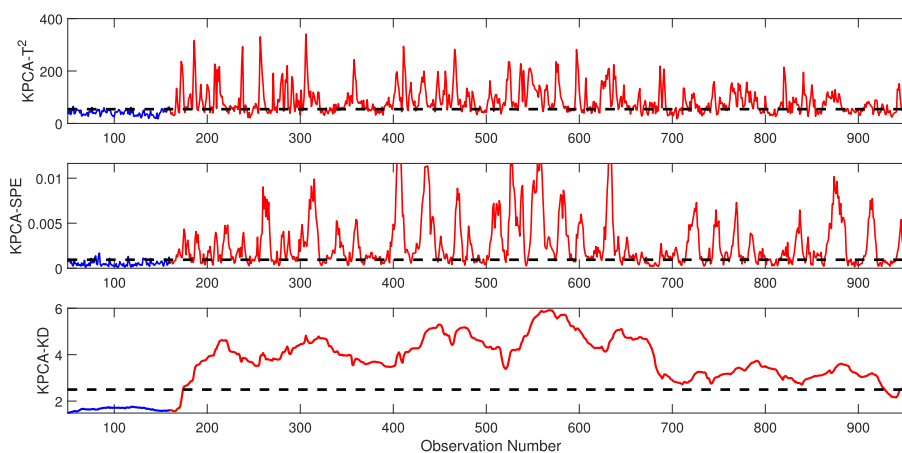

FIGURE 13. KPCA based monitoring plots using T^2, SPE and KD indicators for IDV(11).

FD scheme provides better detection of faults with superior FDR values and minimum FAR values, thus, satisfying the properties of a good fault detection scheme.

C. EXPERIMENTAL DISTILLATION COLUMN PROCESS

An experimental distillation column (DC) process is considered in this section to confirm the effectiveness of the proposed KPCA-KD based FD scheme. A distillation column is an energy-consuming unit operation in the process industry. The experimental distillation column considered in this study is housed in the Department of Chemical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. A pictorial view of the bubble cap distillation column can be observed in Figure 14. The construction details of the distillation column can be found in [57].

1) DATA GENERATION

The data are generated by sequentially perturbing the reflux and feed flow. The distillation column was brought to a nominal operating condition and then, the feed flow was perturbed with a magnitude of 50, while maintaining the reflux flow

constant at nominal condition. Again, the distillation column is brought to the nominal condition followed by perturbing the column with reflux flow with a magnitude of 40 and keeping the flow rate constant. The resultant change in the product quality (output) was recorded as x_D along with six temperatures to monitor the condition of the column. In model development stage, along with the flow rate and feed flow, six temperatures with 2048 observations were included in the input matrix. Next, the generated data were split equally into training and testing data groups. To have a fair comparison, 6 optimum PCs are selected for both PCA as well as KPCA based fault detection strategies during the model development stage.

2) MONITORING RESULTS

The efficacy of the proposed FD scheme is assessed by its ability to monitor bias, intermittent and drift faults. First, a sustained bias was introduced in the temperature variable 2 in between sampling time instants 400 and 1024 of the testing data. The results of PCA and KPCA based fault indicators for monitoring this fault are presented through Figures 15 and 16. The PCA-T^2 and PCA-SPE based fault
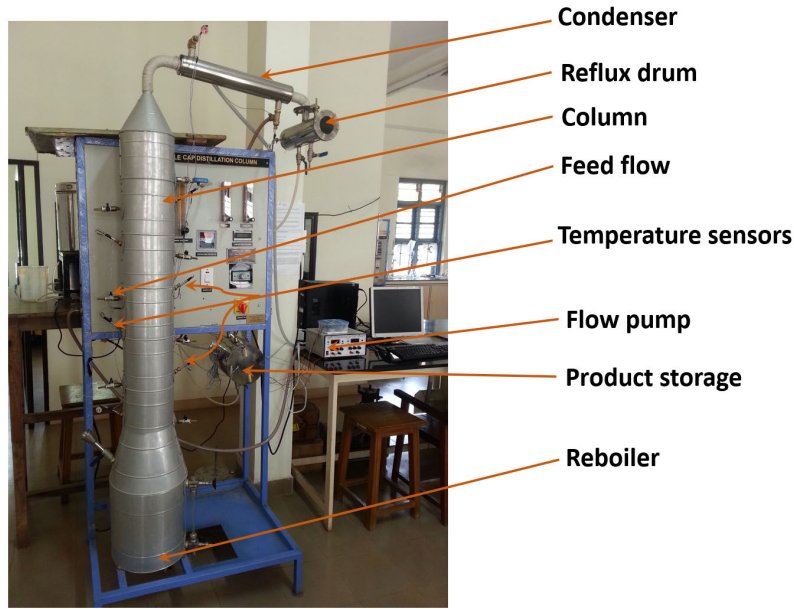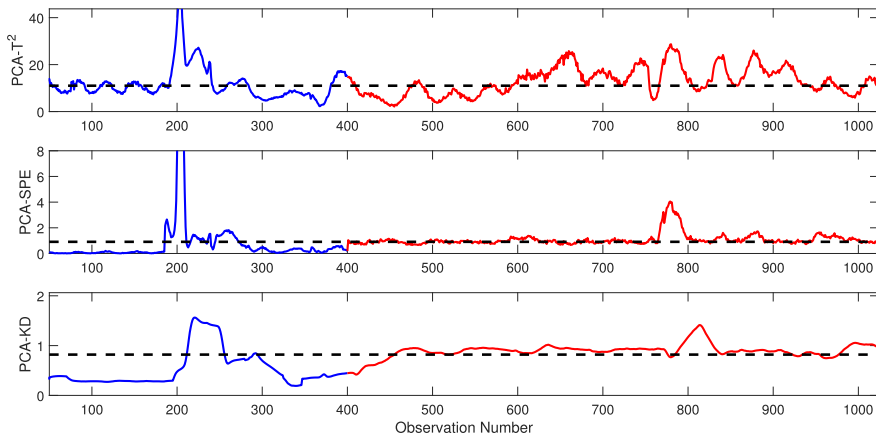
FIGURE 14. A schematic of distillation column set-up.

FIGURE 15. PCA based monitoring plots using T^2, SPE and KD indicators for bias fault.
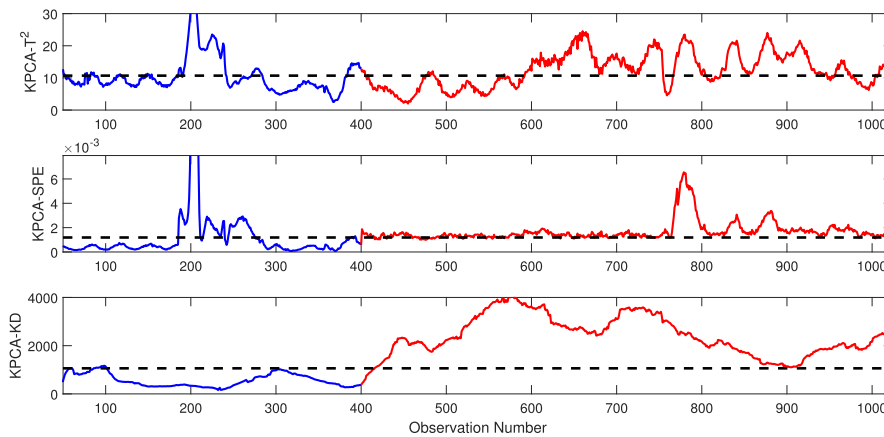

FIGURE 16. KPCA based monitoring plots using T^2, SPE and KD indicators for bias fault.

indicators have few false alarms and are unable to precisely detect the fault in the fault region. The PCA-KD performs slightly better, but few missed detections are visible in the

fault region. Even in the case of KPCA based FD strategy, the T^2 and SPE indicators are unable to demonstrate a good detection performance of the bias fault. In contrast,
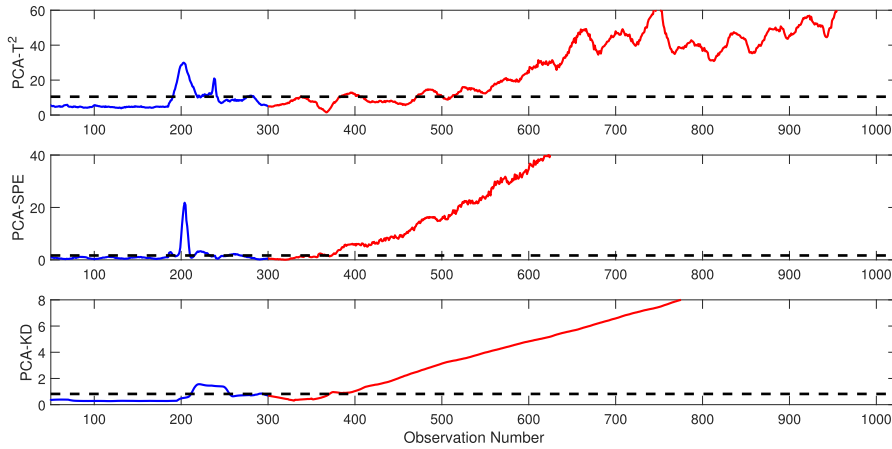
FIGURE 17. PCA based monitoring plots using T^2, SPE and KD indicators for drift fault.
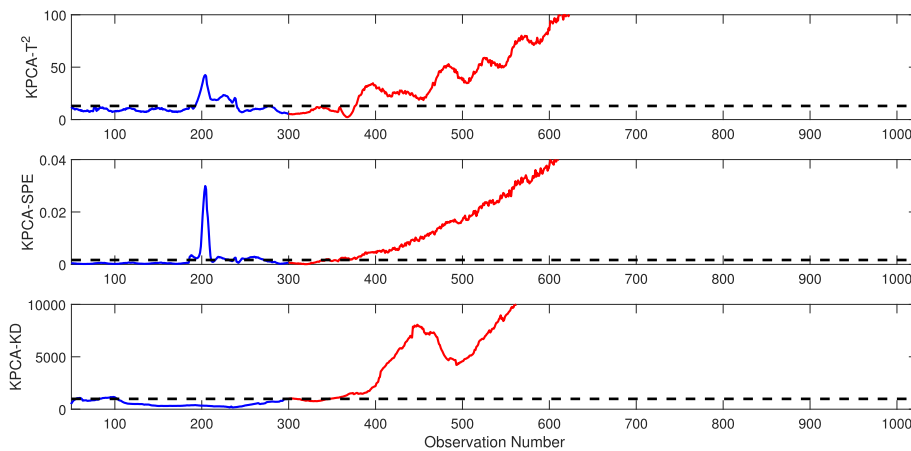

FIGURE 18. KPCA based monitoring plots using T^2, SPE and KD indicators for drift fault.

TABLE 7. Experimental DC process: Fault detection rates (in percentage).

Faults	PCA-T^2	PCA-SPE	PCA-KD	KPCA-T^2	KPCA-SPE	KPCA-KD
Bias	24.88	59.64	85.29	47.51	89.40	98.25
Intermittent	46.24	80.50	83.59	44.25	90.50	97.45
Drift	72.80	88.25	90.33	88.25	92.82	95.25

TABLE 8. Experimental DC process: False alarm rates(in percentage).

Faults	PCA-T^2	PCA-SPE	PCA-KD	KPCA-T^2	KPCA-SPE	KPCA-KD
Bias	12.25	18.75	12.25	18.50	18.75	6.24
Intermittent	22.16	12.89	3.45	21.52	10.33	3.12
Drift	11.33	18.50	16.33	14.33	20.15	7.25

the KPCA-KD FD scheme can clearly detect the bias fault with a good detection rate. Hence, it can be inferred that the KPCA-KD scheme outperforms other schemes with improved detection performance. An intermittent fault are introduced in the temperature variable 4 at sampling time instants [250,350] and [750,850] of testing data and even in this case, the KPCA-KD scheme was found to have enhanced detection performance.

Next, a drift fault was introduced in the temperature variable 3 between sampling time instants 300 and 1024 of the testing data. The monitoring results of the PCA and KPCA-based indicators for monitoring this fault are presented in Figure 17 and Figure 18 respectively. From

the monitoring plots, it was observed that all monitoring strategies were able to detect the drift fault. The PCA-T^2 and KPCA-T^2 based schemes detect the fault with large detection delay in comparison to PCA-SPE, PCA-KD, KPCA-SPE and KPCA-KD based FD schemes. The proposed KPCA-KD FD scheme carries a small advantage because it shows early detection of drift fault in comparison with other schemes. The FDR and FAR of the proposed KPCA-KD FD strategy and corresponding FD strategies are presented in Table 7 and Table 8 respectively. The tables clearly indicate that the KPCA-KD strategy is better equipped for handling the sensor faults in experimental DC process with high FDR values and less FAR values.

VI. CONCLUSION

In this study, a new and effective KPCA-KD based FD scheme was developed for improved and efficient detection of sensor faults. The addressed fault detection problem modeled the data using the KPCA framework and utilized the capability of the KD statistical indicator for fault detection. The KD statistical indicator evaluated the distance between the KPCA residuals of normally operating data and the residuals of the new online data. The KDE approach was used to compute the reference threshold for KD statistical indicator. The efficacy of the proposed KPCA-KD based FD strategy was illustrated using the CSTR process, benchmark TE process and experimental distillation column process. The KPCA-KD fault detection was contrasted against PCA-T^2, PCA-SPE, PCA-KD, KPCA-T^2 and KPCA-SPE based FD schemes. Since the KD indicator was computed in a moving window involving segment-by-segment comparison, sensitive details from the process were captured easily which enabled the enhanced detection of faults. In the case of the TE process, the proposed FD strategy was able to successfully detect the IDV(5) fault which is generally not detected clearly by traditional PCA and KPCA based FD schemes.

Overall, the KD metric was able to precisely detect small magnitude faults with a smooth detection profile. Hence, it can be concluded that the KPCA-KD scheme clearly emerges as a better choice for formulating the FD strategy when compared to traditional fault indicators. As a part of future work, we plan to amalgamate wavelet functions with Kernel PCA scheme to have a novel kernel multi-scale PCA based fault detection strategy where the multi-scale Kernel PCA strategy will be used as modeling framework and the KD metric as fault indicator. Since the process data comes with large amount of noise, one possible way to reduce the effect of noise is by multi-scale representation of the data using wavelet functions. Additionally, it is also planned to perform the fault diagnosis to determine the variable or set of variables that were responsible for the fault. This can be achieved by appropriately evaluating the residuals from the KPCA model using contribution plots.

REFERENCES

[1] X. Han, J. Jiang, A. Xu, X. Huang, C. Pei, and Y. Sun, "Fault detection of pneumatic control valves based on canonical variate analysis," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13603–13615, Jun. 2021.

[2] Z. Lou, Y. Wang, S. Lu, and P. Sun, "Process monitoring using a novel robust PCA scheme," *Ind. Eng. Chem. Res.*, vol. 60, no. 11, pp. 4397–4404, Mar. 2021.

[3] C. Zhao and H. Sun, "Dynamic distributed monitoring strategy for large-scale nonstationary processes subject to frequently varying conditions under closed-loop control," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4749–4758, Jun. 2019.

[4] X. Gao and Y. A. W. Shardt, "Dynamic system modelling and process monitoring based on long-term dependency slow feature analysis," *J. Process Control*, vol. 105, pp. 27–47, Sep. 2021.

[5] W. Li, M. Peng, and Q. Wang, "Fault identification in PCA method during sensor condition monitoring in a nuclear power plant," *Ann. Nucl. Energy*, vol. 121, pp. 135–145, Nov. 2018.

[6] A. Kumar, A. Bhattacharya, and J. Flores-Cerrillo, "Data-driven process monitoring and fault analysis of reformer units in hydrogen plants: Industrial application and perspectives," *Comput. Chem. Eng.*, vol. 136, May 2020, Art. no. 106756.

[7] P. Subbaraj and B. Kannapiran, "Fault detection and diagnosis of pneumatic valve using adaptive neuro-fuzzy inference system approach," *Appl. Soft Comput.*, vol. 19, pp. 362–371, Jun. 2014.

[8] M. Seera, C. P. Lim, C. K. Loo, and H. Singh, "A modified fuzzy min–max neural network for data clustering and its application to power quality monitoring," *Appl. Soft Comput.*, vol. 28, pp. 19–29, Mar. 2015.

[9] A. Rodríguez-Ramos, A. J. da Silva Neto, and O. Llanes-Santiago, "An approach to fault diagnosis with online detection of novel faults using fuzzy clustering tools," *Expert Syst. Appl.*, vol. 113, pp. 200–212, Dec. 2018.

[10] Q. Jiang, X. Yan, and B. Huang, "Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes," *Ind. Eng. Chem. Res.*, vol. 58, no. 29, pp. 12899–12912, Jul. 2019.

[11] N. M. Nor, C. R. C. Hassan, and M. A. Hussain, "A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems," *Rev. Chem. Eng.*, vol. 36, no. 4, pp. 513–553, May 2020.

[12] A. Bakdi, W. Bounoua, S. Mekhilef, and L. M. Halabi, "Nonparametric Kullback-divergence-PCA for intelligent mismatch detection and power quality monitoring in grid-connected rooftop PV," *Energy*, vol. 189, Dec. 2019, Art. no. 116366.

[13] J. Li and X. Yan, "Process monitoring using principal component analysis and stacked autoencoder for linear and nonlinear coexisting industrial processes," *J. Taiwan Inst. Chem. Eng.*, vol. 112, pp. 322–329, Jul. 2020.

[14] M. Madakyaru, F. Harrou, and Y. Sun, "Improved data-based fault detection strategy and application to distillation columns," *Process Saf. Environ. Protection*, vol. 107, pp. 22–34, Apr. 2017.

[15] H. Liu, H. Zhang, Y. Zhang, F. Zhang, and M. Huang, "Modeling of wastewater treatment processes using dynamic Bayesian networks based on fuzzy PLS," *IEEE Access*, vol. 8, pp. 92129–92140, 2020.

[16] Z. Chen, S. X. Ding, T. Peng, C. Yang, and W. Gui, "Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1559–1567, Feb. 2018.

[17] C. Shang, B. Huang, F. Yang, and D. Huang, "Slow feature analysis for monitoring and diagnosis of control performance," *J. Process Control*, vol. 39, pp. 21–34, Mar. 2016.

[18] J.-M. Lee, C. K. Yoo, and I.-B. Lee, "Statistical process monitoring with independent component analysis," *J. Process Control*, vol. 14, no. 5, pp. 467–485, Aug. 2004.

[19] K. R. Kini and M. Madakyaru, "Improved process monitoring scheme using multi-scale independent component analysis," *Arabian J. Sci. Eng.*, Jun. 2021, doi: 10.1007/s13369-021-05822-1.

[20] Y. Dong and S. J. Qin, "A novel dynamic PCA algorithm for dynamic data modeling and process monitoring," *J. Process Control*, vol. 67, pp. 1–11, Jul. 2018.

[21] V. S. Yellapu, V. Vajpayee, and A. P. Tiwari, "Online fault detection and isolation in advanced heavy water reactor using multiscale principal component analysis," *IEEE Trans. Nucl. Sci.*, vol. 66, no. 7, pp. 1790–1803, Jul. 2019.

[22] C. D. Tong and X. F. Yan, "A novel decentralized process monitoring scheme using a modified multiblock PCA algorithm," *TEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1129–1138, Apr. 2017.

[23] L. M. Elshenawy, S. Yin, A. S. Naik, and S. X. Ding, "Efficient recursive principal component analysis algorithms for process monitoring," *Ind. Eng. Chem. Res.*, vol. 49, no. 1, pp. 252–259, Nov. 2009.

[24] C. Miao and Z. Lv, "Nonlinear chemical processes fault detection based on adaptive kernel principal component analysis," *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 350–358, May 2020.

[25] J.-M. Lee, C. K. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chem. Eng. Sci.*, vol. 59, no. 1, pp. 223–234, Jan. 2004.

[26] S. W. Choi, C. K. Lee, J.-M. Lee, J. H. Park, and I.-B. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemometrics Intell. Lab. Syst.*, vol. 75, no. 1, pp. 55–67, Jan. 2005.

[27] Y. Zhang, S. Li, and Y. Teng, "Dynamic processes monitoring using recursive kernel principal component analysis," *Chem. Eng. Sci.*, vol. 72, pp. 78–86, Apr. 2012.

[28] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Fault discriminant enhanced kernel principal component analysis incorporating prior fault information for monitoring nonlinear processes," *Chemometrics Intell. Lab. Syst.*, vol. 162, pp. 21–34, Mar. 2017.

[29] M. Nawaz, A. S. Maulud, H. Zabiri, H. Suleman, and L. D. Tufa, "Multiscale framework for real-time process monitoring of nonlinear chemical process systems," *Ind. Eng. Chem. Res.*, vol. 59, no. 41, pp. 18595–18606, Sep. 2020.

[30] R. Fezai, M. Mansouri, O. Taouali, M. F. Harkat, and N. Bouguila, "Online reduced kernel principal component analysis for process monitoring," *J. Process Control*, vol. 61, pp. 1–11, Jan. 2018.

[31] T. Berghout, M. Benbouzid, S. M. Muyeen, T. Bentrcia, and L.-H. Mouss, "Auto-NAHL: A neural network approach for condition-based maintenance of complex industrial systems," *IEEE Access*, vol. 9, pp. 152829–152840, 2021.

[32] H. Chen, B. Jiang, N. Lu, and J. Wu, "Incipient fault detection method using Kullback–Leibler divergence for CRH5 and some remarks," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2018, pp. 881–885.

[33] S. Wang and Y. Chen, "Sensor validation and reconstruction for building central chilling systems based on principal component analysis," *Energy Convers. Manage.*, vol. 45, no. 5, pp. 673–695, Mar. 2004.

[34] H. Chen, B. Jiang, S. X. Ding, and B. Huang, "Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives," *IEEE Trans. Intell. Transp. Syst.*, early access, Oct. 22, 2020, doi: 10.1109/TITS.2020.3029946.

[35] H. Chen, B. Jiang, and N. Lu, "A newly robust fault detection and diagnosis method for high-speed trains," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2198–2208, Jun. 2019.

[36] F. Harrou, Y. Sun, and M. Madakyaru, "Kullback–Leibler distance-based enhanced detection of incipient anomalies," *J. Loss Prevention Process Ind.*, vol. 44, pp. 73–87, Nov. 2016.

[37] F. Harrou, M. Madakyaru, and Y. Sun, "Improved nonlinear fault detection strategy based on the Hellinger distance metric: Plug flow reactor monitoring," *Energy Buildings*, vol. 143, pp. 149–161, May 2017.

[38] D. Li and S. Martinez, "High-confidence attack detection via Wasserstein-metric computations," *IEEE Control Syst. Lett.*, vol. 5, no. 2, pp. 379–384, Apr. 2021.

[39] S. Kammammettu and Z. Li, "Change point and fault detection using Kantorovich distance," *J. Process Control*, vol. 80, pp. 41–59, Aug. 2019.

[40] B. M. S. Arifin, Z. Li, and S. L. Shah, "Change point detection using the Kantorovich distance algorithm," *IFAC-PapersOnLine*, vol. 51, no. 18, pp. 708–713, 2018.

[41] K. R. Kini and M. Madakyaru, "Improved process monitoring strategy using Kantorovich distance-independent component analysis: An application to Tennessee eastman process," *IEEE Access*, vol. 8, pp. 205863–205877, 2020.

[42] B. Xiao, Y. Li, B. Sun, C. Yang, K. Huang, and H. Zhu, "Decentralized PCA modeling based on relevance and redundancy variable selection and its application to large-scale dynamic process monitoring," *Process Saf. Environ. Protection*, vol. 151, pp. 85–100, Jul. 2021.

[43] F. Bencheikh, M. F. Harkat, A. Kouadri, and A. Bensmail, "New reduced kernel PCA for fault detection and diagnosis in cement rotary kiln," *Chemometric Intell. Lab. Syst.*, vol. 204, Sep. 2020, Art. no. 104091.

[44] S. Kolouri, S. R. Park, M. Thorpe, D. Slepčev, and G. K. Rohde, "Optimal mass transport: Signal processing and machine-learning applications," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 43–59, Jul. 2017.

[45] L. Kantorovich, "On the translocation of masses," *Manage. Sci.*, vol. 5, no. 1, pp. 1–4, Oct. 1958.

[46] A. Takatsu, "Wasserstein geometry of Gaussian measures," *Osaka J. Math.*, vol. 48, no. 48, pp. 1005–1026, 2011.

[47] Q. Chen, R. J. Wynne, P. Goulding, and D. Sandoz, "The application of principal component analysis and kernel density estimation to enhance process monitoring," *Control Eng. Pract.*, vol. 8, no. 5, pp. 531–543, May 2000.

[48] Y. Li, X. Peng, and Y. Tian, "Plant-wide process monitoring strategy based on complex network and Bayesian inference-based multi-block principal component analysis," *IEEE Access*, vol. 8, pp. 199213–199226, 2020.

[49] J. Liang, "Multivariate statistical process monitoring using kernel density estimation," *Develop. Chem. Eng. Mineral Process.*, vol. 13, nos. 1–2, pp. 185–192, May 2008.

[50] J. Prakash, S. C. Patwardhan, and S. Narasimhan, "A supervisory approach to fault-tolerant control of linear multivariable systems," *Ind. Eng. Chem. Res.*, vol. 41, no. 9, pp. 2270–2281, Apr. 2002.

[51] X. Deng, P. Cai, J. Deng, Y. Cao, and Z. Song, "Primary-auxiliary statistical local kernel principal component analysis and its application to incipient fault detection of nonlinear industrial processes," *IEEE Access*, vol. 7, pp. 122204–122912, 2019.

[52] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.

[53] C. Hu, Z. Xu, X. Kong, and J. Luo, "Recursive-CPLS-based quality-relevant and process-relevant fault monitoring with application to the Tennessee eastman process," *IEEE Access*, vol. 7, pp. 128746–128757, 2019.

[54] G. Sun and J. Hou, "An improved principal component regression for quality-related process monitoring of industrial control systems," *IEEE Access*, vol. 8, pp. 34177–34186, 2020.

[55] C. Hu, Z. Xu, X. Kong, and J. Luo, "Generalized principal component analysis-based subspace decomposition of fault deviations and its application to fault reconstruction," *IEEE Access*, vol. 5, pp. 21723–21730, 2019.

[56] S. Yin, S. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *J. Process Control*, vol. 22, no. 9, pp. 1567–1581, Oct. 2012.

[57] M. Madakyaru, F. Harrou, and Y. Sun, "Monitoring distillation column systems using improved nonlinear partial least squares-based strategies," *IEEE Sensors J.*, vol. 19, no. 23, pp. 11697–11705, Dec. 2019.

K. RAMAKRISHNA KINI received the B.E. degree in electronics and communication engineering from NMAMIT, Nitte, and the M.Tech. degree in control systems and the Ph.D. degree in process monitoring from the Manipal Institute of Technology, Manipal Academy of Higher Education, India. He is currently an Assistant Professor in instrumentation and control engineering with the Manipal Institute of Technology, Manipal Academy of Higher Education. He has published five articles in peer-reviewed journals and three international conference proceedings papers. His research interests include fault detection and diagnosis, soft sensor modeling, latent variable regression modeling using wavelets, and advanced process control.

MRUNMAYEE BAPAT received the B.Tech. degree in chemical engineering from the National Institute of Technology, Karnataka. She is currently pursuing the M.Tech. degree in chemical engineering with the Manipal Institute of Technology, Manipal Academy of Higher Education, India. Her research interests include process control and fault detection and diagnosis.

MUDDU MADAKYARU received the B.E. degree in chemical engineering from R.V.C.E., the M.Tech. degree in chemical plant design from the National Institute of Technology Karnataka, India, and the Ph.D. degree in process control from the Indian Institute of Technology Bombay, Mumbai, India. He was a Postdoctoral Researcher with Texas A&M University, Doha, Qatar, for four years. He is currently a Professor with the Department of Chemical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, India. He has published more than 30 articles in peer-reviewed journals and 15 international conference proceedings papers. His research interests include advanced process control, including system identification, fault detection and diagnosis, model predictive control, and latent variable regression modeling using wavelets. He is a fellow of the Institution of Engineers, India, and a Life Member of the Indian Society for Technical Education and the Indian Society of Systems for Science and Engineering.

• • •