Manuscript Details

Manuscript number	CACE_2018_593
Title	Mixed Kernel Canonical Variate Dissimilarity Analysis for Incipient Fault Monitoring in Nonlinear Dynamic Processes
Article type	Full Length Article

Abstract

Incipient fault monitoring in large industrial plants is becoming more important, since the early detection of these faults can prevent an emergency situation. Recently, the Canonical Variate Dissimilarity Analysis (CVDA) detection method was shown to be efficient especially for processes under varying operating conditions. CVDA can be extended to nonlinear processes by introducing kernel-based learning. However, incipient fault monitoring requires kernels with both good interpolation and extrapolation abilities. Unfortunately, conventional single kernels only exhibit one ability or the other, but not both. To overcome this drawback, a Mixed Kernel CVDA method is presented in this study for incipient fault monitoring in nonlinear dynamic processes. Due to the use of mixed kernels, both enhanced detection sensitivity and a better depiction of the growing fault severity in the monitoring charts were achieved. The proposed method remains effective for handling the nonlinear, non-Gaussian, and dynamic nature of the data all at once.

Keywords	fault detection; canonical variate analysis; global kernel; local kernel; kernel density estimation
Manuscript category	Process dynamics, control and monitoring

Corresponding Author	Karl Ezra Pilario
Corresponding Author's Institution	Cranfield University
Order of Authors	Karl Ezra Pilario, Yi Cao, Mahmood Shafiee

Suggested reviewers	Zhiqiang Ge, Steven Ding, Leo Chiang, Geert Gins, Furong Gao

Submission Files Included in this PDF

File Name [File Type]

Cover letter.docx [Cover Letter]

Highlights.docx [Highlights]

Abstract.docx [Abstract]

TXT_CACE2018.pdf [Manuscript File]

FIG_1.pdf [Figure]

FIG_2.pdf [Figure]

FIG_3.pdf [Figure]

FIG_4.pdf [Figure]

FIG_5.pdf [Figure]

FIG_6.pdf [Figure]

FIG_7.pdf [Figure]

Submission Files Not Included in this PDF

File Name [File Type]

elsarticle.dtx [LaTeX Source File]

library.bib [LaTeX Source File]

TXT_CACE2018.tex [LaTeX Source File]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Dear Editor,

On behalf of my co-authors, I would like to submit the attached manuscript entitled: “Mixed Kernel Canonical Variate Dissimilarity Analysis for Incipient Fault Monitoring in Nonlinear Dynamic Processes” to be considered for publication in the Computers & Chemical Engineering journal.

We first highlight these issues in the current industrial process monitoring literature:

(a) sensitive detection of incipient faults; and
(b) the use of kernel-based techniques for nonlinear process monitoring.

Many studies have covered these topics separately. However, there is no existing technique that simultaneously addresses these two issues together. We believe that the reason behind this is the proper choice of kernel.

Our study, hence, emphasizes the importance of using mixed kernels for incipient fault monitoring. Canonical variate dissimilarity analysis (CVDA) is a recent technique developed for incipient fault detection. In our paper, we present Mixed Kernel CVDA (MKCVDA), which is an extension of CVDA that incorporates mixed kernels. Using a continuous stirred-tank reactor (CSTR) case study, we have demonstrated the superior performance of MKCVDA over the linear CVDA and single Kernel CVDA for incipient fault monitoring in nonlinear dynamic processes.

We declare that this work has not been published before and is not being considered for publication elsewhere.

We look forward to hearing from you soon.

Sincerely,
Karl Ezra S. Pilario

Highlights:

- A method consisting of kernel PCA and canonical variate dissimilarity analysis is proposed.
- Current kernel-based process monitoring methods only use single kernels.
- Incipient fault growth is depicted more accurately using a mixture of Gaussian RBF and linear kernel.
- The grid search method is used to optimize kernel parameters.
- The dissimilarity index achieves the earliest detection time compared to conventional statistical indices.

Mixed Kernel Canonical Variate Dissimilarity Analysis for Incipient Fault Monitoring in Nonlinear Dynamic Processes

Abstract

Incipient fault monitoring in large industrial plants is becoming more important, since the early detection of these faults can prevent an emergency situation. Recently, the Canonical Variate Dissimilarity Analysis (CVDA) detection method was shown to be efficient especially for processes under varying operating conditions. CVDA can be extended to nonlinear processes by introducing kernel-based learning. However, incipient fault monitoring requires kernels with both good interpolation and extrapolation abilities. Unfortunately, conventional single kernels only exhibit one ability or the other, but not both. To overcome this drawback, a Mixed Kernel CVDA method is presented in this study for incipient fault monitoring in nonlinear dynamic processes. Due to the use of mixed kernels, both enhanced detection sensitivity and a better depiction of the growing fault severity in the monitoring charts were achieved. The proposed method remains effective for handling the nonlinear, non-Gaussian, and dynamic nature of the data all at once.

Keywords: Fault detection, canonical variate analysis, global kernel, local kernel, kernel density estimation.

Mixed Kernel Canonical Variate Dissimilarity Analysis for Incipient Fault Monitoring in Nonlinear Dynamic Processes

Karl Ezra S. Pilario[a,b,*], Yi Cao[c], Mahmood Shafiee[a]

[a]*School of Water, Energy, and Environment, Cranfield University, Cranfield, United Kingdom*
[b]*Department of Chemical Engineering, University of the Philippines Diliman, Republic of the Philippines*
[c]*Department of Chemical and Biological Engineering, Zhejiang University, People's Republic of China*

Abstract

Incipient fault monitoring in large industrial plants is becoming more important, since the early detection of these faults can prevent an emergency situation. Recently, the Canonical Variate Dissimilarity Analysis (CVDA) detection method was shown to be efficient especially for processes under varying operating conditions. CVDA can be extended to nonlinear processes by introducing kernel-based learning. However, incipient fault monitoring requires kernels with both good interpolation and extrapolation abilities. Unfortunately, conventional single kernels only exhibit one ability or the other, but not both. To overcome this drawback, a Mixed Kernel CVDA method is presented in this study for incipient fault monitoring in nonlinear dynamic processes. Due to the use of mixed kernels, both enhanced detection sensitivity and a better depiction of the growing fault severity in the monitoring charts were achieved. The proposed method remains effective for handling the nonlinear, non-Gaussian, and dynamic nature of the data all at once.

Keywords: Fault detection, canonical variate analysis, global kernel, local kernel, kernel density estimation

*Corresponding author
Email address: k.pilario@cranfield.ac.uk (Karl Ezra S. Pilario)

1. Introduction

Large-scale industrial plants are nowadays highly-integrated and more complex. Hence, the necessary task of process health monitoring becomes more challenging (Chiang et al., 2005). Fortunately, with the rise of new technologies in automation and data acquisition, large data sets from these plants are readily available (Yin et al., 2015). By taking advantage of this, Multivariate Statistical Process Monitoring (MSPM) methods are deemed most favorable for monitoring complex industrial processes (Zhang and Zhang, 2010). Since process variables are highly correlated, MSPM methods are usually dimensionality reduction tools (Chiang et al., 2005) such as principal components analysis (PCA), partial least squares (PLS), independent component analysis (ICA), and canonical variate analysis (CVA). Data-driven methods are attractive because their use avoids the costly and time-consuming process of first-principles modelling for distinguishing between normal and faulty process operating conditions (Yin et al., 2015; Ge et al., 2013).

The key issues in MSPM are outlined by Ge et al. (2013). Plant data was described to be nonlinear, non-Gaussian, and dynamic in nature. Hence, through the decades, the MSPM methods are continuously being enhanced for *nonlinear dynamic process* monitoring. But aside from these, incipient fault monitoring is a more important issue and is recently gaining research attention. As opposed to abrupt faults, incipient faults are slowly developing process anomalies that start at small magnitudes. If not detected early, these faults can lead to an emergency situation or catastrophic failure (Vachtsevanos et al., 2006). Yet early detection is difficult, especially in closed-loop systems where the fault is initially masked by process control, and by noise or disturbances (Zhang et al., 2002).

To address this, nonlinear dynamic MSPM methods with enhanced sensitivity were recently proposed. Shang et al. (2018) used an augmented kernel Mahalanobis distance metric for improved fault detection, which avoids space partitioning in PCA. This produced a more sensitive detection index than CVA and PCA variants when tested in the Tennessee Eastman Plant. Meanwhile, Rato and Reis (2014) proposed sensitivity enhancing transformations, which also uses augmented data for accounting dynamics and nonlinearities. The adaptive kernel PCA by Cheng et al. (2010) uses the multivariate exponentially weighted moving average for capturing small mean shifts in the process. Recently, Pilario and Cao (2018) proposed the Canonical Variate Dissimilarity Analysis (CVDA) to detect incipient faults even at dynamically

varying process operating conditions. However, the nonlinear issue was not addressed sufficiently in that work. Indeed, incipient faults may develop to a point where the process behaves differently in a way that linear models cannot describe. One way to address this is to introduce kernel-based learning in CVDA.

Kernel methods are currently being used to handle the nonlinear issue with promising results. In kernel methods, the idea is to project the data onto a high-dimensional space using kernel functions, so that linear MSPM can be applied to the transformed data. Ever since Schölkopf et al. (1998) laid the foundations of kernel PCA, several other kernel MSPM methods have been reported in the literature. Recent works include the kernel dynamic PCA by Fezai et al. (2018) and Jaffel et al. (2016), the enhanced kernel PCA by Nguyen and Golinval (2010), the kernel PLS based generalized likelihood ratio test by Botre et al. (2016), the kernel dynamic ICA by Fan and Wang (2014), the weighted kernel ICA for non-Gaussian data by Cai et al. (2017), and the kernel CVA by Samuel and Cao (2015). Fault diagnosis using kernels applied to support vector machines (SVM) was also explored in numerous works, as surveyed by Yin and Hou (2016). For example, Zhang (2009) used kernel PCA and kernel ICA features as input to SVM for classifying faults. The most widely used kernel function in these studies, e.g. Cheng et al. (2010); Nguyen and Golinval (2010); Fan and Wang (2014); Samuel and Cao (2015); Bernal-de Lázaro et al. (2016), is the Gaussian radial basis function, or simply, the RBF kernel. Other choices include the polynomial and sigmoid kernels, to name a few.

In this paper, we first highlight some drawbacks in using the RBF kernel or any single kernel on their own for monitoring specifically incipient faults. Also, since CVDA is recognized as a dynamic MSPM method that is sensitive to incipient faults, we extend its applicability to nonlinear processes using kernel methods. As a result, a new kernel MSPM method is presented that is called Mixed Kernel Canonical Variate Dissimilarity Analysis (MK-CVDA). The overall method consists of a kernel PCA (KPCA) followed by CVDA. The grid search method is also used for finding optimal kernel parameters. In MK-CVDA, the same detection indices from CVDA, namely the T^2, Q, and D, are adopted. The non-Gaussianity issue is handled by using kernel density estimation for computing the detection limits of these indices. The new method is intended for monitoring nonlinear dynamic processes under varying operating conditions, where no prior fault information is needed.

The structure of the paper is organized as follows. KPCA is first revisited

in Section 2. Afterwards, mixed kernels are introduced in Section 3. Section 4 discusses the overall MK-CVDA method. The performance of MK-CVDA is evaluated in Section 5. Finally, the work is concluded in Section 6.

2. Kernel PCA Revisited

In general, kernel dynamic MSPM methods consist of: (i) data projection to the kernel space; (ii) augmentation of lagged variables to treat dynamics; and (iii) dimensionality reduction for partitioning the data space into the *state* and *residual* subspaces. For MK-CVDA, step (i) is done using kernel PCA (KPCA) and steps (ii)-(iii) are performed using CVDA. In this section, KPCA is revisited as follows.

Let $\mathbf{x}_k = \begin{bmatrix}\mathbf{u}_k^T & \mathbf{y}_k^T\end{bmatrix}^T \in \Re^m, k = 1, \ldots, N$ denote a data set of N observations of m variables, where $\mathbf{u}$ and $\mathbf{y}$ represent the process inputs and outputs, respectively. The $\mathbf{x}_k$ are normalized to zero mean and unit variance, giving us $\hat{\mathbf{x}}_k$.

In PCA, it is required to analyze features that live in a linear space. Thus, some nonlinear map $\mathbf{\Phi}(\cdot)$ must be used to project the data from the nonlinear input space onto a linear feature space F, i.e. $\mathbf{\Phi} : \Re^m \to F$. Assuming that $\sum_{k=1}^{N} \mathbf{\Phi}(\hat{\mathbf{x}}_k) = 0$, PCA seeks to solve an eigenvalue problem on the sample covariance in F, as follows:

$$\mathbf{C}^F = \frac{1}{N} \sum_{k=1}^{N} \mathbf{\Phi}(\hat{\mathbf{x}}_k) \mathbf{\Phi}(\hat{\mathbf{x}}_k)^T, \tag{1}$$

$$\mathbf{C}^F \mathbf{w} = \lambda \mathbf{w}, \tag{2}$$

where $\mathbf{C}^F$ is the sample covariance in F, $\mathbf{w}$ is an eigenvector, and λ is an eigenvalue.

Many kinds of nonlinear relationships must be accounted to design an ideal $\mathbf{\Phi}(\cdot)$, but it may inevitably result in a large dimensionality in F. So to avoid specifying $\mathbf{\Phi}(\cdot)$ explicitly, Schölkopf et al. (1998) suggested to represent dot products in F using kernel functions K for $(i, j) = 1, \ldots, N$ as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) \triangleq K_{ij} = \langle \mathbf{\Phi}(\mathbf{x}_i), \mathbf{\Phi}(\mathbf{x}_j) \rangle, \tag{3}$$

where $\langle \cdot, \cdot \rangle$ denotes dot product. Then, they modified Eq. (2) into the following set of equations for $k = 1, \ldots, N$:

$$\langle \mathbf{\Phi}(\hat{\mathbf{x}}_k), \mathbf{C}^F \mathbf{w} \rangle = \lambda \langle \mathbf{\Phi}(\hat{\mathbf{x}}_k), \mathbf{w} \rangle. \tag{4}$$

Noting that there exists some $\mathbf{v}$ such that $\mathbf{w} = \langle \mathbf{v}, \mathbf{\Phi}(\hat{\mathbf{x}}_k) \rangle$, the expression in Eq. (4) is then expanded, where all instances of $\langle \mathbf{\Phi}(\mathbf{x}_i), \mathbf{\Phi}(\mathbf{x}_j) \rangle$ are replaced with the kernels in Eq. (3), yielding a different eigenvalue problem:

$$\widehat{\mathbf{K}}\mathbf{v} = N\lambda\mathbf{v}, \tag{5}$$

where $\mathbf{v}$ is an eigenvector, λ is an eigenvalue, $\mathbf{K} \equiv [K_{ij}]$ is an $N \times N$ symmetric *kernel matrix*, and $\widehat{\mathbf{K}}$ is matrix $\mathbf{K}$ mean-centered in F by:

$$\widehat{\mathbf{K}} = \mathbf{K} - \mathbf{1}_N\mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N\mathbf{K}\mathbf{1}_N, \tag{6}$$

where $\mathbf{1}_N \in \mathbb{R}^{N\times N}$ and $(\mathbf{1}_N)_{ij} = 1/N$.

A form of nonlinear PCA now involves solving Eq. (5) instead of Eq. (2). Thus, the need to specify $\mathbf{\Phi}(\cdot)$ is eliminated since the nonlinear mapping is implicitly achieved by a so-called *kernel trick*. However, as it will be discussed in Section 3, not all functions can be used as kernels.

KPCA proceeds by forming the kernel matrix $\mathbf{K}$ from $\hat{\mathbf{x}}_k$ using Eq. (3) and centering $\mathbf{K}$ to $\widehat{\mathbf{K}}$ using Eq. (6). Due to Eq. (5), $\widehat{\mathbf{K}}$ is then diagonalized as

$$\widehat{\mathbf{K}}/N = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^T, \tag{7}$$

where $\mathbf{S} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N] \in \mathbb{R}^{N\times N}$ represents N eigenvectors and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N) \in \mathbb{R}^{N\times N}$ are eigenvalues where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. For the purpose of this study, only r number of principal components (PCs) that explain 99% of the total variance are retained. Denoting $\mathbf{S}_r$ as the first r columns of $\mathbf{S}$, PCs $\mathbf{t}_k$ are finally obtained by using the following equation:

$$\mathbf{T} \equiv [\mathbf{t}_k] = \mathbf{S}_r^T\widehat{\mathbf{K}} \in \mathbb{R}^{r\times N}. \tag{8}$$

In KPCA monitoring, the widely used T^2 and Q indices are computed for monitoring the principal subspace $\mathbf{T}$ and the residual subspace, respectively. However, dynamics in the data are not handled by KPCA alone. Hence, the MK-CVDA method takes the PCs $\mathbf{t}_k$ as input features to CVDA. So aside from data projection to the kernel space, KPCA effectively serves as a data whitening step (as did Fan and Wang (2014)), as well as a way to avoid singular matrices in CVDA afterwards (as did Samuel and Cao (2015)).

In summary, KPCA involves the transformation of training data $\mathbf{x}_k \in \Re^m$ into $\mathbf{t}_k \in \mathbb{R}^r$ by nonlinear projection to a feature space F, and further onto a subspace of F so as to perform whitening in MK-CVDA. In the following section, we discuss the choice of kernel functions in Eq. (3).

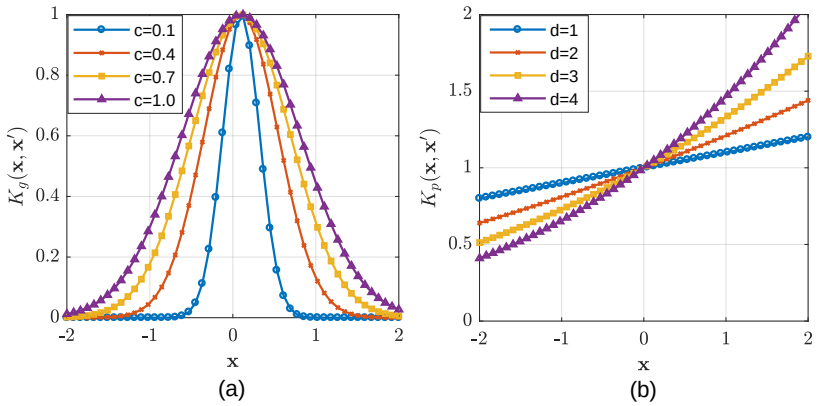


Figure 1: Sample plots of kernel values, where $\mathbf{x}' = 0.1$, for: (a) a local kernel (RBF kernel); and (b) a global kernel (polynomial kernel) (Zhu et al., 2012; Jordaan, 2002).

3. Choice of Kernel

In functional analysis, Mercer's theorem gives conditions for kernel functions that can act as a dot product in a possibly ∞-dimensional space, formally known as a Hilbert space (Cristianini and Shawe-Taylor, 2014). In loose terms, admissible kernels are said to be those that produce a positive semi-definite kernel matrix, $\mathbf{K}$. Although many different functions satisfy this requirement, Jordaan (2002) noted two main types of kernels: *local* and *global*. A typical example of a local kernel is the widely used Gaussian radial basis function (RBF), that is given by:

$$K_g(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{c}\right), \tag{9}$$

where c is the kernel width. It satisfies the Mercer condition for $c > 0$. It also corresponds to an ∞-dimensional space F, because the exponential can be viewed as a polynomial of *infinite* degree, when expressed as a power series. On the other hand, a typical example of a global kernel is the polynomial kernel, given by:

$$K_p(\mathbf{x}, \mathbf{x}') = (\langle \mathbf{x}, \mathbf{x}' \rangle + 1)^d, \tag{10}$$

where d is the kernel parameter that denotes the degree of the polynomial. This kernel satisfies the Mercer condition for $d \in \mathbb{N}$ (Smola et al., 2000). Others have found polynomial kernels more suitable than the RBF kernel for

certain applications, e.g. the penicillin process (Jia et al., 2012; Lee et al., 2004).

Sample plots of local and global kernel values are shown in Fig. 1, where $\mathbf{x}'$ represents a training sample from the normal process and $\mathbf{x}$ represents any unseen test sample to be mapped using K_g or K_p, as in Eqs. (9)-(10). Using these plots, the differences and limitations of each type of kernel are discussed as follows.

For the RBF kernel in Fig. 1(a), the behavior of an exponential function is expected: K_g tends to one as the difference between $\mathbf{x}$ and $\mathbf{x}'$ become zero, and tends to zero when their difference become large. However, the fact that the K_g mapping "vanishes" beyond a certain distance from the training data is undesirable in process monitoring. If two faulty samples differ in fault magnitude, but both are mapped to $K_g = 0$, then they would be perceived as the same. Worse, they may even be mistaken as normal. In most studies, much larger kernel widths c are chosen to increase the spread of K_g, e.g. Fan and Wang (2014) used $c = 500m$ (where m is the number of variables) and Samuel and Cao (2015) used a constant $c = 1720$ for the Tennessee Eastman Plant. However, these mappings still vanish beyond a certain distance from the training data. Although the RBF kernel can learn an effective mapping in the vicinity of the training data, i.e. good interpolation ability, it cannot influence a mapping over the *entire* data space. Thus, whenever c is said to be obtained empirically, it is actually chosen too large, hoping that the RBF kernel would extrapolate well. As a local kernel, however, the RBF kernel loses its interpolation ability at large c. The extent of this occurrence depends on the case study at hand. Hence, local kernels alone cannot exhibit both good interpolation and extrapolation abilities at the same time, as noted by Zhu et al. (2012).

On the other hand, Fig. 1(b) shows that the polynomial kernel extrapolates well because it creates a mapping on the entire data space, regardless of where it was trained. However, it only interpolates well at higher d. Thus, when global kernels are used alone, good extrapolation and interpolation abilities cannot be achieved at the same time either, as noted by Zhu et al. (2012).

In practice, a kernel that has both good interpolation and extrapolation abilities, i.e. good generalization, is desired. In a particular development of soft sensors, Jordaan (2002) proposed the use of a mixture of local and global kernels, which was also proven to satisfy Mercer's condition. In that work, a

convex combination of kernels was formed as given by:

$$K_{\text{mix}} = \omega K_p + (1 - \omega) K_g, \tag{11}$$

where $\omega \in [0, 1]$ is the mixture weight. Note that the mixed kernel reduces to the polynomial and RBF kernels at $\omega = 1$ and $\omega = 0$, respectively. Since then, more studies on mixed kernels have also been published. For instance, mixed kernel canonical correlation analysis (MKCCA) was proposed by Zhu et al. (2012) for various learning applications. Also, Zhong and Carr (2016) used mixed kernels for a support vector regression model. Moreover, a grammar for combining kernels was studied by Duvenaud (2014) for Gaussian process models. In this paper, the mixed kernel in Eq. (11) is used for nonlinear process monitoring of incipient faults. Having both interpolation and extrapolation abilities, mixed kernels are able to handle the nonlinear issue in process monitoring and also depict the notion of incipient fault growth properly as the process degrades in time.

According to Jordaan (2002), the weighted sum of a linear ($d = 1$) and RBF kernel is sufficient to balance good interpolation and extrapolation abilities. Hence, we adopt $d = 1$ for the rest of this paper. After the KPCA step in Section 2, we continue the MK-CVDA algorithm description in the following section, including a discussion on how to choose parameters in Eq. (11).

4. Mixed Kernel CVDA

CVDA is a framework based on canonical variate analysis (CVA), which is an effective dynamic MSPM method (Odiowei and Cao, 2010). CVDA aims to enhance CVA for incipient fault detection (Pilario and Cao, 2018). In this paper, the proposed MK-CVDA consists of a KPCA followed by CVDA. As KPCA was discussed in Section 2, we proceed with CVDA as follows.

4.1. Dimensionality reduction by CVDA

In CVDA, data are first arranged into past and future matrix blocks. However, only the process output variables must appear in the future data, considering that future inputs are independent from the past data. Thus, KPCA must be performed for $\mathbf{x}_k$ only, and another KPCA for $\mathbf{y}_k$ only.

Let $\mathbf{t}_k^{(1)} \in \mathbb{R}^{r_1}$ denote the PCs from $\mathbf{x}_k$, and $\mathbf{t}_k^{(2)} \in \mathbb{R}^{r_2}$ denote the PCs from $\mathbf{y}_k$. Although r_1 and r_2 are each chosen using the same cutoff criteria, they

are not necessarily equal. Lagged variables are formed in Hankel matrices as in Eqs. (12)-(13):

$$\mathbf{Y}_p = \begin{bmatrix} \mathbf{t}_p^{(1)} & \mathbf{t}_{p+1}^{(1)} & \mathbf{t}_{p+2}^{(1)} & \cdots & \mathbf{t}_{p+M-1}^{(1)} \\ \mathbf{t}_{p-1}^{(1)} & \mathbf{t}_p^{(1)} & \mathbf{t}_{p+1}^{(1)} & \cdots & \mathbf{t}_{p+M-2}^{(1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{t}_1^{(1)} & \mathbf{t}_2^{(1)} & \mathbf{t}_3^{(1)} & \cdots & \mathbf{t}_M^{(1)} \end{bmatrix}^T, \tag{12}$$

$$\mathbf{Y}_f = \begin{bmatrix} \mathbf{t}_{p+1}^{(2)} & \mathbf{t}_{p+2}^{(2)} & \mathbf{t}_{p+3}^{(2)} & \cdots & \mathbf{t}_{p+M}^{(2)} \\ \mathbf{t}_{p+2}^{(2)} & \mathbf{t}_{p+3}^{(2)} & \mathbf{t}_{p+4}^{(2)} & \cdots & \mathbf{t}_{p+M+1}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{t}_{p+f}^{(2)} & \mathbf{t}_{p+f+1}^{(2)} & \mathbf{t}_{p+f+2}^{(2)} & \cdots & \mathbf{t}_N^{(2)} \end{bmatrix}^T, \tag{13}$$

where $\mathbf{Y}_p \in \mathbb{R}^{M \times r_1 p}$ and $\mathbf{Y}_f \in \mathbb{R}^{M \times r_2 f}$ are respectively the past and future data matrices, p and f are respectively the number of lags in the past and future, and $M = N - p - f + 1$. Here, the amount of lags are chosen using autocorrelation analysis (Odiowei and Cao, 2010), but it must be ensured that $r_1 p < M$ and $r_2 f < M$ for results to make sense (Samuel and Cao, 2015). The Hankel matrices are normalized using the mean and standard deviation of each column, giving us $\widehat{\mathbf{Y}}_p$ and $\widehat{\mathbf{Y}}_f$. Canonical correlation analysis (CCA) is then performed between $\widehat{\mathbf{Y}}_p$ and $\widehat{\mathbf{Y}}_f$. But since they may be rank-deficient after KPCA, they must first be factored QR decomposition:

$$\widehat{\mathbf{Y}}_p = \mathbf{Q}_p \mathbf{R}_p \mathbf{\Pi}_p^T, \tag{14}$$

$$\widehat{\mathbf{Y}}_f = \mathbf{Q}_f \mathbf{R}_f \mathbf{\Pi}_f^T, \tag{15}$$

where $\mathbf{Q}_p$ and $\mathbf{Q}_f$ are $M \times M$ column orthogonal matrices, $\mathbf{R}_p$ and $\mathbf{R}_f$ are upper triangular matrices of size $M \times r_1 p$ and $M \times r_2 f$ respectively, and $\mathbf{\Pi}_p$ and $\mathbf{\Pi}_f$ are permutation matrices of size $r_1 p \times r_1 p$ and $r_2 f \times r_2 f$, respectively. The latter are used to permute the rows of $\mathbf{R}$ to have non-increasing absolute value of diagonal elements.

Let $\rho_1 = \text{rank}(\widehat{\mathbf{Y}}_p)$ and $\rho_2 = \text{rank}(\widehat{\mathbf{Y}}_f)$. Only the first ρ_1 columns of $\mathbf{Q}_p$ and the first ρ_2 columns of $\mathbf{Q}_f$ are taken and denoted as $\mathbf{Q}'_p$ and $\mathbf{Q}'_f$, respectively. The corresponding top-left $\rho_1 \times \rho_1$ submatrix of $\mathbf{R}_p$ and top-left

$\rho_2 \times \rho_2$ submatrix of $\mathbf{R}_f$ are also taken as $\mathbf{R}'_p$ and $\mathbf{R}'_f$, respectively. Now, CCA involves the singular value decomposition (SVD) of the sample correlation matrix $\mathbf{H}$ as follows:

$$\mathbf{H} = (\mathbf{Q}'_f)^T \mathbf{Q}'_p = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{16}$$

where $\mathbf{U}$ and $\mathbf{V}$ consist of the left and right singular columns of $\mathbf{H}$, respectively, and $\mathbf{\Sigma}$ is a diagonal matrix of sorted singular values, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_\rho$ with $\rho = \min(\rho_1, \rho_2)$. The singular vectors are rescaled and reordered by:

$$\mathbf{U}^* = \begin{bmatrix} (\mathbf{R}'_f)^{-1}\mathbf{U}\sqrt{M-1} \\ \mathbf{0} \end{bmatrix} \mathbf{\Pi}_f^T \in \mathbb{R}^{r_2 f \times r_2 f}, \tag{17}$$

$$\mathbf{V}^* = \begin{bmatrix} (\mathbf{R}'_p)^{-1}\mathbf{V}\sqrt{M-1} \\ \mathbf{0} \end{bmatrix} \mathbf{\Pi}_p^T \in \mathbb{R}^{r_1 p \times r_1 p}, \tag{18}$$

wherein the zero rows would appear only if dependent columns exist in $\widehat{\mathbf{Y}}_p$ or $\widehat{\mathbf{Y}}_f$. The CCA implementation above is same as the CCA built-in function in MATLAB®.

Since only n (with $n < \rho$) dominant singular values explain the system dynamics (Pilario and Cao, 2018), only the first n columns of $\mathbf{U}^*$ and $\mathbf{V}^*$ are collected and denoted as $\mathbf{U}_n^*$ and $\mathbf{V}_n^*$, respectively. Projection matrices $\mathbf{J}$, $\mathbf{L}$ and $\mathbf{F}$ are formed as:

$$\mathbf{J} = (\mathbf{V}_n^*)^T \in \mathbb{R}^{n \times r_1 p}, \tag{19}$$

$$\mathbf{L} = (\mathbf{U}_n^*)^T \in \mathbb{R}^{n \times r_2 f}, \tag{20}$$

$$\mathbf{F} = (\mathbf{I} - \mathbf{J}^T\mathbf{J})^T \in \mathbb{R}^{r_1 p \times r_1 p}, \tag{21}$$

which are used to reveal the state $\mathbf{Z}$ and residual $\mathbf{E}$ subspaces, as follows:

$$\mathbf{Z} \equiv [\mathbf{z}_k] = \mathbf{J}\widehat{\mathbf{Y}}_p^T \in \mathbb{R}^{n \times M}, \tag{22}$$

$$\mathbf{E} \equiv [\mathbf{e}_k] = \mathbf{F}\widehat{\mathbf{Y}}_p^T \in \mathbb{R}^{r_1 p \times M}, \tag{23}$$

where $\mathbf{z}_k$ are the state variables and $\mathbf{e}_k$ are the residual variables for $k = 1, \ldots, M$. Lastly, dissimilarity features $\mathbf{D}$ are computed as:

$$\mathbf{D} \equiv [\mathbf{d}_k] = \mathbf{L}\widehat{\mathbf{Y}}_f^T - \mathbf{\Sigma}_n \mathbf{J}\widehat{\mathbf{Y}}_p^T \in \mathbb{R}^{n \times M}, \tag{24}$$

where $\mathbf{\Sigma}_n = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ from Eq. (16).

4.2. MK-CVDA monitoring

The same detection indices from the CVDA framework are adopted for MK-CVDA, defined as follows:

$$T_k^2 = \mathbf{z}_k^T \mathbf{z}_k, \tag{25}$$

$$Q_k = \mathbf{e}_k^T \mathbf{e}_k, \tag{26}$$

$$D_k = \mathbf{d}_k^T (\mathbf{I} - \boldsymbol{\Sigma}_n^2)^{-1} \mathbf{d}_k. \tag{27}$$

Upper control limits (UCL), denoted by T^2_{UCL}, Q_{UCL}, and D_{UCL}, are computed using kernel density estimation (KDE) as explained in Odiowei and Cao (2010) and Pilario and Cao (2018). In KDE, the distributions of the indices are estimated, which may not necessarily be Gaussian. Given a significance level, α, the UCLs are solved such that $P(J < J_{\text{UCL}}) = \alpha$ where $J \in \{T^2, Q, D\}$. The kth sample is considered faulty if any of T_k^2, Q_k, or D_k exceeded T^2_{UCL}, Q_{UCL}, or D_{UCL}, respectively. In this paper, we adopted a 99.9% significance level for all monitoring tasks.

During online monitoring, a data block of the last $l = p + f$ test samples $\mathbf{x}_i^{\text{test}} \in \Re^m, i = (k - l + 1), \ldots, k$ are taken at the kth sampling instant. This data is normalized using the mean and standard deviation of the training set, giving us $\hat{\mathbf{x}}_i^{\text{test}}$. The $\hat{\mathbf{x}}_i^{\text{test}}$ is projected to the feature space F using $\mathbf{K}^{\text{test}} = K(\hat{\mathbf{x}}_i^{\text{test}}, \hat{\mathbf{x}}_j) \in \mathbb{R}^{l \times N}$ where $\hat{\mathbf{x}}_j$ represents all training samples $j = 1, \ldots, N$. $\mathbf{K}^{\text{test}}$ is then centered as

$$\widehat{\mathbf{K}}^{\text{test}} = \mathbf{K}^{\text{test}} - \mathbf{1}_N^{\text{test}} \mathbf{K} - \mathbf{K}^{\text{test}} \mathbf{1}_N + \mathbf{1}_N^{\text{test}} \mathbf{K} \mathbf{1}_N, \tag{28}$$

where $\mathbf{1}_N^{\text{test}} \in \mathbb{R}^{l \times N}$ and $(\mathbf{1}_N^{\text{test}})_{ij} = 1/N$. The KPCA step gives the PCs obtained from the test data as

$$\mathbf{T}^{\text{test}} \equiv [\mathbf{t}_i^{\text{test}}] = (\widehat{\mathbf{K}}^{\text{test}} \mathbf{S}_r)^T \in \mathbb{R}^{r \times l}. \tag{29}$$

The above projections are done to the output variables $\hat{\mathbf{y}}_i^{\text{test}}$ as well, using the same kernel parameters. For the CVDA step, lagged data vectors $\mathbf{y}_p$ and $\mathbf{y}_f$ are formed as in the first rows of Eq. (12) and Eq. (13). After normalizing, $\widehat{\mathbf{y}}_p$ and $\widehat{\mathbf{y}}_f$ are subjected to projection matrices, $\mathbf{J}$, $\mathbf{F}$, and $\mathbf{L}$, to yield $\mathbf{z}_k$, $\mathbf{e}_k$, and $\mathbf{d}_k$ as in Eqs. (22)-(24). Finally, detection indices are obtained using Eqs. (25)-(27) for fault detection. In summary, we outline the overall algorithm of MK-CVDA in Fig. 2.

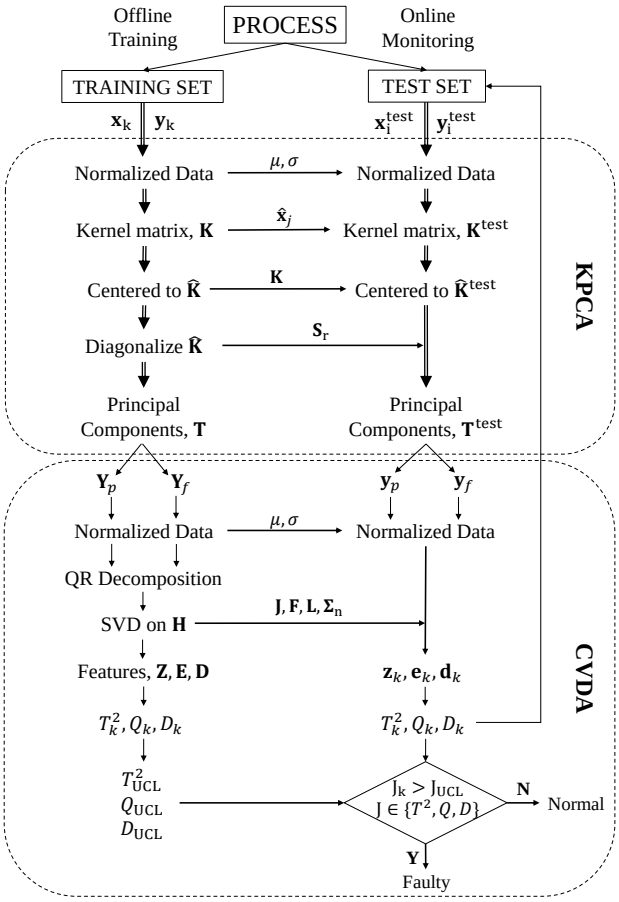


Figure 2: MK-CVDA algorithm for nonlinear process monitoring.

4.3. Grid search method

The parameters that must be set prior to the application of MK-CVDA are: the mixed kernel parameters c and ω, and the number of dominant singular values, or states, n in CVDA. Because these parameters are difficult to determine automatically, they must be subjected to an optimization procedure, where the objective might be to select $[c, \omega, n]$ that best distinguishes normal from faulty process conditions (as did Bernal-de Lázaro et al. (2016)). However, we assume that no prior fault information is available for checking this criteria. So to choose $[c, \omega, n]$, we take two different data sets, both from the normal operation of the process: SET 1 (the training set) is used to train an MK-CVDA model and SET 2 (the validation set) is used to evaluate the

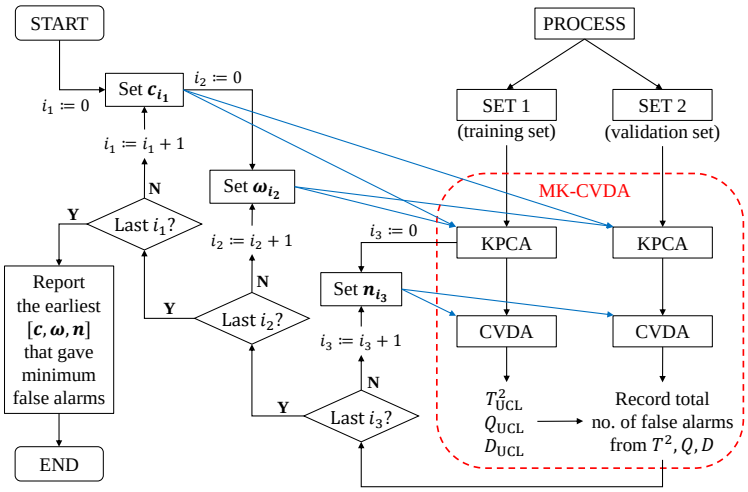


Figure 3: MK-CVDA offline training with grid search algorithm.

model trained from SET 1. The optimal $[c, \omega, n]$ is defined as that which minimizes the combined false alarms incurred by the T^2, Q, and D indices in monitoring SET 2.

Several approaches exist for optimizing kernel parameters. Bernal-de Lázaro et al. (2016) used differential evolution (DE) and particle swarm optimization (PSO), and Jia et al. (2012) used genetic algorithms (GAs), to name a few. Although the use of these metaheuristics is attractive, the precision of results is not worth the computational effort. In reality, only a small range of $[c, \omega, n]$ needs to be explored, since under- or overitting may occur outside these ranges (Zhu et al., 2012). For example, too small c makes the RBF kernel sensitive to noise, while too large c creates a smooth mapping that may behave as linear (Bernal-de Lázaro et al., 2016). A similar case for choosing n is discussed by Ruiz-Cárcel et al. (2015). Hence, the grid search method is adopted to find optimal parameters (see Zhu et al. (2012)). In grid search, combinations from only a finite set of values of $[c, \omega, n]$ are explored. The set of values are pre-defined manually depending on the problem.

In summary, the grid search method is used to decide kernel parameters c and ω, and the number of states, n, in MK-CVDA, by way of minimizing false alarms in a validation data set. After defining the sets of $[c, \omega, n]$ to explore, grid search is performed as represented in Fig. 3.

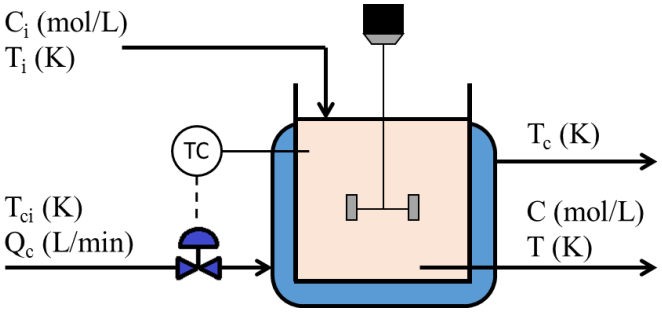


Figure 4: Schematic of the CSTR case study.

5. Case Study

In this section, the proposed MK-CVDA is evaluated using a closed-loop continuous stirred-tank reactor (CSTR) case study, described in Pilario and Cao (2018). A schematic of the CSTR is given in Fig. 4. Simulated data from this process are produced by the following nonlinear state-space model:

$$\frac{dC}{dt} = \frac{Q}{V}(C_i - C) - akC + \nu_1, \tag{30}$$

$$\frac{dT}{dt} = \frac{Q}{V}(T_i - T) - a\frac{(\Delta H_r)kC}{\rho C_p} - b\frac{UA}{\rho C_p V}(T - T_c) + \nu_2, \tag{31}$$

$$\frac{dT_c}{dt} = \frac{Q_c}{V_c}(T_{ci} - T_c) + b\frac{UA}{\rho_c C_{pc} V_c}(T - T_c) + \nu_3, \tag{32}$$

where the inputs are $\mathbf{u} = \begin{bmatrix} C_i & T_i & T_{ci} \end{bmatrix}^T$, the outputs are $\mathbf{y} = \begin{bmatrix} C & T & T_c & Q_c \end{bmatrix}^T$, and $k = k_0 \exp\left(\frac{-E}{RT}\right)$. Here, the same controller settings and parameter values in Eqs. (30)-(32) were used as those in Pilario and Cao (2018). Simulations of normal and faulty data were carried out under varying operating conditions every 60 min. In this paper, we investigate only the incipient faults listed in Table 1. Fault 1 is a slow decay in catalyst activity, introduced by decreasing the value of a in Eq. (30) and Eq. (31) to zero. Fault 2 is a fouling fault in the cooling jacket, introduced by decreasing the value of b in Eq. (31) and Eq. (32) to zero. Lastly, fault 3 is a drift in the readings of reactor temperature, which is the controlled variable. This fault produces oscillations to the coolant flow rate, as the controller becomes saturated.

Table 1: Incipient fault scenarios in the CSTR

Fault	Description†	Nominal Values	Name
1	$a = a_0 \exp(-0.0005\, t)$	$a_0 = 1$	Catalyst Decay
2	$b = b_0 \exp(-0.001\, t)$	$b_0 = 1$	Fouling
3	$T = T_0 + 0.01\, t$	$T_0 = 430\,\mathrm{K}$	Sensor Drift

† All t in minutes.

In the following, we first demonstrate the importance of mixed kernels visually using KPCA on 2-dimensional CSTR data, and then present MK-CVDA online monitoring results using the entire input-output CSTR data.

5.1. Illustrative study of mixed kernels for 2D data

Consider a data set of 300 normal samples and 300 faulty samples of only the reactor temperature T and concentration C in the CSTR, where the fault condition is fault 1 in Table 1. Fig. 5(a) illustrates the raw data in the original data space. Due to process control, the normal samples are expected to cluster at the desired setpoint, which is $[T_{sp}, C_{sp}] = [430\ \mathrm{K}, 0.1\ \mathrm{mol/L}]$. These samples are then normalized to zero mean and unit variance prior to KPCA. To illustrate, the dashed box in Fig. 5(a) represents the range $[\mu - 3\sigma, \mu + 3\sigma]$ in each dimension of the normal data samples.

Following the kernel parameter recommendations of Choi et al. (2005), projection maps of the first 3 KPCA components of the normalized data using an RBF kernel of width $c = 10$ are shown in Fig. 5(b). Due to the use of only a local kernel with an arbitrarily chosen kernel parameter, the normal data samples were able to influence mappings *only* within their vicinity in the data space. When these mappings are applied to the faulty samples, most of the samples will be undesirably projected to zero kernel value. In effect, the notion of increasing fault severity as test samples move farther from the normal data is not encoded. Even worse, far-away faulty samples may be perceived as normal. Although increasing c may enlarge the influence of the mappings, an influence over the entire data space is not guaranteed. Also, at large c, the RBF kernel loses its interpolation ability within the vicinity of the normal data samples. Thus, the ability of KPCA to handle the nonlinear issue is compromised.

On the other hand, Fig. 5(c) shows the projection maps of the first 3 KPCA components using a mixed kernel. Several advantages of using mixed

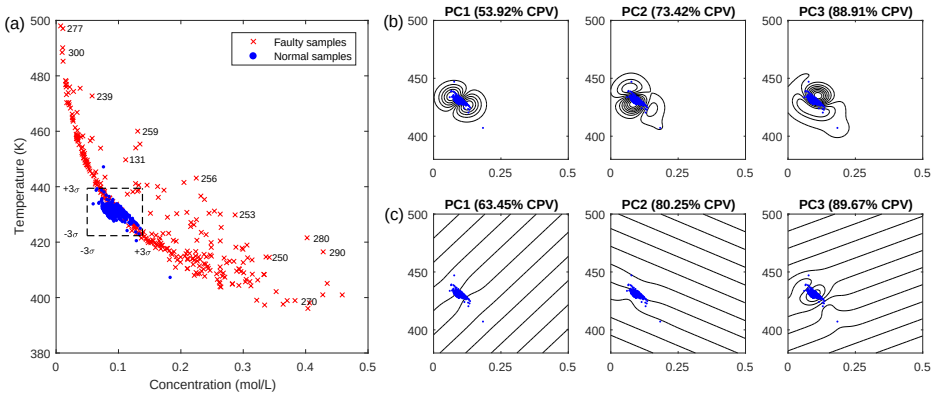


Figure 5: KPCA monitoring of CSTR data: (a) raw data in T-C space; (b) projection maps from the first 3 principal components using RBF kernel ($c = 10$); (c) projection maps from the first 3 principal components using Mixed kernel ($\omega = 0.1, d = 1, c = 10$). Normal and faulty samples are shown in blue and red, respectively.

kernels can be deduced from the figure: (i) the interpolation ability of the mixed kernel is retained as evidenced by nonlinear contours within the vicinity of the normal data samples; (ii) the extrapolation ability of the mixed kernel is also retained as evidenced by influencing a mapping over the entire data space; and, (iii) the notion of increasing fault severity as test samples move farther from the normal data is encoded. The latter is required in any kernel-based incipient fault monitoring method since the kernel-based model should exhibit a degradation behavior as the fault grows in time. Hence, mixed kernels must be used for incipient fault monitoring rather than single (local or global) kernels.

For the rest of this section, the performance of MK-CVDA is discussed using the entire CSTR input-output data.

5.2. MK-CVDA offline training

To proceed with offline training, a training set and a validation set is generated from the CSTR, each consisting of 1200 samples of the 7 variables. The sampling interval is 1 min. Random seeds for input disturbances and noise are different between the training and validation sets. Grid search (Fig. 3) was then used to find MK-CVDA parameters, $[c, \omega, n]$, such that false alarms in the validation set are minimized. For this case, we considered the sets: $c \in \{1, 0.5, 2, \ldots, 10\}$, $\omega \in \{0, 0.05, 0.1, \ldots, 1\}$, and $n \in \{2, 3, \ldots, 12\}$.

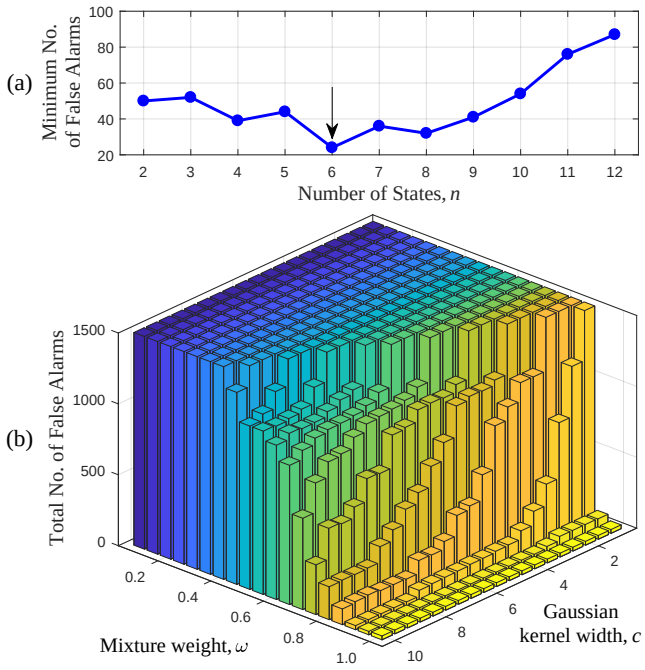


Figure 6: Grid search results for the CSTR: (a) no. of false alarms against n; (b) no. of false alarms against $[c, \omega]$ at $n = 6$. (Note: The upper portion of the bar graph is tapered at 1500 in the z-axis.)

Besides, if smaller increments are used in the c and ω sets, a change in the number of false alarms may not occur between adjacent choices.

In Fig. 6(a), the minimum number of false alarms ever recorded for each n (at any $[c, \omega]$) are plotted, indicating that the CSTR must have $n = 6$ states. Further in Fig. 6(b), a bar graph of the number of false alarms against choices of $[c, \omega]$ for $n = 6$ is shown. These results agrees with those from (Jordaan, 2002; Zhu et al., 2012) in that the choice of mixture weight is desirable at $\omega \geq 0.9$, i.e. only a "pinch" of the RBF kernel needs to be added to improve the interpolation ability of a low-order polynomial kernel. In other words, the influence of each kernel type in the mixture may not necessarily balance at $\omega = 0.5$. At low c and ω, Fig. 6(b) shows that the resulting MK-CVDA models are not suitable since the detection indices are found above the detection limits most of the time. In the end, the grid search found $[c, \omega, n] = [4.5, 0.95, 6]$ as the optimal parameters for MK-CVDA monitoring of the CSTR.

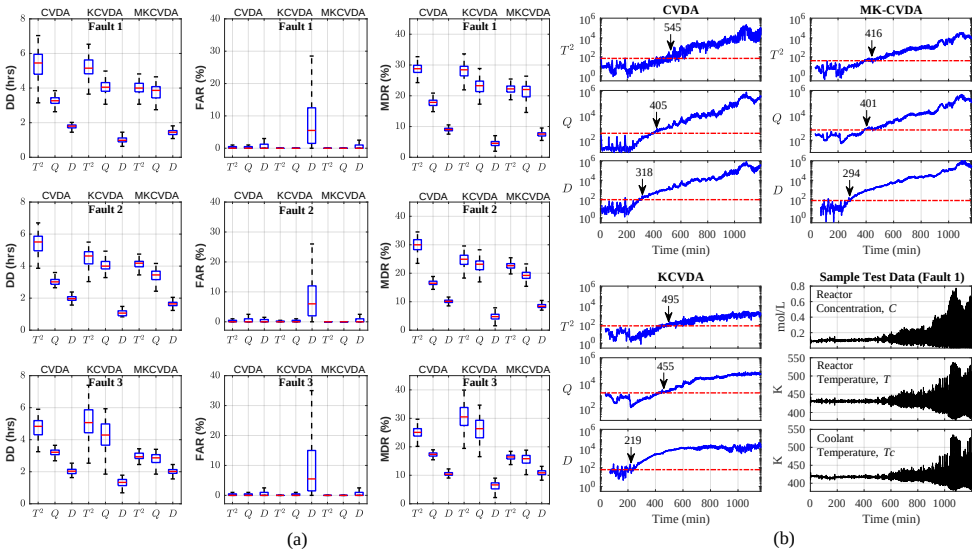


Figure 7: Monitoring performance on CSTR incipient faults: (a) box plots comparing CVDA, KCVDA, and MK-CVDA using DD, FAR, and MDR (left to right) on 3 faults (top to bottom); (b) sample monitoring charts (with detection times) and plots of C, T, T_c data set from Fault 1. Note: In (a), MATLAB's `boxplot` function without outliers was used. The top and bottom box edges correspond to the 75th and 25th percentiles, respectively.

5.3. MK-CVDA online monitoring

The performance of any process monitoring method can be evaluated using detection delays (DD), false alarm rates (FAR), and missed detection rates (MDR). In this paper, detection time is defined as the first time when 10 consecutive alarms were raised from the start of operation. Hence, DD is the period between the start of fault and the detection time. Also, standard FAR and MDR definitions are given as:

$$\text{FAR} = \frac{\text{no. of samples } (J > J_{\text{UCL}} | \text{fault-free})}{\text{total samples (fault-free)}} \times 100\%, \tag{33}$$

$$\text{MDR} = \frac{\text{no. of samples } (J < J_{\text{UCL}} | \text{fault})}{\text{total samples (fault)}} \times 100\%. \tag{34}$$

Using these performance metrics, we compared our proposed MK-CVDA method with linear CVDA and KCVDA (which uses RBF kernels only). For linear CVDA, the settings found in (Pilario and Cao, 2018) are adopted. For

KCVDA, we adopt $c = 350$ and $n = 8$ as obtained from the guidelines by Choi et al. (2005).

For a robust comparison, 200 test data sets were generated for each fault scenario in Table 1, while varying the random seeds for disturbances and noise. For all test data sets, an initial 200 min of normal operation is followed by the introduction of each incipient fault until 1200 min. After monitoring all test data, performance results are summarized in Fig. 7(a) as box plots. Each row of box plots correspond to a fault scenario, while each column of box plots correspond to DD, FAR, and MDR results. Within a box plot, the detection indices from CVDA, KCVDA, and MK-CVDA are compared with each other. In Fig. 7(b), sample monitoring charts for the Fault 1 scenario (catalyst decay) are given, along with part of the test data that was monitored in these charts. In general, a good detection index must have low DD, FAR, and MDR, and must also depict the severity of the fault properly above the detection limit. Using these criteria, the detection indices were evaluated as follows.

KCVDA has significant merits over linear CVDA. Due to the handling of nonlinear data, KCVDA is expected to create tighter bounds around the normal data. Hence, T^2_{KCVDA} and D_{KCVDA} detected faults 1 and 2 earlier than T^2_{CVDA} and D_{CVDA} in Fig. 7(a). Also, less false alarms were incurred by T^2_{KCVDA} and Q_{KCVDA} compared to T^2_{CVDA} and Q_{CVDA} for all faults. However, T^2_{KCVDA} and Q_{KCVDA} performed worse than T^2_{CVDA} and Q_{CVDA} for fault 3, in terms of DD and MDR. More importantly, since the RBF kernel lost its interpolation ability at a large c ($= 350$), D_{KCVDA} gave an unacceptably large FAR in all fault scenarios. In the KCVDA monitoring chart (Fig. 7(b)), D_{KCVDA} achieved a low DD precisely because of prevalent false alarms. Lastly, the KCVDA indices do not reflect the fault severity properly above the detection limit. Note that the incipient fault continues to degrade the process (as seen in the C, T, and T_c profiles), especially at 600-1200 min of operation. Yet, the KCVDA indices remain levelled during these times. Although this can be resolved by increasing c further, it will be at the expense of the interpolation ability of the RBF kernel. This demonstrates the limitation of KCVDA using the RBF kernel alone.

Now, MK-CVDA also improves upon CVDA by handling nonlinear data: Earlier detection and lesser MDR were achieved by T^2_{MKCVDA} and D_{MKCVDA} compared to those of CVDA in faults 1 and 2. For fault 3, all MK-CVDA indices performed better than their CVDA counterparts in terms of all the metrics. As seen in Fig. 7(b), the spiking behavior found in the CVDA

monitoring charts (caused by the abrupt disturbance change every 60 min (Pilario and Cao, 2018)), was eliminated in any of the MK-CVDA indices. Hence, the MK-CVDA indices resulted in less false alarms than CVDA as seen in Fig. 7(a). This demonstrates the good interpolation ability of the mixed kernel. More importantly, fault severity is reflected properly in all the MK-CVDA indices, which demonstrates good extrapolation ability. Due to this, we argue that any kernel MSPM method must consider mixed kernels for a more general approach to process monitoring. Finally, we emphasize the advantage of the canonical variate dissimilarity index, D, observed in all results above: It provides enhanced sensitivity for incipient faults as evidenced by having the earliest detection time among the three indices in all fault scenarios.

In summary, using the CSTR case study, the improved performance of the proposed MK-CVDA over the linear CVDA and RBF-kernel based KCVDA is realized.

6. Conclusion

In this paper, a new Mixed Kernel Canonical Variate Dissimilarity Analysis (MK-CVDA) method was proposed for process monitoring. The proposed method has the capability of handling the nonlinear, non-Gaussian, and dynamic nature of the data all at once, while being sensitive to incipient faults at the same time. For addressing nonlinearities using kernel methods, we first examined the drawbacks of using the RBF kernel or the polynomial kernel on their own. To address these, we adopted the use of mixed kernels, where the RBF and linear kernels were combined to achieve good generalization ability. To decide parameters for MK-CVDA, the grid search optimization method was used. A closed-loop continuous stirred tank reactor (CSTR) case study demonstrated the improved performance of the MK-CVDA indices over the linear CVDA indices in terms of detection delay, false alarm rates, and missed detection rates. More importantly, the growth of a fault across time is better depicted by the MK-CVDA indices beyond their detection limits.

The MK-CVDA method in this study can be extended in the future for fault diagnosis and prognosis. More importantly, the idea of combining kernels for enhancing generalization capabilities (Duvenaud, 2014) of kernel-based methods offers new directions for nonlinear process monitoring research.

Acknowledgement

The authors gratefully acknowledge the support from the DOST-ERDT Faculty Development Fund of the Republic of the Philippines.

References

Botre, C., Mansouri, M., Nounou, M., Nounou, H., Karim, M.N., 2016. Kernel PLS-based GLRT method for fault detection of chemical processes. J. Loss Prev. Process Ind. 43, 212–224. doi:10.1016/j.jlp.2016.05.023.

Cai, L., Tian, X., Chen, S., 2017. Monitoring Nonlinear and Non-Gaussian Processes Using Gaussian Mixture Model-Based Weighted Kernel Independent Component Analysis. IEEE Trans. Neural Networks Learn. Syst. 28, 122–135. doi:10.1109/TNNLS.2015.2505086.

Cheng, C.Y., Hsu, C.C., Chen, M.C., 2010. Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes. Ind. Eng. Chem. Res. 49, 2254–2262. doi:10.1021/ie900521b.

Chiang, L.H., Russell, E.L., Braatz, R.D., 2005. Fault Detection and Diagnosis in Industrial Systems. Springer-Verlag, London.

Choi, S.W., Lee, C., Lee, J.M., Park, J.H., Lee, I.B., 2005. Fault detection and identification of nonlinear processes based on kernel PCA. Chemom. Intell. Lab. Syst. 75, 55–67. doi:10.1016/j.chemolab.2004.05.001.

Cristianini, N., Shawe-Taylor, J., 2014. Support Vector Machines and other kernel-based learning methods. Cambridge University Press.

Duvenaud, D.K., 2014. Automatic Model Construction with Gaussian Processes. Ph.D. thesis. Cambridge University.

Fan, J., Wang, Y., 2014. Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis. Inf. Sci. (Ny). 259, 369–379. doi:10.1016/j.ins.2013.06.021.

Fezai, R., Mansouri, M., Taouali, O., Harkat, M.F., Bouguila, N., 2018. Online reduced kernel principal component analysis for process monitoring. J. Process Control 61, 1–11. doi:10.1016/j.jprocont.2017.10.010.

Ge, Z., Song, Z., Gao, F., 2013. Review of Recent Research on Data-based Process Monitoring. Ind. Eng. Chem. Res. 52, 3543–3562.

Jaffel, I., Taouali, O., Harkat, M.F., Messaoud, H., 2016. Moving window KPCA with reduced complexity for nonlinear dynamic process monitoring. ISA Trans. 64, 184–192. doi:10.1016/j.isatra.2016.06.002.

Jia, M., Xu, H., Liu, X., Wang, N., 2012. The optimization of the kind and parameters of kernel function in KPCA for process monitoring. Comput. Chem. Eng. 46, 94–104. doi:10.1016/j.compchemeng.2012.06.023.

Jordaan, E.M., 2002. Development of robust inferential sensors: Industrial applications of support vector machines for regression. Ph.D. thesis. Technische Universiteit Eindhoven. doi:10.6100/IR561175.

Bernal-de Lázaro, J.M., Llanes-Santiago, O., Prieto-Moreno, A., Knupp, D.C., Silva-Neto, A.J., 2016. Enhanced dynamic approach to improve the detection of small-magnitude faults. Chem. Eng. Sci. 146, 166–179. doi:10.1016/j.ces.2016.02.038.

Lee, J.M., Yoo, C.K., Lee, I.B., 2004. Fault detection of batch processes using multiway kernel principal component analysis. Comput. Chem. Eng. 28, 1837–1847. doi:10.1016/j.compchemeng.2004.02.036.

Nguyen, V.H., Golinval, J.C., 2010. Fault detection based on Kernel Principal Component Analysis. Eng. Struct. 32, 3683–3691. doi:10.1016/j.engstruct.2010.08.012.

Odiowei, P.E., Cao, Y., 2010. Nonlinear Dynamic Process Monitoring Using Canonical Variate Analysis and Kernel Density Estimations. IEEE Trans. Ind. Informatics 6, 36–45. doi:10.1109/TII.2009.2032654.

Pilario, K.E.S., Cao, Y., 2018. Canonical Variate Dissimilarity Analysis for Process Incipient Fault Detection. IEEE Trans. Ind. Informatics , 1–1doi:10.1109/TII.2018.2810822.

Rato, T.J., Reis, M.S., 2014. Sensitivity enhancing transformations for monitoring the process correlation structure. J. Process Control 24, 905–915. doi:10.1016/j.jprocont.2014.04.006.

Ruiz-Cárcel, C., Cao, Y., Mba, D., Lao, L., Samuel, R.T., 2015. Statistical process monitoring of a multiphase flow facility. Control Eng. Pract. 42, 74–88. doi:10.1016/j.conengprac.2015.04.012.

Samuel, R.T., Cao, Y., 2015. Kernel canonical variate analysis for nonlinear dynamic process monitoring. IFAC-PapersOnLine 28, 605–610. doi:10.1016/j.ifacol.2015.09.034.

Schölkopf, B., Smola, A., Müller, K.R., 1998. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Comput. 10, 1299–1319. doi:10.1162/089976698300017467.

Shang, J., Chen, M., Zhang, H., 2018. Fault detection based on augmented kernel Mahalanobis distance for nonlinear dynamic processes. Comput. Chem. Eng. 109, 311–321. doi:10.1016/j.compchemeng.2017.11.010.

Smola, A.J., Ovari, Z.L., Williamson, R.C., 2000. Regularization with Dot-Product Kernels, in: Proc. Neural Inf. Process. Syst., MIT Press. pp. 308–314.

Vachtsevanos, G., Lewis, F.L., Roemer, M., Hess, A., Wu, B., 2006. Intelligent Fault Diagnosis and Prognosis for Engineering Systems. John Wiley & Sons, Inc.

Yin, S., Li, X., Gao, H., Kaynak, O., 2015. Data-based techniques focused on modern industry: An overview. IEEE Trans. Ind. Electron. 62, 657–667. doi:10.1109/TIE.2014.2308133.

Yin, Z., Hou, J., 2016. Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. Neurocomputing 174, 643–650. doi:10.1016/j.neucom.2015.09.081.

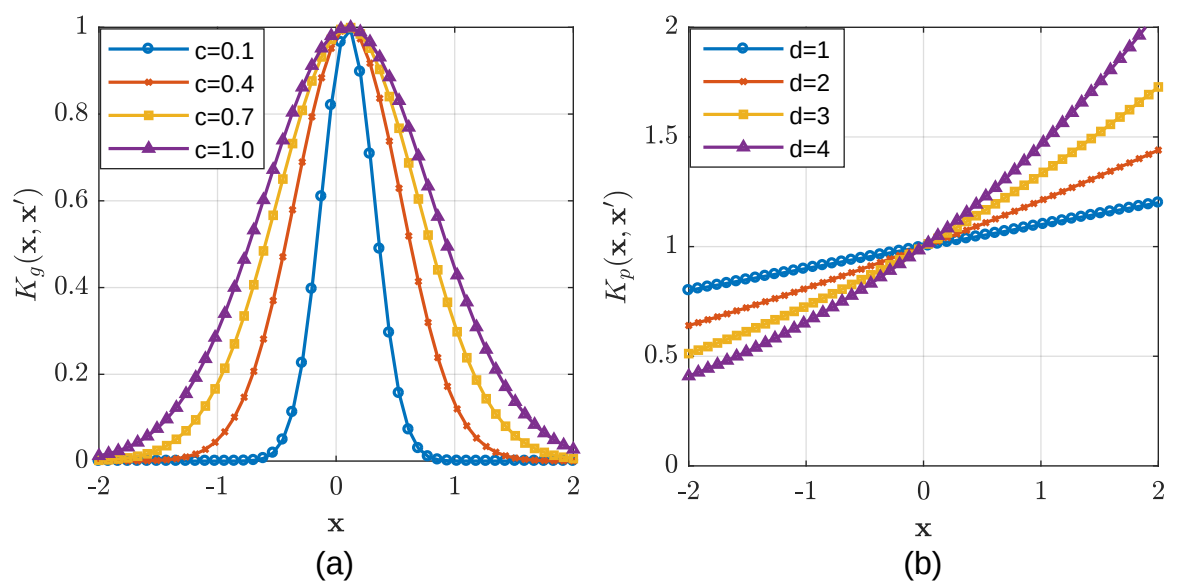
Zhang, X., Polycarpou, M.M., Parisini, T., 2002. A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems. IEEE Trans. Automat. Contr. 47, 576–593. doi:10.1109/9.995036.

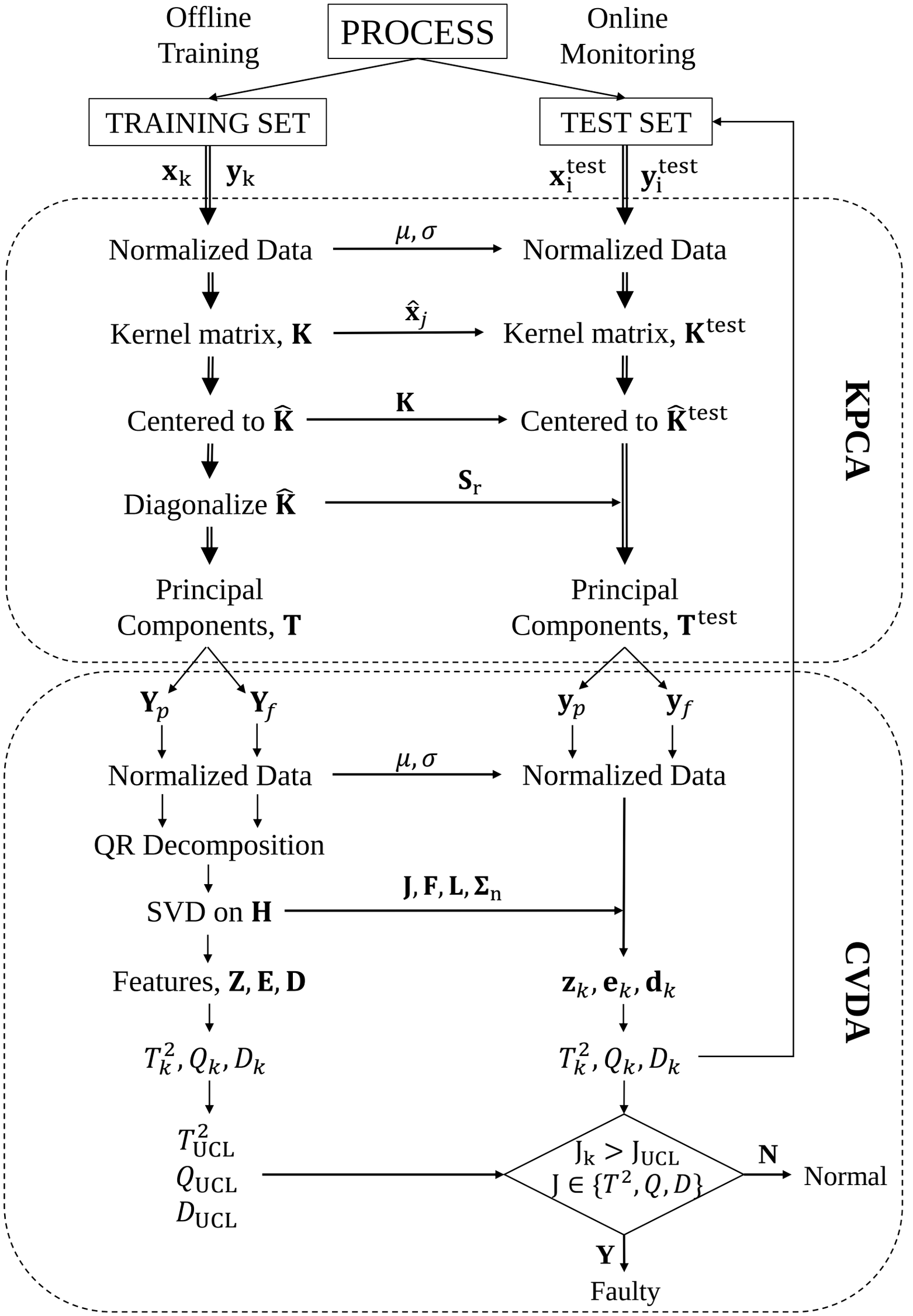
Zhang, Y., 2009. Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM. Chem. Eng. Sci. 64, 801–811. doi:10.1016/j.ces.2008.10.012.

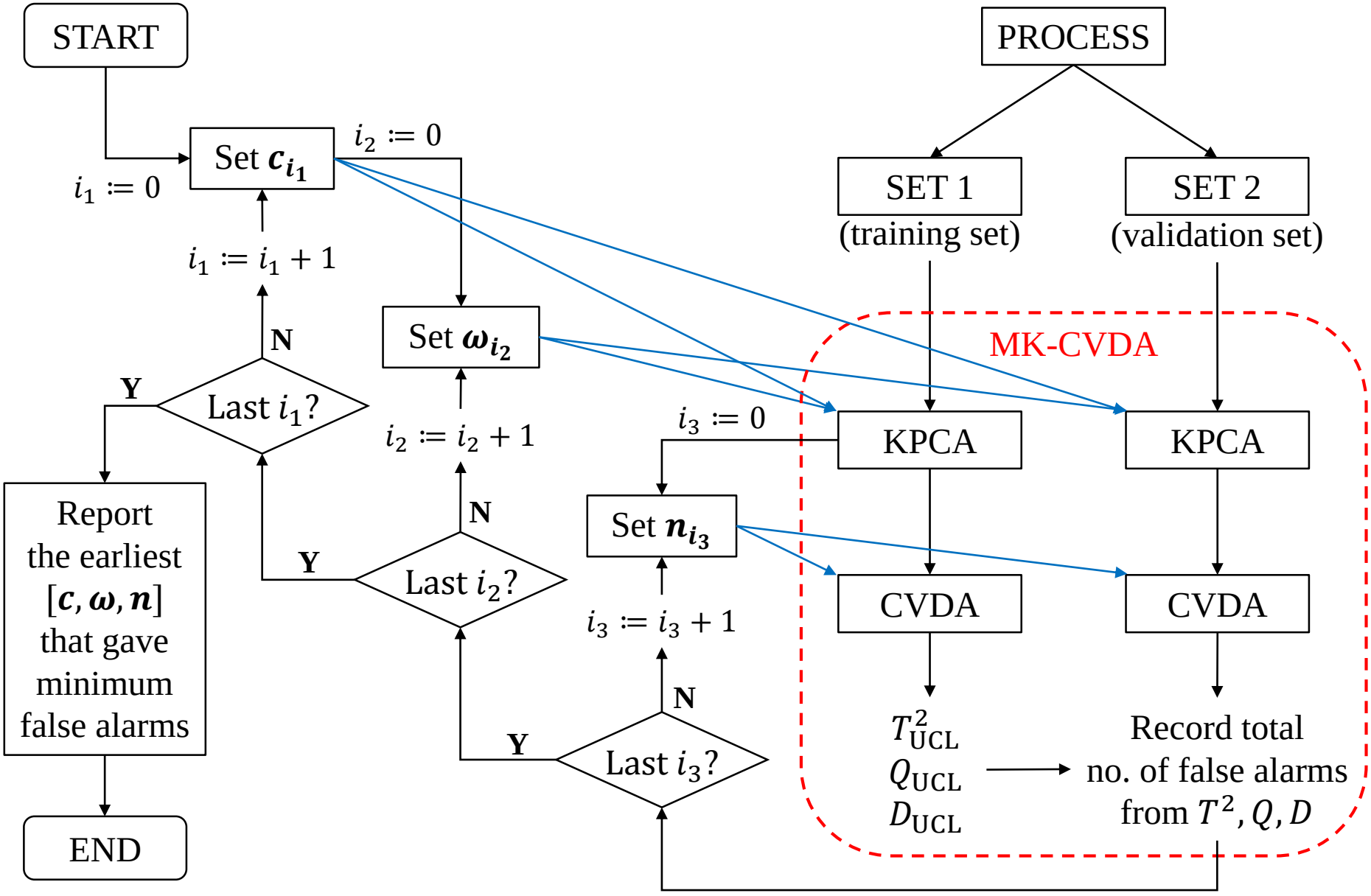
Zhang, Y., Zhang, Y., 2010. Process Monitoring, Fault Diagnosis and Quality Prediction Methods Based on the Multivariate Statistical Techniques. IETE Tech. Rev. 27, 406. doi:10.4103/0256-4602.62226.

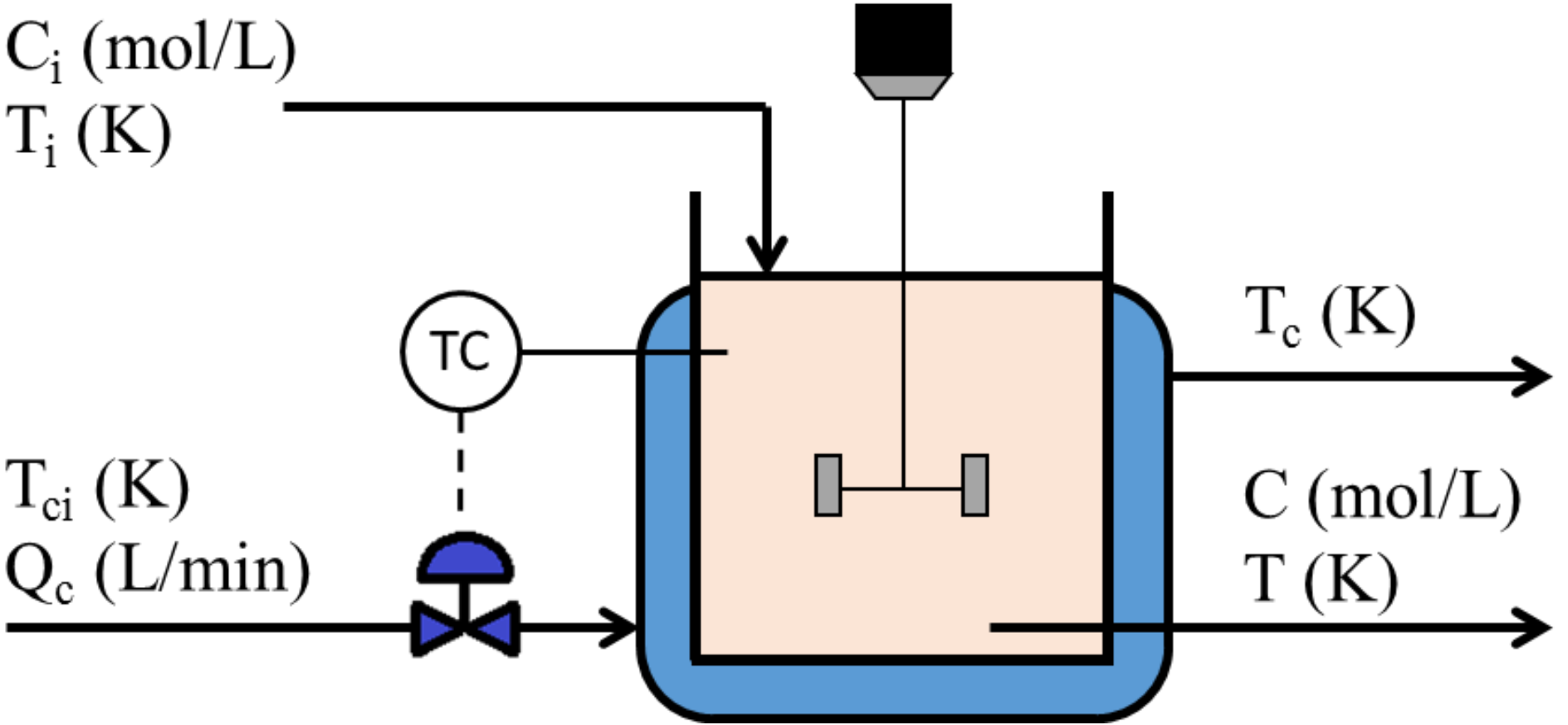
Zhong, Z., Carr, T.R., 2016. Application of mixed kernels function (MKF) based support vector regression model (SVR) for CO2 Reservoir oil minimum miscibility pressure prediction. Fuel 184, 590–603. doi:10.1016/j.fuel.2016.07.030.

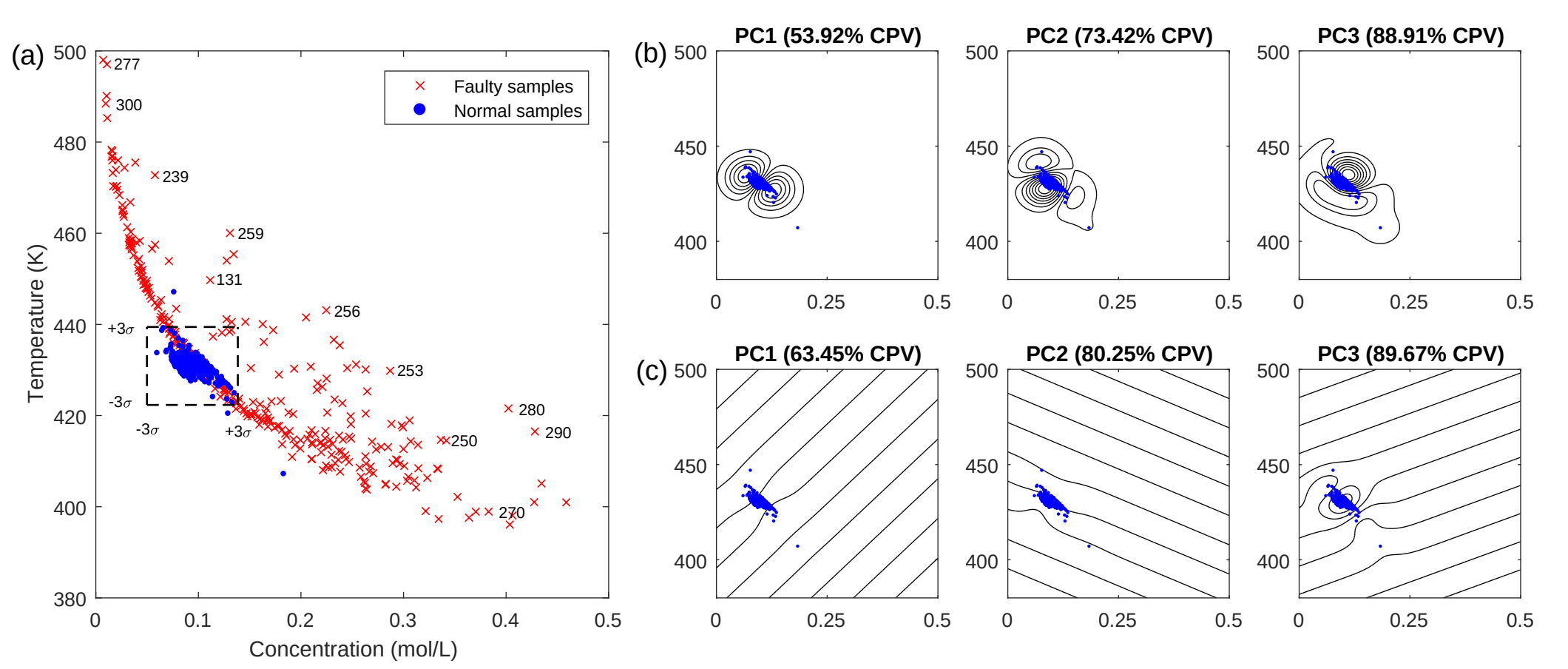
Zhu, X., Huang, Z., Tao Shen, H., Cheng, J., Xu, C., 2012. Dimensionality reduction by Mixed Kernel Canonical Correlation Analysis. Pattern Recognit. 45, 3003–3016. doi:10.1016/j.patcog.2012.02.007.


c=0.1
c=0.4
c=0.7
c=1.0
$K_g(\mathbf{x}, \mathbf{x}')$
x
(a)
d=1
d=2
d=3
d=4
$K_p(\mathbf{x}, \mathbf{x}')$
x
(b)

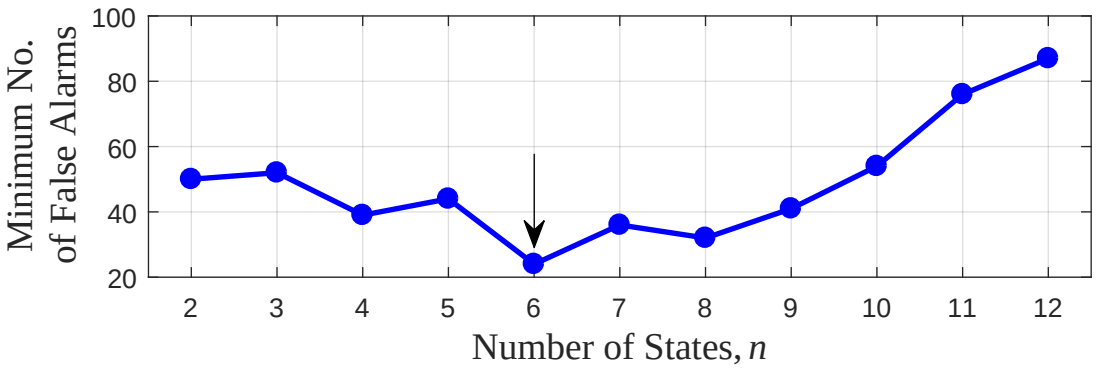

Offline Training
PROCESS
Online Monitoring
TRAINING SET
TEST SET
$\mathbf{x}_k$ $\mathbf{y}_k$
$\mathbf{x}_i^{test}$ $\mathbf{y}_i^{test}$
Normalized Data
μ, σ
Normalized Data
Kernel matrix, $\mathbf{K}$
$\hat{\mathbf{x}}_j$
Kernel matrix, $\mathbf{K}^{test}$
Centered to $\widehat{\mathbf{K}}$
$\mathbf{K}$
Centered to $\widehat{\mathbf{K}}^{test}$
Diagonalize $\widehat{\mathbf{K}}$
$\mathbf{S}_r$
Principal Components, $\mathbf{T}$
Principal Components, $\mathbf{T}^{test}$
KPCA
$\mathbf{Y}_p$ $\mathbf{Y}_f$
$\mathbf{y}_p$ $\mathbf{y}_f$
Normalized Data
μ, σ
Normalized Data
QR Decomposition
SVD on $\mathbf{H}$
$\mathbf{J}, \mathbf{F}, \mathbf{L}, \mathbf{\Sigma}_n$
Features, $\mathbf{Z}, \mathbf{E}, \mathbf{D}$
$\mathbf{z}_k, \mathbf{e}_k, \mathbf{d}_k$
T_k^2, Q_k, D_k
T_k^2, Q_k, D_k
T_{UCL}^2
Q_{UCL}
D_{UCL}
$J_k > J_{UCL}$
$J \in \{T^2, Q, D\}$
N
Normal
Y
Faulty
CVDA


START
PROCESS
Set c_{i_1}
$i_1 := 0$
$i_2 := 0$
SET 1
(training set)
SET 2
(validation set)
$i_1 := i_1 + 1$
N
Y
Last i_1?
Set ω_{i_2}
MK-CVDA
$i_2 := i_2 + 1$
$i_3 := 0$
KPCA
KPCA
Report the earliest $[c, \omega, n]$ that gave minimum false alarms
N
Y
Last i_2?
Set n_{i_3}
CVDA
CVDA
$i_3 := i_3 + 1$
N
Y
Last i_3?
T^2_{UCL}
Q_{UCL}
D_{UCL}
Record total no. of false alarms from T^2, Q, D
END

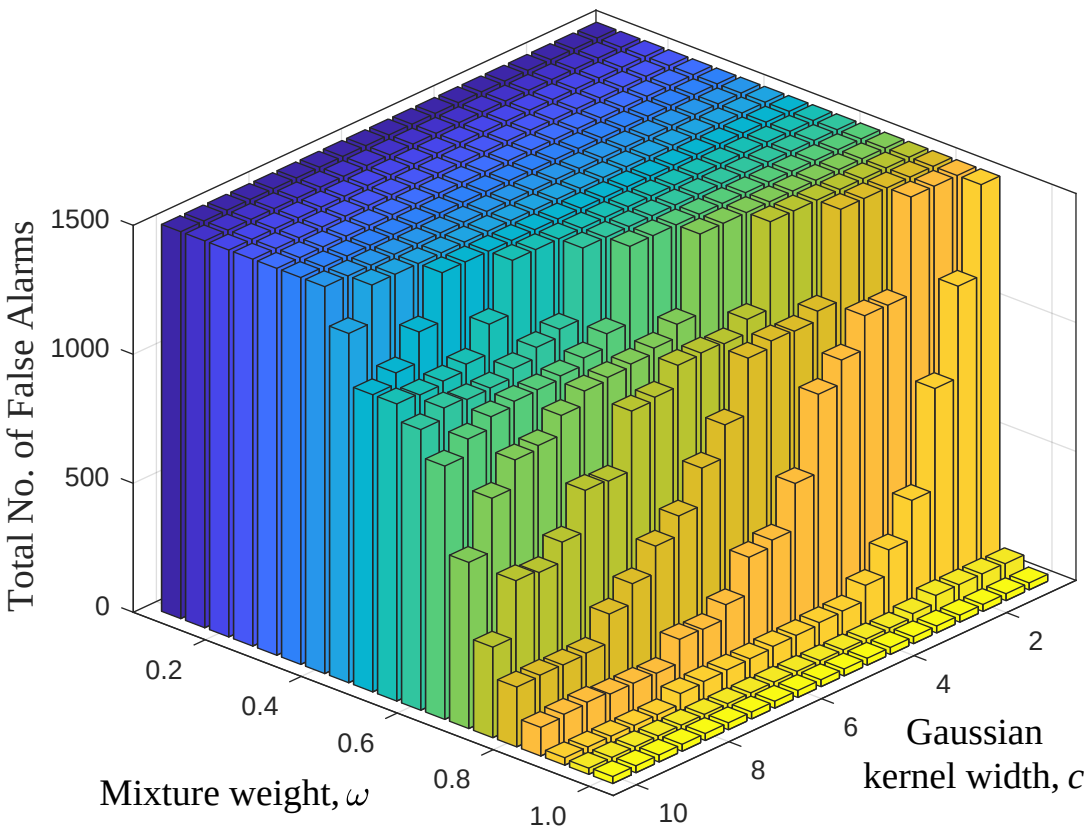

C_i (mol/L)
T_i (K)
TC
T_c (K)
T_{ci} (K)
Q_c (L/min)
C (mol/L)
T (K)

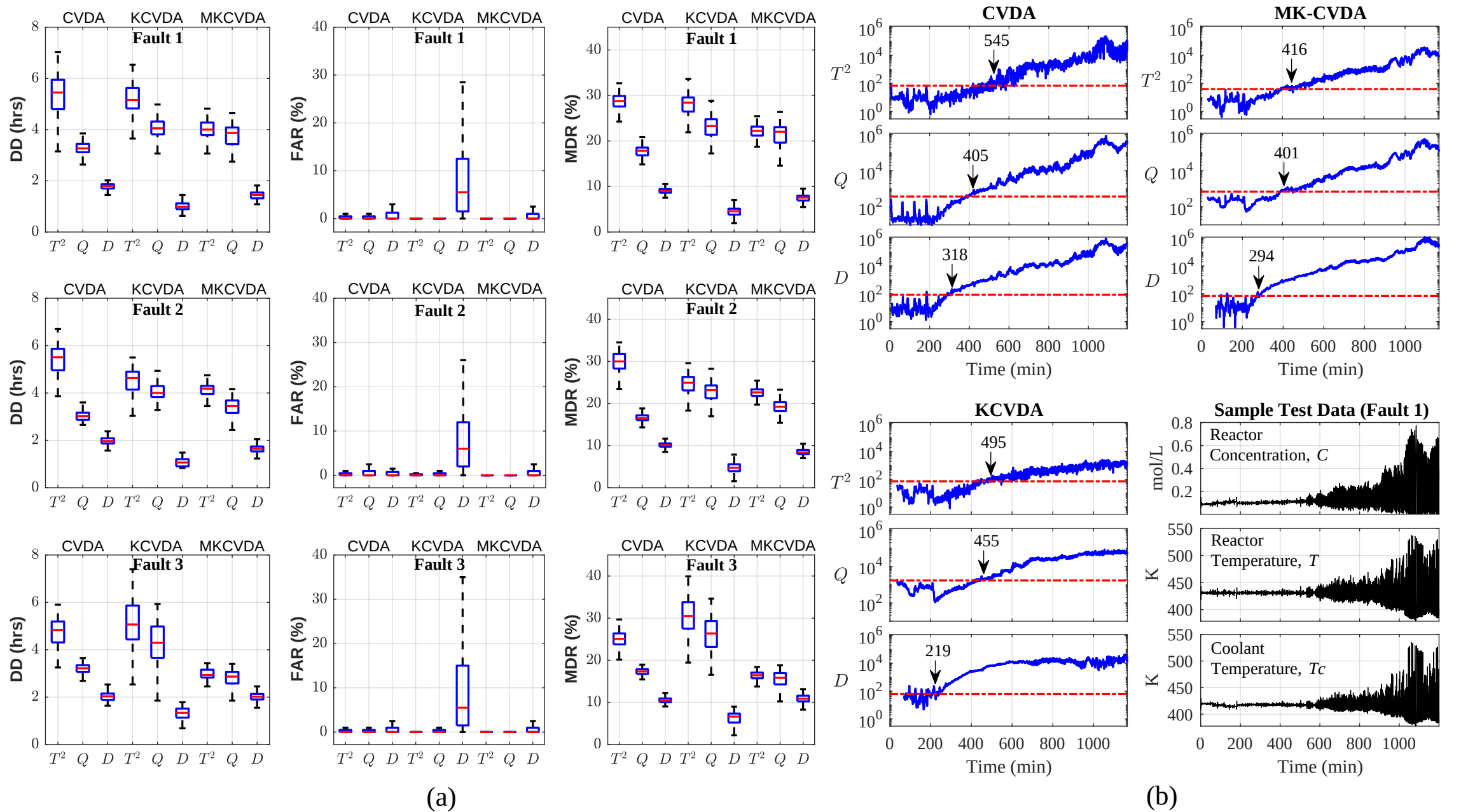

(a)
Faulty samples
Normal samples
277
300
239
259
131
256
253
280
290
250
270
+3σ
-3σ
-3σ
+3σ
Temperature (K)
Concentration (mol/L)
(b)
PC1 (53.92% CPV)
PC2 (73.42% CPV)
PC3 (88.91% CPV)
(c)
PC1 (63.45% CPV)
PC2 (80.25% CPV)
PC3 (89.67% CPV)

(a)


Minimum No. of False Alarms
100
80
60
40
20
2
3
4
5
6
7
8
9
10
11
12
Number of States, n

(b)


Total No. of False Alarms
1500
1000
500
0
0.2
0.4
0.6
0.8
1.0
Mixture weight, ω
2
4
6
8
10
Gaussian kernel width, c


CVDA
KCVDA
MKCVDA
Fault 1
Fault 2
Fault 3
DD (hrs)
FAR (%)
MDR (%)
T^2
Q
D
(a)
CVDA
MK-CVDA
KCVDA
545
416
405
401
318
294
495
455
219
Time (min)
Sample Test Data (Fault 1)
Reactor Concentration, C
mol/L
Reactor Temperature, T
Coolant Temperature, Tc
K
(b)