Data-Driven Incipient Fault Detection via Canonical Variate Dissimilarity and Mixed Kernel Principal Component Analysis

Wu, Ping; Ferrari, Riccardo M.G.; Liu, Yichao; Van Wingerden, Jan Willem

DOI
10.1109/TII.2020.3029900

Publication date
2021

Document Version
Final published version

Published in
IEEE Transactions on Industrial Informatics

Data-Driven Incipient Fault Detection via Canonical Variate Dissimilarity and Mixed Kernel Principal Component Analysis

Ping Wu, Riccardo M. G. Ferrari, Yichao Liu, and Jan-Willem van Wingerden, *Senior Member, IEEE*

***Abstract*—Incipient fault detection plays a crucial role in preventing the occurrence of serious faults or failures in industrial processes. In most industrial processes, linear, and nonlinear relationships coexist. To improve fault detection performance, both linear and nonlinear features should be considered simultaneously. In this article, a novel hybrid linear-nonlinear statistical modeling approach for data-driven incipient fault detection is proposed by closely integrating recently developed canonical variate dissimilarity analysis and mixed kernel principal component analysis (MKPCA) using a serial model structure. Specifically, canonical variate analysis (CVA) is first applied to estimate the canonical variables (CVs) from the collected process data. Linear features are extracted from the estimated CVs. Then, the canonical variate dissimilarity (CVD) which quantifies model residuals in the CVA state-subspace is calculated using the estimated CVs. To explore the nonlinear features, the nonlinear principal components are extracted as nonlinear features through performing MKPCA on CVD. Fault detection indices are formed based on Hotelling's T^2 as well as Q statistics from the extracted linear and nonlinear features. Moreover, kernel density estimation is utilized to determine the control limits. The effectiveness of the proposed method is demonstrated by the comparisons with other relevant methods via simulations based on a closed-loop continuous stirred-tank reactor process.**

***Index Terms*—Canonical variate analysis (CVA), dissimilarity analysis, incipient fault detection, kernel principal component analysis (KPCA), mixed kernel.**

Manuscript received June 28, 2020; accepted October 1, 2020. Date of publication October 9, 2020; date of current version May 3, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61703371, and in part by the Social Development Project of Zhejiang Provincial Public Technology Research under Grant LGF19F030004. Paper no. TII-20-3135. *(Corresponding author: Ping Wu.)*

Ping Wu is with the Department of Automation, Faculty of Mechanical Engineering and Automation, Zhejiang Sci-Tech University, Zhejiang 310018, China (e-mail: pingwu@zstu.edu.cn).

Riccardo M. G. Ferrari, Yichao Liu, and Jan-Willem van Wingerden are with the Delft Center for Systems and Control, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, 2628 CD Delft, Netherlands (e-mail: R.Ferrari@tudelft.nl; Y.Liu-17@tudelft.nl; J.W.vanWingerden@tudelft.nl).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.3029900

I. Introduction

RECENTLY, data-driven fault detection techniques, especially multivariate statistical process monitoring (MSPM) methods have attracted considerable interest from both the academic and industrial spheres. Compared with model-based or knowledge-based methods, MSPM methods are developed to operate exclusively on process data without detailed first-principle models or expert experience, which is usually infeasible or time-consuming to obtain in practice [1]–[3].

Widely used MSPM methods include principal component analysis (PCA), partial least squares (PLS), canonical variate analysis (CVA), [4]–[11]. A major limitation of PCA- and PLS-based approaches is that both PCA and PLS rely on the assumption that the process data are not time-dependent [12]. However, most real industrial processes are dynamic. Compared to PCA and PLS, CVA is a state-space based method that takes both serial correlation and relationship between correlated process variables into consideration. Therefore, CVA is more suitable for dynamic process modeling [13]–[15].

Although MSPM methods have been successfully applied in fault detection, dealing with incipient faults is still a major challenge. The main reason for this is that incipient faults often have small amplitudes and are slowly developing changes, as opposed to abrupt faults [16]. Incipient faults are easily compensated by feedback control during their initial stage [17]. Unfortunately, incipient faults can slowly affect the process behavior and gradually evolve into serious faults, even system failures. Thus, incipient fault detection plays a crucial role in the maintenance activities where timely and effective detection of incipient faults can avoid more serious consequences [18]. Conventional MSPM methods as mentioned above are not sensitive to incipient faults, resulting in a high missed detection rate (MDR) and long detection delay (DD) time.

To detect incipient faults, Harmouche *et al.* [19] combined Kullback–Leibler divergence (KLD) with PCA. A dissimilarity measure is established by comparing the probability density of each of the latent scores to a reference one using the KLD. In a similar work, Chen *et al.* [20] presented an incipient fault detection and diagnosis method based on KLD under probability-relevant PCA, where KLD and Bayesian inference is integrated. Another dissimilarity measure for process data called DISSIM method was proposed by Kano *et al.* [21]. DISSIM method is to evaluate the difference between distributions of data sets based on the Karhunen–Loeve expansion. Zhao *et al.* [22], [23] proposed a subspace distribution monitoring strategy to evaluate the changes of linear and nonlinear stationary and nonstationary distribution structures based on the DISSIM method for incipient fault detection. To observe the variation of process data statistics, Shang *et al.* [24] proposed recursive dynamic transformed component statistical analysis where the higher-order statistics of projected data are monitored from a sliding window of process data. Ji *et al.* [25] developed a generic fault detection index in MSPM by using moving average and exponentially weighted moving average for incipient fault detection. It should be noticed that these dissimilarity-based techniques require a large window width of samples for computing statistical patterns.

In recent work, Pilario *et al.* [26] proposed a method called canonical variate dissimilarity analysis (CVDA) to cope with incipient fault detection. In CVDA-based incipient fault detection the model residuals in the CVA state-subspace, namely canonical variate dissimilarity (CVD) between past-projected and future-projected canonical variables, are formed through traditional CVA. Then a detection index is defined as the squared Mahalanobis distance of the CVD for fault detection. Furthermore, the kernel density estimation (KDE) method was utilized to compute control limits. In [27], a combined index that combines Hotelling's T^2 statistic, Q statistic, and CVD-based statistic was developed. Furthermore, Pilario *et al.* [28] extended CVDA to nonlinear CVDA by preprocessing the original data with a kernel principal component analysis (KPCA) method. Then, CVDA was performed on the extracted nonlinear principal components (NPCs). A mixed kernel was adopted to enhance the interpolation and extrapolation abilities of single kernel-based learning. The method was referred to as MKCVDA in [28]. Since mixed kernel principal component analysis (MKPCA) was first performed, only nonlinear features are explored in MKCVDA.

Both linear and nonlinear relationships always coexist in complex industrial processes [29]–[31]. Using a single nonlinear model may not be optimal for statistical modeling in process monitoring and fault diagnosis [29]. A combined strategy would thus be preferable, by exploring linear and nonlinear features. Such a hybrid structure was successfully applied to describe the underlying relationship for time series forecasting [32]. Chen combined the linear and nonlinear statistical models to forecast time series with possibly nonlinear characteristics [33]. In [34], a linear model was first built via a projection algorithm, then a feedforward neural network was used to model the unmodeled dynamics. Recently, Deng *et al.* [29] integrated linear PCA and kernel PCA methods in a serial model structure to extract linear and nonlinear features. However, hybrid linear-nonlinear statistical modeling is still little investigated for incipient fault detection.

Motivated by the above discussions, we propose a novel data-driven fault detector using a hybrid linear–nonlinear statistical modeling approach. The main spirit of the proposed method is to use CVDA to build a linear dynamic model from process data and then extract nonlinear features from the CVD. This way, both linear and nonlinear features are simultaneously leveraged for fault detection. To extract the nonlinear features, neural networks and kernel-based methods are widely used and studied [34]. Compared to neural network methods, kernel-based methods have their foundation in the solid mathematical framework of reproducing kernel Hilbert spaces. Kernel methods yield convex optimization problems, can be used as universal nonlinear approximators, and require only moderate computational complexity [35]–[38]. Among the kernel-based methods, KPCA is a powerful technique, widely applied in process monitoring and fault diagnosis [1], [3], [39]–[41]. However, the commonly used Gaussian radial basis function (RBF) may suffer from overfitting problem, due to its lack of extrapolation ability, particularly while an inappropriate kernel width is selected [28], [42], [43]. The combination of RBF and polynomial kernels can provide enhanced modeling performance [44]. Following this idea, we adopt MKPCA to extract the nonlinear features from the obtained CVD for incipient fault detection. Moreover, five fault detection indices are designed by computing Hotelling's T^2 and Q statistics based on the extracted linear and nonlinear features. Therefore, here the proposed method is referred to as canonical variate dissimilarity mixed kernel principal component analysis (CVD-MKPCA).

CVD-MKPCA combines the merits of CVDA and MKPCA methods. Compared to the recently developed MKCVDA [28], CVD-MKPCA has two advantages. First, linear and nonlinear features are simultaneously extracted in a natural way. MKCVDA only considers nonlinear features, as the original data is first projected into a nonlinear high-dimensional space. In CVD-MKPCA, linear features are extracted by CVDA, and then MKPCA extracts the nonlinear features from CVD. A more reliable fault index can thus be derived for nonlinear dynamic processes, compared to MKCVDA. Second, the computational cost of the proposed CVD-MKPCA is lower than MKCVDA in the online monitoring stage, since two mixed kernel matrices are required to be computed for inputs and outputs in MKCVDA versus only one for CVD in CVD-MKPCA.

The main contributions of this article lie in the following:

1) A hybrid statistical modeling approach is presented by integrating CVDA and MKPCA in a serial model structure. Linear and nonlinear features are simultaneously extracted from process data for incipient fault detection.
2) An improved incipient fault detection performance can be attained for nonlinear dynamic processes. Furthermore, a lower computational cost is required, compared to the recently developed MKCVDA method.

Moreover, canonical correlation analysis (CCA) based fault detection methods have been developed for a variety of industrial applications [45]–[47]. These methods can be improved through utilizing the similar statistical data modeling framework proposed in this study.

The remainder of this article is structured as follows. The basic idea of CVDA-based incipient fault detection is described in the next section. Section III presents the proposed CVD-MKPCA method in detail. Section IV gives the case study description, results, and discussion. Finally, Section V concludes this article.

II. Brief Review of the CVDA

Denote $\boldsymbol{u}(k) \in \mathbb{R}^{n_u}$ and $\boldsymbol{y}(k) \in \mathbb{R}^{n_y}$ as the process inputs and outputs at time instant k. The past data vector $\boldsymbol{z}_p(k) \in$

$\mathbb{R}^{(n_u+n_y)p}$ containing the past inputs and outputs, and the future data vector $\boldsymbol{y}_f(k) \in \mathbb{R}^{n_y f}$ which consists of the future outputs are defined

$$\boldsymbol{z}_p(k) = \begin{bmatrix} \boldsymbol{u}(k-1) \\ \boldsymbol{u}(k-2) \\ \vdots \\ \boldsymbol{u}(k-p) \\ \boldsymbol{y}(k-1) \\ \boldsymbol{y}(k-2) \\ \vdots \\ \boldsymbol{y}(k-p) \end{bmatrix} \quad \boldsymbol{y}_f(k) = \begin{bmatrix} \boldsymbol{y}(k) \\ \boldsymbol{y}(k+1) \\ \vdots \\ \boldsymbol{y}(k+f-1) \end{bmatrix}$$

where p and f are the numbers of time lags in past and future data vectors $\boldsymbol{z}_p(k)$ and $\boldsymbol{y}_f(k)$, respectively.

Supposed that a training data set with N measurements of $\boldsymbol{u}(k)$ and $\boldsymbol{y}(k)$, $k = 1, 2, \ldots, N$ are collected under normal operating condition, the past and future Hankel matrices $\boldsymbol{Z}_p$ and $\boldsymbol{Y_f}$ are constructed from $\boldsymbol{z}_p(k)$ and $\boldsymbol{y}_f(k)$ for all $k \in [p+1, p+M]$ as follows:

$$\boldsymbol{Z}_p = \begin{bmatrix} \boldsymbol{z}_p(p+1) & \boldsymbol{z}_p(p+2) & \cdots & \boldsymbol{z}_p(p+M) \end{bmatrix} \quad (1)$$

$$\boldsymbol{Y}_f = \begin{bmatrix} \boldsymbol{y}_f(p+1) & \boldsymbol{y}_f(p+2) & \cdots & \boldsymbol{y}_f(p+M) \end{bmatrix} \quad (2)$$

where $M = N - p - f + 1$. The sample covariance matrices of the past and future vectors and cross-covariance matrix can be estimated

$$\boldsymbol{\Sigma}_{\text{pp}} = \frac{1}{M-1}\boldsymbol{Z}_p\boldsymbol{Z}_p^T \quad (3)$$

$$\boldsymbol{\Sigma}_{\text{ff}} = \frac{1}{M-1}\boldsymbol{Y}_f\boldsymbol{Y}_f^T \quad (4)$$

$$\boldsymbol{\Sigma}_{\text{fp}} = \frac{1}{M-1}\boldsymbol{Y}_f\boldsymbol{Z}_p^T. \quad (5)$$

The goal of CVA is to find the projection matrices $\boldsymbol{J}$ and $\boldsymbol{L}$ to maximize the correlation between $\boldsymbol{L}\boldsymbol{y}_f(k)$ and $\boldsymbol{J}\boldsymbol{z}_p(k)$, where $\boldsymbol{L}\boldsymbol{y}_f(k)$ and $\boldsymbol{J}\boldsymbol{z}_p(k)$ are also called canonical variables. Generally, the projection matrices $\boldsymbol{J}$ and $\boldsymbol{L}$ can be computed by performing singular value decomposition (SVD)

$$\boldsymbol{\Sigma}_{\text{ff}}^{-1/2}\boldsymbol{\Sigma}_{\text{fp}}\boldsymbol{\Sigma}_{\text{pp}}^{-1/2} = \mathcal{U}\mathcal{S}\mathcal{V}^T \quad (6)$$

where $\mathcal{U}$ and $\mathcal{V}$ are the matrices consisting of the left and right singular vectors, respectively. The diagonal matrix $\mathcal{S}$ consists of ordered singular values. From the result of SVD, the projection matrices $\boldsymbol{J}$ and $\boldsymbol{L}$ are formed by

$$\boldsymbol{J} = \mathcal{V}^T\boldsymbol{\Sigma}_{\text{pp}}^{-1/2} \quad (7)$$

$$\boldsymbol{L} = \mathcal{U}^T\boldsymbol{\Sigma}_{\text{ff}}^{-1/2}. \quad (8)$$

Further, the canonical variables $\boldsymbol{c}(k)$ and $\boldsymbol{x}(k)$ at time instant k are obtained

$$\boldsymbol{c}(k) = \boldsymbol{L}\boldsymbol{y}_f(k) \quad (9)$$

$$\boldsymbol{x}(k) = \boldsymbol{J}\boldsymbol{z}_p(k). \quad (10)$$

In CVA-based fault detection method [48], [49], the state vectors $\boldsymbol{x}_n(k)$ are extracted from the past data vectors to represent the process status

$$\boldsymbol{x}_n(k) = \boldsymbol{J}_n\boldsymbol{z}_p(k) \quad (11)$$

where $\boldsymbol{J}_n = \mathcal{V}_n^T\boldsymbol{\Sigma}_{\text{pp}}^{-1/2} \in \mathbb{R}^{n\times(n_u+n_y)p}$. $\mathcal{V}_n$ contains the first n columns of $\mathcal{V}$. The value for n can be determined by analyzing the plot of the singular values curve from the result of SVD in (6). In [26], n is selected as the point where a knee appears in the singular values curve.

Additionally, the model residual vectors $\boldsymbol{e}(k)$ which span the residual subspace is derived

$$\boldsymbol{e}(k) = (\boldsymbol{I} - \mathcal{V}_n\mathcal{V}_n^T)\boldsymbol{\Sigma}_{\text{pp}}^{-1/2}\boldsymbol{z}_p(k) \quad (12)$$

where $\boldsymbol{I}$ is the identity matrix of appropriate dimension.

Remark 1: CVA is usually employed as a standard method for system identification where the state space vector is different from the $\boldsymbol{x}_n$ in (11) [50]. Particularly, the estimation of the state vector from (11) is biased in the closed-loop case. However, in the process of monitoring and fault diagnosis framework, (11) only builds the vector of the canonical variables for residuals generation. As pointed out in [50], as far as the collected process data do cover the major process operation scenarios, $\boldsymbol{x}_n$ can be used for process monitoring and fault diagnosis.

Two fault detection indices including Hotelling's T_s^2 and Q_s statistics are computed at time instant k

$$T_s^2(k) = \boldsymbol{x}_n(k)\boldsymbol{x}_n(k)^T \quad (13)$$

$$Q_s(k) = \boldsymbol{e}(k)\boldsymbol{e}(k)^T. \quad (14)$$

Here, T_s^2 measures the variations of state vectors $\boldsymbol{x}_n(k)$, while Q_s measures the variations of model residual vectors $\boldsymbol{e}(k)$.

It is noticed that the predictability of future canonical variables from past canonical variables can effectively reflect the small shifts in process data. To detect incipient faults such as decay in process parameters, sensor drifts, the CVD between the past-projected and future-projected canonical variables is employed in CVDA [26]. The CVD $\boldsymbol{d}_n(k)$ at time instant k is defined as follows:

$$\boldsymbol{d}_n(k) = \boldsymbol{L}_n\boldsymbol{y}_f(k) - \mathcal{S}_n\boldsymbol{J}_n\boldsymbol{z}_p(k) \quad (15)$$

where $\boldsymbol{L}_n = \mathcal{U}_n^T\boldsymbol{\Sigma}_{\text{ff}}^{-1/2} \in \mathbb{R}^{n\times n_y f}$. $\mathcal{U}_n$ contains the first n columns of $\mathcal{U}$. $\mathcal{S}_n$ consists of the n largest singular values $\mathcal{S}_n = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. For CCA-based methods [45]–[47], it is noted that the residuals are generated in a similar way as (15) for fault detection.

As presented in [26], [50], the covariance of $\boldsymbol{d}_n$ can be estimated by

$$\boldsymbol{\Sigma}_{\text{dd}} = \boldsymbol{I} - \mathcal{S}_n\mathcal{S}_n^T. \quad (16)$$

To detect incipient faults a fault detection index T_{dc}^2 is introduced, based on the squared Mahalanobis distance of $\boldsymbol{d}_n$ [26]

$$T_{\text{dc}}^2(k) = \boldsymbol{d}_n(k)^T\boldsymbol{\Sigma}_{\text{dd}}^{-1}\boldsymbol{d}_n(k). \quad (17)$$

KDE method is often employed to determine the upper control limits (UCLs) [13], particularly for nonlinear or non-Gaussian distributed process data. In CVDA-based fault detection, KDE is utilized to estimate the probability distributions of T_s^2, Q_s, and T_{dc}^2. The widely used kernel function in KDE is the Gaussian kernel function which is defined by

$$K(g) = \frac{1}{\sqrt{2\pi}} \exp^{-g^2/2}. \tag{18}$$

Given a specific significance level α, the UCL J_{UCL} can be calculated by solving the following problem:

$$P(J < J_{\text{UCL}}) = \int_{-\infty}^{J_{\text{UCL}}} \frac{1}{Mh} \sum_{k=1}^{M} K\left(\frac{J - J(k)}{h}\right) dJ = \alpha \tag{19}$$

where $J(k), k = 1, 2, \ldots, M$ represents the samples of fault detection index $J \in \{T_s^2, Q_s, T_{\text{dc}}^2\}$ under normal operating conditions and h is the kernel bandwidth. J_{UCL} represents the corresponding UCL $J_{\text{UCL}} \in \{T_{\text{UCL},s}^2, Q_{\text{UCL},s}, T_{\text{UCL,dc}}^2\}$. More details for KDE can be found in [13]. In the proposed CVD-MKPCA based fault detection, we also adopt KDE to determine UCLs.

Remark 2: The parametric approach of probability distributions estimation relies on the assumption of specific probability distributions. KDE is a nonparametric one. Thus, KDE has more flexibility for the determination of UCLs. A drawback of KDE is that the kernel function and its parameters should be selected appropriately. In KDE, the problem of finding the appropriate bandwidth h is a key concern. Several approaches have been proposed to find the optimal bandwidth such as the least squares cross-validation, contrast methods [51]. In [52], a simple estimation of bandwidth was developed from minimizing the approximation of the mean integrated squared error

$$h = 1.06\sigma M^{-0.2} \tag{20}$$

where σ is the standard deviation of the established fault detection indices using the collected process data under normal conditions. It has been proved that this selection method (20) can provide a promising performance in CVA-based fault detection methods [13], [26].

In the online monitoring stage, fault detection indices T_s^2, Q_s, and T_{dc}^2 at every sampling instant are calculated using (13), (14), and (17). For CVA-based fault detection, the occurrence of a fault is detected when any one of T_s^2, Q_s exceeds its corresponding UCL, $T_{\text{UCL},s}^2, Q_{\text{UCL},s}$, respectively. For CVDA-based fault detection, the occurrence of a fault is detected when any one of $T_s^2, Q_s, T_{\text{dc}}^2$ exceeds its corresponding UCL, $T_{\text{UCL},s}^2, Q_{\text{UCL},s}, T_{\text{UCL,dc}}^2$, respectively.

III. Proposed Method

Although the fault detection index T_{dc}^2 has proved its effectiveness for incipient fault detection as shown in [26], it can only evaluate the variations of linear features in process data. Nonlinear features usually occur in the residuals of the linear model [29], [32] and their effect cannot be separated by that of other normally occurring uncertainties. This leads to high UCLs and, thus, low detectability of small faults such as incipient ones at early stages. To extract these nonlinear features and improve detectability, it is worthwhile to further analyze the CVD, which

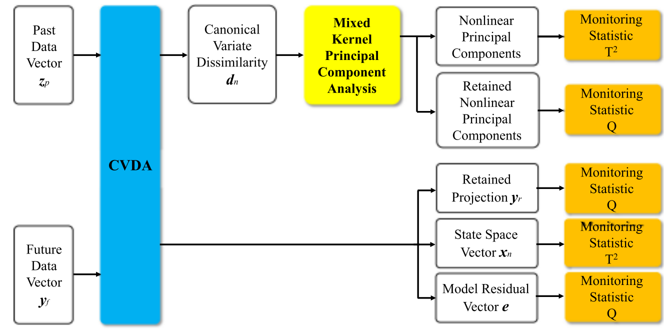


Fig. 1. Schematic diagram of CVD-MKPCA statistical modeling.

is the model residuals in the CVA state-subspace, through nonlinear features extraction methods. Given the main objective of this article and the simplicity of kernel-based methods, MKPCA is applied for this goal. Along with this concept, MKPCA is performed to examine the nonlinear features for fault detection in the proposed CVD-MKPCA method. The proposed method consists of two main steps, as shown in Fig. 1.

In Section II, the derivation of CVD has been introduced. Besides $\boldsymbol{x}_n$ and $\boldsymbol{e}$, the residuals $\boldsymbol{y}_r$ of $\boldsymbol{y}$ onto the state subspace can also be used to construct a fault index, where

$$\boldsymbol{y}_r(k) = (\boldsymbol{I} - \mathcal{U}_n\mathcal{U}_n^T)\boldsymbol{\Sigma}_{ff}^{-1/2}\boldsymbol{y}_f(k). \tag{21}$$

Similarly to what has been done before, the Q_y statistic can be introduced

$$Q_y(k) = \boldsymbol{y}_r(k)\boldsymbol{y}_r(k)^T. \tag{22}$$

As shown in Fig. 1, fault indices T_s^2, Q_s, Q_y are established from linear features through CVDA model. To extract the nonlinear features, $\boldsymbol{d}_n$ is further investigated. Assumed that $\boldsymbol{d}_n$ is implicitly mapped onto a high-dimensional feature space $\mathcal{F}$ through a nonlinear function map $\phi(\boldsymbol{d}_n) : \mathbb{R}^n \to \mathcal{F}$, then the sample covariance of high-dimensional features can be calculated

$$\mathcal{C} = \frac{1}{M}\sum_{i=1}^{M} \phi(\boldsymbol{d}_n(i))\phi(\boldsymbol{d}_n(i))^T \tag{23}$$

where $\sum_{i=1}^{M} \phi(\boldsymbol{d}_n(i)) = \boldsymbol{0}$. In KPCA, the loading vector $\boldsymbol{\nu}$ in the high-dimensional feature space can be computed by solving the below eigenvalue problem

$$\lambda\boldsymbol{\nu} = \mathcal{C}\boldsymbol{\nu} = \frac{1}{M}\sum_{i=1}^{M}(\phi(\boldsymbol{d}_n(i))\boldsymbol{\nu}^T)\phi(\boldsymbol{d}_n(i)) \tag{24}$$

where $\lambda > 0$ and $\boldsymbol{\nu} \neq \boldsymbol{0}$. However, since $\phi(\boldsymbol{d}_n(i))$ can not be expressed explicitly, the eigenvalue problem (24) cannot be directly solved via eigenvalue decomposition. It is known that $\boldsymbol{\nu}$ lies in the subspace spanned by $\phi(\boldsymbol{d}_n(i))$. Thus, there exist some $\boldsymbol{\gamma}$ where $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_M]^T$ such that

$$\boldsymbol{\nu} = \sum_{i=1}^{M} \gamma_i \phi(\boldsymbol{d}_n(i)). \tag{25}$$

Substitute (25) into (24), and multiply $\phi(\boldsymbol{d}_n(j))$ with the left of both sides in (24),

$$\lambda\phi(\boldsymbol{d}_n(j))\boldsymbol{\nu} = \phi(\boldsymbol{d}_n(j))\frac{1}{M}\sum_{i=1}^{M}\phi(\boldsymbol{d}_n(i))\phi(\boldsymbol{d}_n(i))^T\boldsymbol{\nu}. \quad (26)$$

Moreover, the kernel matrix $\boldsymbol{K} \in \mathbb{R}^{M\times M}$ with kernel function κ is defined as $\boldsymbol{K}_{i,j} = \kappa(\boldsymbol{d}_n(i), \boldsymbol{d}_n(j)) = \langle\phi(\boldsymbol{d}_n(i)), \phi(\boldsymbol{d}_n(j))\rangle, i, j = 1, 2, \ldots, M$ where $\langle\cdot,\cdot\rangle$ represents the inner-product operator. Then, the eigenvalue problem (26) can be expressed in terms of the dot products of two mappings to derive the eigenvectors $\boldsymbol{\gamma}$

$$M\lambda\boldsymbol{\gamma} = \boldsymbol{K}\boldsymbol{\gamma}. \quad (27)$$

The detailed explanation, discussion, and implementation of KPCA can readily be found in the literature [28], [39], [53].

Based on Mercer's theorem, the inner products are to be calculated in a possible infinite-dimensional space, known as the Hilbert space [53]. An appropriate kernel function should make the kernel matrix $\boldsymbol{K}$ a positive semidefinite. Two representative kernel functions, the Gaussian RBF and polynomial kernel are widely used in process monitoring and fault diagnosis. The RBF is defined by

$$\kappa_{\text{rbf}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{s}\right) \quad (28)$$

where s is the kernel width. The polynomial kernel is given as

$$\kappa_{\text{poly}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i\boldsymbol{x}_j^T + 1)^{\mu} \quad (29)$$

where μ is the user-defined degree of the polynomial.

For RBF kernel, only the data points in the neighborhood of the test points are affected. The RBF kernel has good interpolation ability but lacks extrapolation ability. Thus, it is considered a local kernel. The overfitting problem may occur in the learning while a single RBF kernel is employed. On the other hand, the polynomial kernel can be considered as a global kernel [43]. The polynomial kernel has good extrapolation ability but poor interpolation ability. In [44], the mixtures of kernels were proposed by combining RBF and polynomial kernels to enhance the modeling performance of the support vector machine for regression. To improve the performance of incipient fault detection, a mixed kernel was applied in [28]. Inspired by these ideas, MKPCA is adopted in our study to extract nonlinear features from CVD.

The mixed kernel is constructed by using a convex combination of RBF and polynomial kernels

$$\kappa_{\text{mix}} = \beta\kappa_{\text{rbf}} + (1-\beta)\kappa_{\text{poly}} \quad (30)$$

where $\beta(0 \le \beta \le 1)$ is the mixing coefficient to balance the interpolation and extrapolation abilities.

Assumed that the mixed kernel $\boldsymbol{K}_{\text{mix}}$ has been centered [54], then (26) is equivalent to

$$M\lambda\boldsymbol{\gamma} = \boldsymbol{K}_{\text{mix}}\boldsymbol{\gamma}. \quad (31)$$

For the mixed kernel, three important parameters should be determined including the degree of the polynomial μ, the kernel width s, and the mixing coefficient β. A large value of s would weaken its interpolation ability of RBF kernel but strengthen the extrapolation ability. Similarly, an appropriate μ should be determined by considering the tradeoff between interpolation and extrapolation abilities. Meanwhile, the mixing coefficient β is of importance to achieve the optimal performance of the learning task. Although several optimization methods such as genetic algorithm, particle swarm optimization, have been developed for finding the optimal kernel parameters, they require much effort and computational costs. A practical method is using a grid search strategy to determine the optimal parameters of a mixed kernel [28]. To find the optimal parameters, we use false alarm rate (FAR) as a criterion in the offline training stage. FAR is the ratio of the false alarming samples over all the fault-free samples. The optimal parameters should be chosen to obtain a FAR as lower as possible. Since μ is an integer, it is easy to choose through cross-validation. In the case study, $\mu = 2$ is adopted. The other two parameters s and β are chosen through the results of the grid search.

Remark 3: While the mixing coefficient β is set as 1, the mixed kernel $\boldsymbol{K}_{\text{mix}}$ becomes a single RBF kernel $\boldsymbol{K}_{\text{rbf}}$. Usually, a regularization term is imposed to deal with the ill-conditioned kernel matrix which is constructed by a single RBF kernel

$$\lambda\boldsymbol{\gamma} = \left(\frac{1}{M}\boldsymbol{K}_{\text{rbf}} + \zeta\boldsymbol{I}\right)\boldsymbol{\gamma} \quad (32)$$

where ζ is the regularization parameter. A cross-validation can be used to determine ζ. In KPCA-based fault detection methods using RBF kernel function, s usually is specified as $500l$ [29], where l is the dimension of process variables.

For a test $\boldsymbol{d}_n(k)$, its retained NPCs $t_{\text{cm},i}(k), i = 1, 2, \ldots, m$ which are with the first m eigenvalues are extracted by

$$t_{\text{cm},i}(k) = \sum_{j=1}^{M}\boldsymbol{\gamma}_j^i\kappa_{\text{mix}}(\boldsymbol{d}_n(j), \boldsymbol{d}_n(k)). \quad (33)$$

Denote $\mathbf{t}_{\text{cm},m}(k) = [t_{\text{cm},1}(k), \ldots, t_{\text{cm},m}(k)]$. A fault detection index is formed by using Hotelling's T^2 statistic to monitor the variation of retained NPCs

$$T_{\text{dm}}^2(k) = \mathbf{t}_{\text{cm},m}(k)\boldsymbol{\Lambda}_{\text{cm}}^{-1}\mathbf{t}_{\text{cm},m}(k)^T \quad (34)$$

where $\boldsymbol{\Lambda}_{\text{cm}}$ is the sample covariance of $\mathbf{t}_{\text{cm},m}$.

The rest NPCs can be monitored by establishing the following Q_{dm} statistic as in [39]:

$$Q_{\text{dm}}(k) = \mathbf{t}_{\text{cm},M}\mathbf{t}_{\text{cm},M}^T - \mathbf{t}_{\text{cm},m}(k)\mathbf{t}_{\text{cm},m}(k)^T \quad (35)$$

where $\mathbf{t}_{\text{cm},M} = [t_{\text{cm},1}, \ldots, t_{\text{cm},M}]$.

Remark 4: Similar to linear PCA, the number of retained NPCs m can be determined by using the cumulative percent variance (CPV) method. In the case study, the selection of m is to achieve the predetermined percentage variation of 98%.

The UCLs $T_{\text{UCL,dm}}^2$ and $Q_{\text{UCL,dm}}$ of T_{dm}^2 and Q_{dm} are obtained by the KDE method, similarly to what was done in Section II. Under the CVD-MKPCA based fault detection framework, all five indices T_s^2, Q_s, Q_y, T_{dm}^2 and Q_{dm}, will be used to detect incipient faults. The fault detection logic is that a fault is detected when any one of $T_s^2, Q_s, Q_y, T_{\text{dm}}^2, Q_{\text{dm}}$ exceeds its corresponding UCL, $T_{\text{UCL},s}^2, Q_{\text{UCL},s}, Q_{\text{UCL},y}, T_{\text{UCL,dm}}^2, Q_{\text{UCL,dm}}$, respectively.

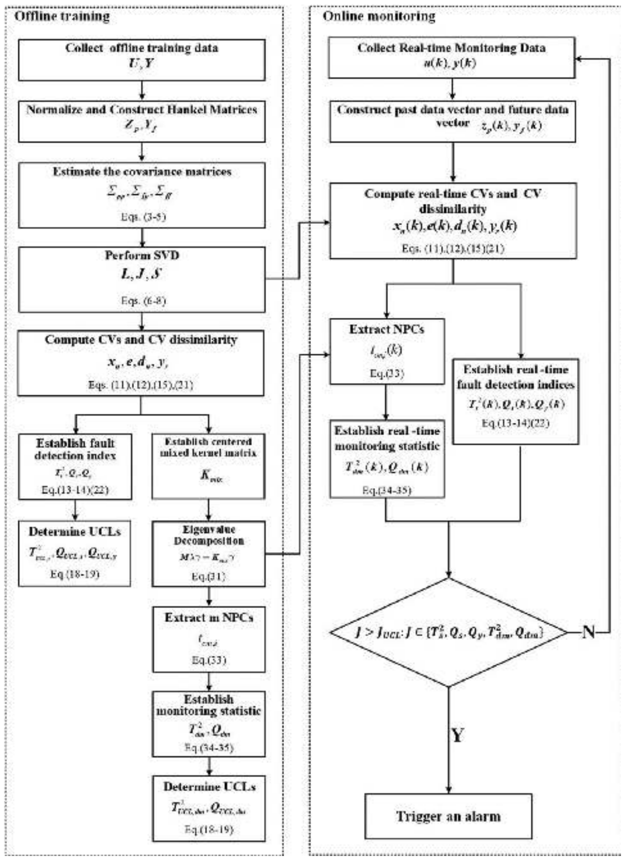


Fig. 2. Procedure of the proposed CVD-MKPCA based fault detection.

In summary, the procedure of the proposed CVD-MKPCA based incipient fault detection is described as follows. 1) In the offline training stage, the CVD-MKPCA model is built from the collected process data and the corresponding UCLs are established through KDE. 2) In the online monitoring stage, the real-time fault detection indices are computed with the continuous collection of a moving window of samples of length $p + f$. The process is determined to be normal or faulty by comparing real-time indices with their respective UCLs. The detailed procedure of the proposed CVD-MKPCA method is depicted in Fig. 2.

IV. Case Study

In this section, a closed-loop CSTR process is used to verify the performance of the proposed CVD-MKPCA based incipient fault detection method. The studied CSTR process is particularly designed by Pilario *et al.* for simulating incipient faults [26]. Fig. 3 plots the diagram of the closed-loop CSTR process. The mechanism of the CSTR process is mainly described by the following equations:

$$\begin{cases} \frac{dC}{dt} = \frac{Q}{V}(C_i - C) - akC + v_1 \\ \frac{dT}{dt} = \frac{Q}{V}(T_i - T) - a\frac{(\Delta H_r)kC}{\rho C_p} - b\frac{UA}{\rho C_p V}(T - T_c) + v_2 \\ \frac{dT_c}{dt} = \frac{Q_c}{V_c}(T_{ci} - T_c) + b\frac{UA}{\rho_c C_{pe} V_c}(T - T_c) + v_3 \end{cases} \tag{36}$$

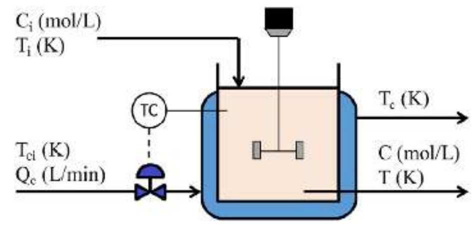


Fig. 3. Diagram of the closed-loop CSTR [26].

TABLE I
Model Parameters of the CSTR Process

Parameter	Description	Value
Q_c	Coolant Inlet flow rate	100 L/min
V	Tank volume	150L
V_c	Jacket volume	10L
ΔH_r	Reaction heat	-2×10^5cal/mol
UA	Heat transfer coefficient	7×10^5cal/min
k_0	Pre-exponential kinetic constant	7.2×10^{10}/min
E/R	Activation energy	1×10^4K
ρ	Fluid density	1000g/L
ρ_c	Fluid density	1000g/L
C_P	Fluid heat capacity	1cal/g/K
C_{pc}	Fluid heat capacity	1cal/g/K

TABLE II
Description of the Incipient Faults in the CSTR Process

No.	Description	ξ	Type
1	Sensor drift $T = T + \xi t$	0.005	Additive
2	Catalyst decay $a = a_0 e^{-\xi t}$	5×10^{-4}	Multiplicative
3	Fouling $b = b_0 e^{-\xi t}$	1×10^{-3}	Multiplicative

where C_i is the concentration of the reactant. T_i and T_{ci} are the temperature of the reactant and inlet temperature of the coolant, respectively. v_i are process noise. $k = k_0 exp^{-E/RT}$ is an Arrhenius-type rate. Due to the existence of Arrhenius-type rate k, it can be observed that there are linear and nonlinear relationships in the closed-loop CSTR process as shown in (36). The model parameters of the CSTR process are given in Table I. Similar to [26], we select $\boldsymbol{u} = [C_i\ T_i\ T_{ci}]$ and $\boldsymbol{y} = [C\ T\ T_c\ Q_c]$. The CSTR simulation model in Matlab Simulink used in this study can be downloaded from https://www.mathworks.com/matlabcentral/fileexchange/66189-feedback-controlled-cstr-process-for-fault-simulation.

For evaluating the fault detection performance, three typical incipient faults are considered [28]. These incipient fault scenarios are described in Table II. To simulate the saturation faults, a and b are decayed from 1.00 at the normal operation to 0. It can be used to simulate incipient faults such as catalyst decay and heat transfer fouling. Another incipient fault is a sensor drift in T.

The sampling interval for all variables is 1 min. The offline training dataset is collected during 20 h under normal operation stimulated by randomly varying inputs $\boldsymbol{u}$ around their nominal values every 1 h. Therefore, 1200 samples are generated for training models. These samples are correlated and non-Gaussian distributed owing to the dynamic and nonlinear behavior of the closed-loop CSTR process. Each fault scenario also has 1200

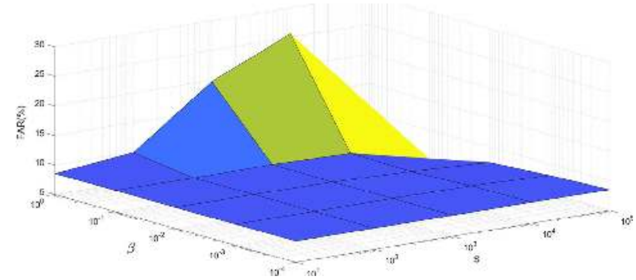


Fig. 4. FAR results versus [s, β] where s is the kernel width of RBF and β is the mixing coefficient.

samples which are generated during 20 h of process operation. The fault is injected after 200 min under each fault scenario. A 99.9% control limit is used to determine the UCLs for all methods.

For comparison, CVA T^2 and Q [13] (which are the same T_s^2 and Q_s in CVD-MKPCA method), CVDA D [26], CVDA T_c^2 [27], KCDVA D, and MKCVDA D [28] are employed. Besides, CVD-KPCA T_d^2 is adopted for comparison. CVD-KPCA T_d^2 is derived through a single Gaussian RBF kernel.

A. CVD-MKPCA Training

We use two fault-free data sets as the training data to build the CVD-MKPCA model and determine the related parameters. As discussed in [26], the numbers of time lag p and f can be determined by using auto-correlation analysis. n is then selected as the point where a knee appears by screening the plot of the singular value curve. In this study, we adopt the same values of p, f, n as in [26] for comparison, where $p = f = 5$, and $n = 8$. With the sets $s = 10^i, \beta = 10^{-j} (i, j = 1, 2, \ldots, 5)$ and predefined $\mu = 2$, the FARs against choices of $[s, \beta]$ are plotted in Fig. 4. Through Fig. 4, the parameters s and β are chosen as $s = 100$ and $\beta = 0.01$. By calculating the CPV from the result of (31), m is set as 40. For CVD-KPCA with a single Gaussian RBF kernel, the kernel width is set as 4000, and the regularization coefficient ζ is set to 0.0001 in (32) through cross-validation. For CVDA, KCVDA, and MKCVDA, the parameters are determined with a similar procedure described in [28].

B. CVD-MKPCA Monitoring

Fault 1 is a sensor drift. As shown in Fig. 5(a) and (b), it can be found that the amplitudes of the change of C and T are relatively small. Notwithstanding that the DDs of all indices are long, the MDR of CVD-MKPCA $T_{\rm dm}^2$ is lower than other indices. From Fig. 5(c)–(j), it can be observed that the detection time by CVD-MKPCA $T_{\rm dm}^2$ is 440 min, while other indices require more time to detect the occurrence of Fault 1 such as 480 min for MKCVDA D, 580 min for KCVDA D and 445 min for CVD-KPCA T_d^2. For Fault 2, it is a catalyst decay fault. In the beginning, the variations of process variables such as C and T are not obvious. After a few hours, the deviation of process variables between under normal and abnormal conditions would gradually become huge as shown in Fig. 6 (a) and (b). As shown in Fig. 6(i), CVD-KPCA T_d^2 changes around a constant after 700 min. However, Fault 2 is actually becoming more severe. The reason is that there may exist an overfitting problem while using a single RBF kernel in CVD-KPCA. On the contrary, this issue is addressed by introducing the mixed kernel. As shown in Fig. 6(j), it can be seen that CVD-MKPCA $T_{\rm dm}^2$ can follow the variation trend of the severity of Fault 2. And the detection time is 300 min for CVD-MKPCA $T_{\rm dm}^2$. It is longer than the DD of MKCVDA D (290 min). However, CVD-MKPCA $T_{\rm dm}^2$ obtains shorter detection time than most of the indices in this case. For Fault 3, the fouling parameter b would gradually become zero. It can be found that the performance of CVD-MKPCA $T_{\rm dm}^2$ is better than other indices. Especially, CVD-MKPCA $T_{\rm dm}^2$ can detect Fault 3 much earlier than MKCVDA D where the detection time is 285 min for CVD-MKPCA $T_{\rm dm}^2$ and 305 min for MKCVDA D. As plotted in Fig. 7(a) and (b), there is a spike around 1000 min in C and T. However, CVD-KPCA T_d^2 can not detect this severe change due to the overfitting problem as shown in Fig. 7(i). Similar to Fault 2, CVD-MKPCA $T_{\rm dm}^2$ works well to capture the trends of process variables C and T due to the adoption of mixed kernel , as shown in Fig. 7(j).

To evaluate the performance robustly, a Monte Carlo simulation of 15 realizations with different random seeds for the process noises, measurement noises, and input disturbances for each fault scenario. Three indices are utilized to quantify the fault detection performance: 1) DD, the elapsed time since the fault has been injected until it is detected—to confirm the occurrence of incipient faults, the detection time is defined as the first time after ten consecutive alarms were raised as in [28]; 2) FAR; and 3) MDR, the ratio of the undetected samples over all the faulty samples. For a robust comparison, 15 test data sets are generated for each fault scenario. In Table III, the medians of DD, FAR, and MDR across 15 faulty data sets in each fault scenario are listed. To make the comparison of DD time more clear, the unit of DD is converted to hours.

As presented in Table III, it can be observed that the fault detection indices relying on linear features have similar performance except for CVA T_s^2 and Q_s. Although CVA Q_s has the same level of DDs and MDRs as CVDA D, CVDA T_c^2, its FARs are higher. In general, for fault detection indices using linear features, the monitoring index based on CVD can provide better performance than other indices. It can also be found that the performance using the fault detection indices based on nonlinear features is superior over these indices based on linear features. For example, the MDRs and DDs of KCVDA D, MKCVDA D, and CVD-MKPCA $T_{\rm dm}^2$ are much lower and shorter than CVDA D and CVDA T_c^2. The CSTR process used in this study includes both linear and nonlinear relationships. Compared to CVDA and MKCVDA, CVD-MKPCA can obtain a more accurate statistical model using a serial model structure. The T^2 statistic of the NPCs with dominant eigenvalues can capture the change of the process status more accurately. CVD-MKPCA $T_{\rm dm}^2$ can derive lower MDRs and shorter DDs for Fault 1 and Fault 3 scenarios, and the same level of performance for Fault 2, compared to KCVDA D and MKCVDA D. From the data in Table III, it is also observed that CVD-MKPCA $Q_{\rm dm}$ can provide better DDs and MDRs for Fault 2 and Fault 3 scenarios. However, like CVA Q_s, the FARs of CVD-MKPCA $Q_{\rm dm}$ are higher than other

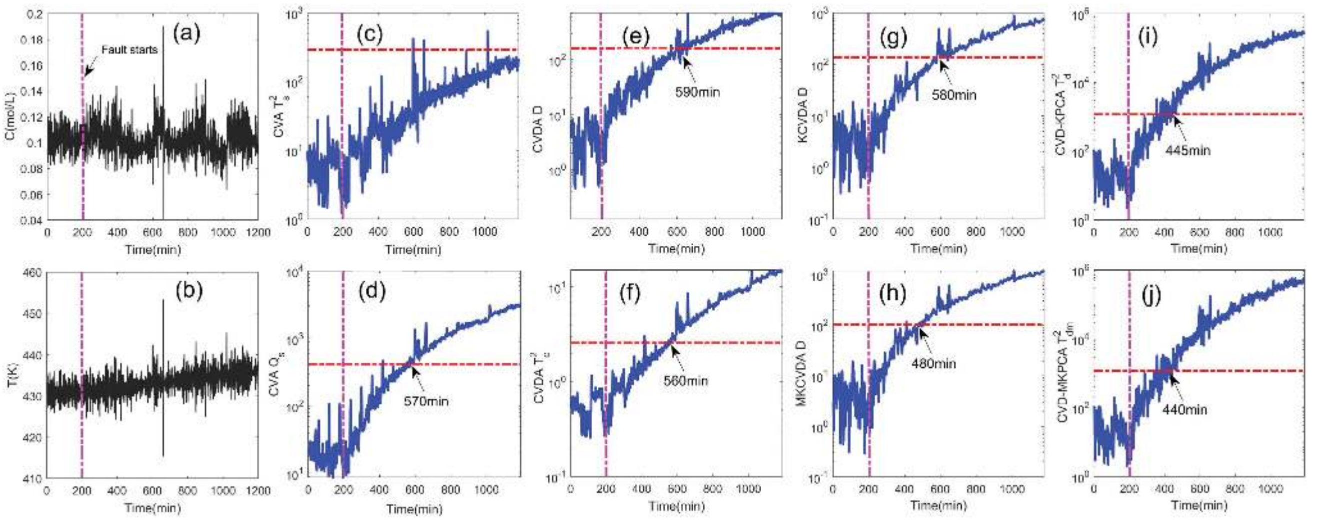


Fig. 5. Trends of sample at Fault 1 condition: (a) C and (b) T; monitoring charts: (c) CVA T_s^2 (d) CVA Q_s (e) CVDA D (f) CVDA T_c^2 (g) KCVDA D (h) MKCVDA D (i) CVD-KPCA T_d^2 (j) CVD-MKPCA T_{dm}^2. Legend: Red Dash dot - UCL; Solid - statistical index at Fault 1 condition; Pink Dash dot - start of fault.

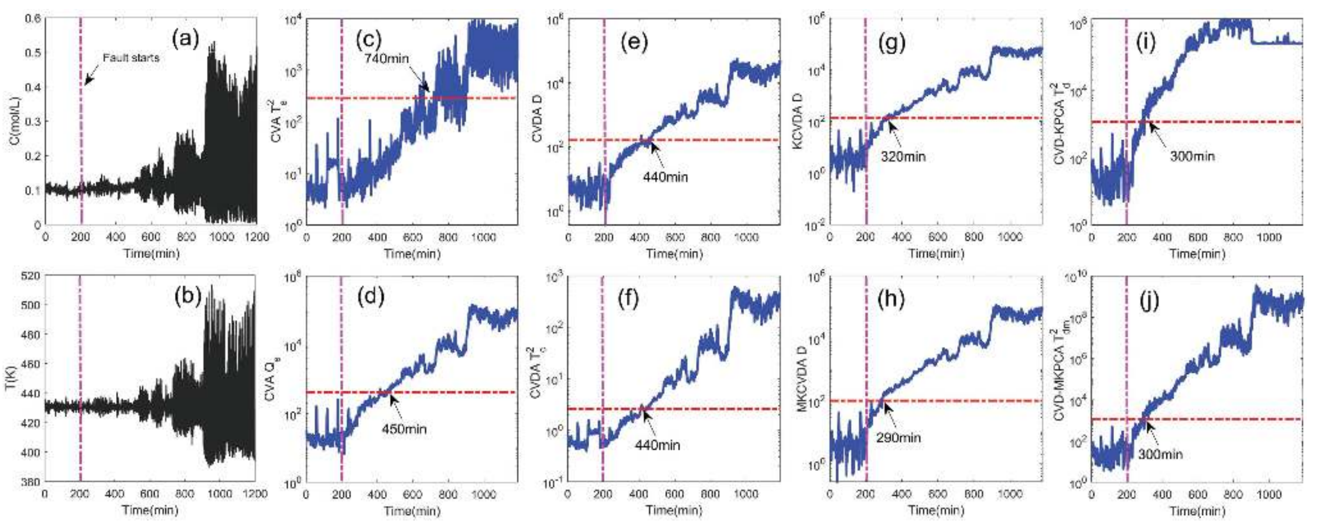


Fig. 6. Trends of sample at Fault 2 condition: (a) C and (b) T; monitoring charts: (c) CVA T_s^2 (d) CVA Q_s (e) CVDA D (f) CVDA T_c^2 (g) KCVDA D (h) MKCVDA D (i) CVD-KPCA T_d^2 (j) CVD-MKPCA T_{dm}^2. Legend: Red Dash dot - UCL; Solid - statistical index at Fault 2 condition; Pink Dash dot - start of fault.

TABLE III
COMPARISON OF FAULT DETECTION PERFORMANCE FOR THE INCIPIENT FAULTS IN CSTR PROCESS†

Fault No.	CVDA	CVDA	KCVDA	MKCVDA	CVD-KPCA	CVD-MKPCA				
	D	T_c^2	D	D	T_d^2	T_s^2	Q_s	Q_y	T_{dm}^2	Q_{dm}
1	4.38 [a]	4.48	3.65	3.32	3.32	7.61	4.40	7.98	3.21	1.92
	0 [b]	0	0	0	0	0	1.04	0	0	4.71
	23.00 [c]	24.30	20.50	19.11	17.20	42.3	24.00	43.5	17.10	24.0
2	2.13	2.65	1.55	1.50	1.70	6.12	2.90	6.05	1.70	1.28
	0	0	0	0	0	0	0	0	0	4.71
	11.40	14.50	8.42	8.00	8.89	34.50	15.30	36.5	8.20	5.50
3	2.11	2.26	1.82	1.69	1.45	7.40	2.27	11.85	1.45	1.20
	0	0	0	0	0	0	0.52	0	0	8.37
	11.00	11.60	10.20	9.50	7.10	39.20	11.90	76.30	7.00	4.90

†All results were medians from the results across 15 faulty data sets monitored in each fault scenario. [a] First row: DDs (DD, hours) consistently for ten consecutive sampling times; [b] Second row: (FAR, %); [c] Third row: (MDR,%).

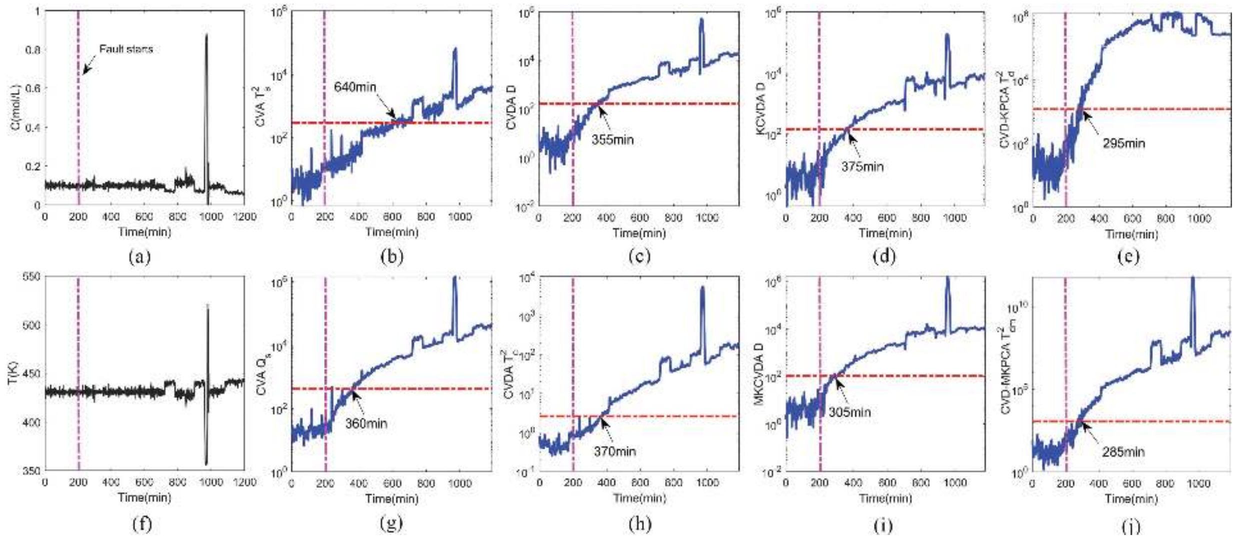


Fig. 7. Trends of sample at Fault 3 condition: (a) C and (b) T; monitoring charts: (c) CVA T_s^2 (d) CVA Q_s (e) CVDA D (f) CVDA T_c^2 (g) KCVDA D (h) MKCVDA D (i) CVD-KPCA T_d^2 (j) CVD-MKPCA T_{dm}^2. Legend: Red Dash dot - UCL; Solid - statistical index at Fault 3 condition; Pink Dash dot - start of fault.

indices. Although CVD-MKPCA Q_{dm} can not provide reliable results due to high FARs. The FARs of CVD-MKPCA T_{dm}^2 are much lower than CVD-MKPCA Q_{dm} as listed in Table III. Compared to CVD-MKPCA Q_{dm}, CVD-MKPCA T_{dm}^2 is more reliable.

In the following discussion, the comparative results are analyzed between CVD-MKPCA T_{dm}^2 and other indices based on nonlinear features such as KCVDA D, MKCVDA D, and CVD-KPCA T_d^2. Compared to other indices, the DD and MDR derived by CVD-MKPCA T_{dm}^2 are superior for Fault 1 and Fault 3 scenarios. And its FARs are zero for all faults. Despite MKCVDA D and KCVDA D can provide slightly better performance for Fault 2 scenario than CVD-MKPCA T_{dm}^2, CVD-MKPCA T_{dm}^2 can still outperform over other fault detection indices.

Based on the results listed in Table III, it is shown that CVD-KPCA T_d^2 and CVD-MKPCA T_{dm}^2 can provide better performance than KCVDA D and MKCVDA D for Fault 1 and Fault 3 scenarios. In summary, it can be concluded that the combination of CVDA and MKPCA via a serial model structure is more effective for incipient fault detection for nonlinear dynamic processes, compared to CVDA and MKCVDA. As shown in Fig. 6(i) and (j) and Fig. 7(i) and (j), it can be seen that CVD-MKPCA T_{dm}^2 is a more reliable index than CVD-KPCA T_d^2 for detecting incipient faults. Nonetheless, CVD-MKPCA T_{dm}^2 is the most powerful index for incipient fault detection among the comparable indices in terms of combined FARs, DDs, and MDRs.

The computational cost should also be a concern in real-time fault detection, particularly while the kernel-based methods are introduced. In order to compare the computational costs of the proposed CVD-MKPCA method with other kernel-based methods such as KCVDA and MKCVDA, we list the elapsed time of the establishment of online fault detection indices in Table IV. The simulation environment is under Matlab 2019a with Intel Core i7-8750H CPU @2.20 GHz and 32 GB RAM. As listed

TABLE IV
Comparison of Computation Time in the Online Monitoring Phase

	KCVDA D	MKCVDA D	CVD-MKPCA T_{dm}^2
Computation time(s)	0.0018	0.0088	0.0042

in Table IV, the computation time of MKCVDA D is 0.0088 s. Since only a single kernel is adopted in calculating KCVDA D, the computation time is shorter than MKCVDA D as listed in Table IV. On the other hand, the computation time of calculating CVD-MKPCA T_{dm}^2 is 0.0042 s. As analyzed in Section III, Only one kernel matrix is needed to compute CVD-MKPCA T_{dm}^2 in the online monitoring stage. The computation time of calculating CVD-MKPCA T_{dm}^2 is shorter than MKCVDA D.

V. Conclusion

In this article, a novel data-driven incipient fault detection method using CVDA and MKPCA in a serial model structure was proposed. Except for the linear features extracted from CVDA, nonlinear PCs were extracted from the CVD between past-projected and future-projected canonical variables. The proposed CVD-MKPCA takes both the advantages of CVDA and MKPCA. Fault detection indices using Hotelling's T^2 and Q statistics were established based on the extracted linear and nonlinear features for incipient fault detection. The UCLs were determined using KDE. Simulation results have confirmed the superior performance of the proposed method over the related techniques. It can also be noticed that although CVD-MKPCA T_{dm}^2 can provide better performance than other indices, a further study on the utilization of all the extracted features or statistics is suggested. Additionally, this article mainly focused on fault detection. Fault identification and diagnosis can be developed using the proposed CVD-MKPCA statistical modeling framework for incipient fault monitoring in the future.

Acknowledgment

The authors would like to thank the anonymous reviewers for their comments and suggestions.

References

[1] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, 2012.

[2] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques-part ii: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015.

[3] S. Yin, H. Gao, and O. Kaynak, "Data-driven control and process monitoring for industrial applications-part ii," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 583–586, Jan. 2015.

[4] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.

[5] Q. Jiang, X. Yan, and B. Huang, "Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and bayesian inference," *IEEE Trans. Ind. Electron.*, vol. 63, no. 1, pp. 377–386, Jan. 2016.

[6] B. Zhou, H. Ye, H. Zhang, and M. Li, "Process monitoring of iron-making process in a blast furnace with PCA-based methods," *Control Eng. Pract.*, vol. 47, pp. 1–14, 2016.

[7] G. Wang and S. Yin, "Quality-related fault detection approach based on orthogonal signal correction and modified PLS," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 398–405, Apr. 2015.

[8] M. Mansouri, M. N. Nounou, and H. N. Nounou, "Multiscale kernel PLS-based exponentially weighted-GLRT and its application to fault detection," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 1, pp. 49–58, Feb. 2019.

[9] Y. Cao, Y. Hu, X. Deng, and X. Tian, "Quality-relevant batch process fault detection using a multiway multi-subspace CVA method," *IEEE Access*, vol. 5, pp. 23 256–23 265, 2017.

[10] S. Stubbs, J. Zhang, and J. Morris, "Fault detection in dynamic processes using a simplified monitoring-specific CVA state space modelling approach," *Comput. Chem. Eng.*, vol. 41, pp. 77–87, 2012.

[11] B. Jiang and R. D. Braatz, "Fault detection of process correlation structure using canonical variate analysis-based correlation features," *J. Process Control*, vol. 58, pp. 131–138, 2017.

[12] Z. Chen, S. X. Ding, K. Zhang, Z. Li, and Z. Hu, "Canonical correlation analysis-based fault detection methods with application to alumina evaporation process," *Control Eng. Pract.*, vol. 46, pp. 51–58, 2016.

[13] P. Odiowei and Y. Cao, "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations," *IEEE Trans. Ind. Informat.*, vol. 6, no. 1, pp. 36–45, 2010.

[14] R. T. Samuel and Y. Cao, "Kernel canonical variate analysis for nonlinear dynamic process monitoring," *IFAC-PapersOnLine*, vol. 48, no. 8, pp. 605–610, 2015.

[15] C. Ruiz-Cárcel, L. Lao, Y. Cao, and D. Mba, "Canonical variate analysis for performance degradation under faulty conditions," *Control Eng. Pract.*, vol. 54, pp. 70–80, 2016.

[16] R. Isermann, "Model-based fault-detection and diagnosis – status and applications," *Annu. Rev. Control*, vol. 29, no. 1, pp. 71–85, 2005.

[17] M. A. Demetriou and M. M. Polycarpou, "Incipient fault diagnosis of dynamical systems using online approximators," *IEEE Trans. Autom. Control*, vol. 43, no. 11, pp. 1612–1617, Nov. 1998.

[18] S. Gautam, P. K. Tamboli, V. H. Patankar, K. Roy, and S. P. Duttagupta, "Sensors incipient fault detection and isolation using kalman filter and kullback-leibler divergence," *IEEE Trans. Nucl. Sci.*, vol. 66, no. 5, pp. 782–794, May 2019.

[19] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part i," *Signal Process.*, vol. 94, pp. 278–287, 2014.

[20] H. Chen, B. Jiang, and N. Lu, "An improved incipient fault detection method based on Kullback-Leibler divergence," *ISA Trans.*, vol. 79, pp. 127–136, 2018.

[21] M. Kano, S. Hasebe, I. Hashimoto, and H. Ohno, "Statistical process monitoring based on dissimilarity of process data," *AlChE J.*, vol. 48, no. 6, pp. 1231–1240, 2002.

[22] C. Zhao and F. Gao, "A sparse dissimilarity analysis algorithm for incipient fault isolation with no priori fault information," *Control Eng. Pract.*, vol. 65, pp. 70–2, 2017.

[23] C. Zhao and B. Huang, "Incipient fault detection for complex industrial processes with stationary and nonstationary hybrid characteristics," *Ind. Eng. Chem. Res.*, vol. 57, no. 14, pp. 5045–5057, 2018.

[24] J. Shang, M. Chen, H. Ji, and D. Zhou, "Recursive transformed component statistical analysis for incipient fault detection," *Automatica*, vol. 80, pp. 313–327, 2017.

[25] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient fault detection with smoothing techniques in statistical process monitoring," *Control Eng. Pract.*, vol. 62, pp. 11–21, 2017.

[26] K. E. S. Pilario and Y. Cao, "Canonical variate dissimilarity analysis for process incipient fault detection," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5308–5315, Dec. 2018.

[27] X. Li, X. Yang, Y. Yang, I. Bennett, A. Collop, and D. Mba, "Canonical variate residuals-based contribution map for slowly evolving faults," *J. Process Control*, vol. 76, pp. 87–97, 2019.

[28] K. E. S. Pilario, Y. Cao, and M. Shafiee, "Mixed kernel canonical variate dissimilarity analysis for incipient fault monitoring in nonlinear dynamic processes," *Comput. Chem. Eng.*, vol. 123, pp. 143–154, 2019.

[29] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Nonlinear process fault diagnosis based on serial principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 560–572, Mar. 2018.

[30] S. J. Qin and L. H. Chiang, "Advances and opportunities in machine learning for process data analytics," *Comput. Chem. Eng.*, vol. 126, pp. 465–473, 2019.

[31] Q. Jiang and X. Yan, "Parallel PCA–KPCA for nonlinear process monitoring," *Control Eng. Pract.*, vol. 80, pp. 17–25, 2018.

[32] G. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[33] K.-Y. Chen, "Combining linear and nonlinear model in forecasting tourism demand," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10368–10376, 2011.

[34] Y. Zhang, T. Chai, and D. Wang, "An alternating identification algorithm for a class of nonlinear dynamical systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1606–1617, Jul. 2017.

[35] S. Van Vaerenbergh, "Kernel methods for nonlinear identification, equalization and separation of signals," Ph.D. dissertation, Dept. Commun. Eng., Univ. Cantabria, Santander, Spain, Feb. 2010.

[36] K. Zhong, M. Han, T. Qiu, and B. Han, "Fault diagnosis of complex processes using sparse kernel local fisher discriminant analysis," *IEEE Trans. Neural Netw. Learn. Syst*, vol. 31, no. 5, pp. 1581–1591, May 2020.

[37] L. Cai, X. Tian, and S. Chen, "Monitoring nonlinear and non-gaussian processes using gaussian mixture model-based weighted kernel independent component analysis," *IEEE Trans. Neural Netw. Learn. Syst*, vol. 28, no. 1, pp. 122–135, Jan. 2017.

[38] J. Huang and X. Yan, "Quality relevant and independent two block monitoring based on mutual information and KPCA," *IEEE Trans. Ind. Informat.*, vol. 64, no. 8, pp. 6518–6527, Aug. 2017.

[39] J.-M. Lee, C. 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, 2004.

[40] J. Ni, C. Zhang, and S. X. Yang, "An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBS," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1960–1971, Jul. 2011.

[41] C.-Y. Cheng, C.-C. Hsu, and M.-C. Chen, "Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes," *Ind. Eng. Chem. Res.*, vol. 49, no. 5, pp. 2254–2262, 2010.

[42] K. E. Pilario, M. Shafiee, Y. Cao, L. Lao, and S.-H. Yang, "A review of kernel methods for feature extraction in nonlinear process monitoring," *Processes*, vol. 8, no. 1, 2019.

[43] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

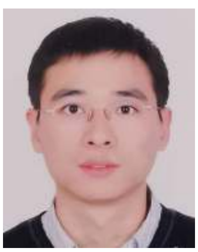
[44] G. F. Smits and E. M. Jordaan, "Improved SVM regression using mixtures of kernels," in *Proc. Int. Joint Conf. Neural Netw. IJCNN'02 (Cat. No.02CH37290)*, vol. 3, 2002, pp. 2785–2790.

[45] Q. Jiang, S. Ding, Y. Wang, and X. Yan, "Data-driven distributed local fault detection for large-scale processes based on ga-regularized canonical correlation analysis," *IEEE Trans. Ind. Electron.*, vol. 64, pp. 8148–8157, Oct. 2017.

[46] Z. Chen, Y. Cao, S. X. Ding, K. Zhang, T. Koenings, T. Peng, C. Yang, and W. Gui, "A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2710–2720, May 2019.

[47] Y. Liu, B. Liu, X. Zhao, and M. Xie, "A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6478–6486, Aug. 2018.

[48] B. C. Juricek, D. E. Seborg, and W. E. Larimore, “Fault detection using canonical variate analysis,” *Ind. Eng. Chem. Res.*, vol. 43, no. 2, pp. 458–474, 2004.
[49] B. Jiang, D. Huang, X. Zhu, F. Yang, and R. D. Braatz, “Canonical variate analysis-based contributions for fault identification,” *J. Process Control*, vol. 26, pp. 17–25, 2015.
[50] S. X. Ding, *Data Driven Design of Fault Diagnosis and Fault Tolerant Control Systems*. Newyork, NY, USA: Springer, 2014.
[51] A. Mugdadi and I. A. Ahmad, “A bandwidth selection for kernel density estimation of functions of random variables,” *Comput. Statist. Data Anal.*, vol. 47, no. 1, pp. 49–62, 2004.
[52] S. Xiaoping and A. Sonali, “Kernel density estimation for an anomaly based intrusion detection system,” in *Proc. World Congr. Comput. Sci. Comput. Eng. Appl. Comput.*, 2006, Art. no. 161.
[53] B. Schölkopf, A. Smola, and K. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.
[54] R. T. Samuel and Y. Cao, “Nonlinear process fault detection and identification using kernel PCA and kernel density estimation,” *Syst. Sci. Control Eng.*, vol. 4, no. 1, pp. 165–174, 2016.



Ping Wu received the B.S. and Ph.D. degrees in control theory and control engineering from Zhejiang University, Hangzhou, China, in 2003 and 2009, respectively.

He is currently an Associate Professor with the Faculty of Mechanical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou, China. His major research interests include fault diagnosis, machine learning, and industrial intelligence.

Riccardo M. G. Ferrari received the Laurea degree (cum laude and printing honours) in electronic engineering. and the Ph.D. degree in information engineering from University of Trieste, Trieste, Italy, in 2004 and 2009, respectively.

He is currently a Marie Curie alumnus and currently an Assistant Professor with the Delft Center for Systems and Control, Delft University of Technology, The Netherlands. His research interests include wind power fault tolerant control and fault diagnosis and attack detection in large-scale cyber-physical systems, with applications to electric vehicles, cooperative autonomous vehicles and industrial control systems.

Dr. Ferrari is the recipient of the 2005 Giacomini Award of the Italian Acoustic Society and he obtained the 2nd place in the Competition on Fault Detection and Fault Tolerant Control for Wind Turbines during IFAC 2011. Furthermore, he was awarded an Honorable Mention for the Pauk M. Frank Award at the IFAC SAFEPROCESS in 2018 and won an Airbus Award at IFAC 2020 for the best contribution to the competition on Aerospace Industrial Fault Detection. He has hold both academical and industrial R&D positions, in particular as Researcher in the field of process instrumentation and control for the steel-making sector.

Yichao Liu received the B.S. degree (hons.) in marine technology from Dalian Maritime University, Dalian, China, in 2014 and the Ph.D. degree in environmental science and engineering from Tsinghua University, Beijing, China, in 2018.

He is currently a Postdoctoral Researcher and Marie-Curie fellow at the Delft University of Technology, Delft, The Netherlands, in the Delft Center for Systems and Control. His current research interests include dynamical modelling, wind turbine control, and fault-tolerant control.

Jan-Willem van Wingerden (Senior Member, IEEE) was born in Ridderkerk, The Netherlands, in 1980. He received the B.S. and Ph.D. (cum laude) degrees in mechanical engineering and in control engineering from the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands, in 2004 and 2008, respectively. His Ph.D. thesis was entitled Smart Dynamic Rotor Control for Large Offshore Wind Turbines.

He is currently a Professor with the Delft University of Technology. His current research interests include linear parameter varying identification, subspace identification, smart structures and control, and identification of wind turbines and wind farms.