Rebooting Kernel CCA method for Nonlinear Quality-relevant Fault Detection in Process Industries

Hongchao Cheng[a,b] , Yiqi Liu[a*], Daoping Huang[a], Baoping Cai[c], Qilin Wang[b]

[a]School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China.

[b]Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia.

[c] College of Mechanical and Electronic Engineering, China University of Petroleum, Qingdao, Shan -dong, China .

Corresponding author：

aulyq@scut.edu.cn（YiqiLiu*）

Abstract

Process monitoring is essential and important strategy for ensuring process safety and product quality. However, due to the nonlinear characteristics and multiple working conditions in process industries, the traditional process monitoring method cannot be effectively applied. Therefore, we propose a novel process monitoring framework, termed as mixture enhanced kernel canonical correlation analysis framework (M-NAKCCA). The innovations and advantages of M-NAKCCA are as follows: 1). The traditional CCA method is re-boosted as a new method, M-NAKCCA, to better nonlinear fault detection. Also, a matter-element model (MEm) is assimilated into M-NAKCCA to make the information more refined. 2). To overcome the curse of

dimensionality that usually occurs in the high-dimensional dataset, M-NAKCCA uses the Nyström approximation technology to compress the kernel matrix. Moreover, the T^2 control chart is reconstructed and the corresponding control upper limit is re-configured to improve the method sensitivity and to better the fault detection performance. 3). The proposed M-NAKCCA framework is firstly used to monitor a wastewater treatment plant (WWTP) and chemical plant with diverse process behaviors. The experimental results showed that the M-NAKCCA framework achieved the best performance for both of case studies.

Keywords: Kernel canonical correlation analysis; Nyström approximation; Matter-element; Fault detection; Nonlinear industrial process.

1. Introduction

Since industrial processes become more and more complicated, the safety of process industry system has received attention gradually (Ge et al., 2010; Kramer, 1991; Liu et al., 2016; Ma et al., 2019; Olsson, 2012). Taking the WWTP as an example, WWTP is a complex system that contains chemical, physical, and biological reactions. Such complexity, in turn, makes the generated data usually exhibit the characteristics of multi-modality, time-varying parameters, multi-variable coupling, and nonlinearity. Moreover, the wastewater treatment process involves a large amount of sensors (temperature, alkalinity, flow, *etc*.), brakes, and biochemical reaction devices. Simultaneously, different types of faults will occur because the equipment is always exposed to the harsh environment (Olsson, 2012). These faults will directly lead to the degradation of effluent quality. Poorly treated wastewater flowing into the river will

have severe effects on the environment, and may also cause the irreversible damage to citizens' living. Therefore, it is necessary to take the quality-relevant faults into account during the plant management. Fortunately, the data generated and stored by the WWTPs contains important information and knowledge about the equipments. If the abnormal information of equipments can be extracted from the dataset, the refined knowledge can be used to build a model to monitor and support the management of the WWTPs.

Therefore, the data-driven process monitoring methods have attracted much attention from the industrial communities and academic (Baklouti et al., 2018; Chen et al., 2016; Cheng et al., 2019a; Ge et al., 2010; Jiang et al., 2017; Liu and Xie, 2020). Moreover, unsupervised statistical methods, such as principal component analysis (PCA), CCA, and forecastable component analysis (ForeCA) have been more studied widely (Cheng et al., 2019a; Liu et al., 2018; Shen et al., 2012; Zhu et al., 2016). Yu et al. proposed sparse PCA to monitor the Tennessee Eastman chemical process (TECP), the experiment showed that sparse PCA is more robust in corrupted data when performing feature extraction (Yu et al., 2016). Adisbaba et al. combined PCA with Bayesian network to detect the fault of the crude oil distillation unit operation system (Adedigba et al., 2017). Ghosh et al. combined the copula function and Bayesian network to predict the fault of multivariate time dependent process system. Over the past several decades, CCA achieves excellent performance and popularity because of its ability to monitor the quality-relevant fault accurately (Zhu et al., 2016). In the field of wastewater treatment, the quality-relevant faults are the most concerning types for WWTP. Therefore, it is necessary to study the application of CCA in wastewater

process monitoring. Cheng et al. proposed to use the combination of CCA and transfer learning to monitor the quality-relevant failures of WWTP (Cheng et al., 2020a). However, due to the complexity of a wastewater treatment process, the collected data are usually nonlinear. It is a pity that the basic statistical methods can show superior performance only for the linear data. To overcome this issue, industrial communities have done a lot of research works (Cheng et al., 2020b; Dong and Mcavoy, 1996; Ge et al., 2010; Kramer, 1991; Ma et al., 2019). Kramer et al. proposed to use a five-layer neural network to approximate the nonlinear relationship among data (Kramer, 1991). Nevertheless, the main defect is that it is time-consuming. To solve the time-consuming problem, Dong et al. proposed a nonlinear model combining the principal curve and the neural network. Although the time consumption is alleviated, the poor generalization performance makes the method unable to be further applied widely (Dong and Mcavoy, 1996). Besides, Ge et al. proposed to use the Bayesian inference to fuse multiple linear subspaces models, whose purpose is to approximate the nonlinear relationship among data (Ge et al., 2010). However, the accuracy of the nonlinear approximation method cannot be effectively guaranteed. Recently, the kernel track-based techniques have gained more attention. Because there is no need to know the mapping function, and to a certain extent, it can avoid the dimensional disasters (Yu et al., 2019). Therefore, KPCA, SVM, and other nonlinear statistical methods are gradually being studied in the field of process monitoring (Cheng et al., 2019a; Jiang and Yan, 2015). Cheng et al. used a novel ensemble adaptive sparse Bayesian transfer learning machine for nonlinear process monitoring(Cheng et al., 2020b). Recently, Samuel et al. used the kernel

canonical variate analysis (KCVA) based method to monitor a TE process(Samuel and Cao, 2015). Inspired by this, this paper embeds kernel trick into the CCA method. The difference with KCVA is that KCCA can capture the structural information of the nonlinear system among input and output space, while KCVA is mainly to describe the relevant information on the time node. Therefore, the KCCA method is suited to explain quality-relevant faults. However, the KCCA method still has the following disadvantages. Firstly, although the idea of the kernel track method can avoid the dimensional disaster, the time consumption of operations in high-dimensional space cannot be completely avoided (Yu et al., 2019). Secondly, the chemical process not only has a complex internal reaction but also suffers from the uncertain external disturbances. Therefore, the fault signal is easy to be hidden by other signals (Cheng et al., 2019b), which leads to the general method cannot effectively extract the fault information.

To tackle the overly time-consuming problem of the kernel-based methods, Cheng et al. use the forecast component analysis to reduce the computational burden of SVM (Cheng et al., 2019a). Generally, these methods reduce the space dimension by an external strategy. It is noteworthy that the external strategy does not consider the best approximation of the sparse matrix, which will cause the information lost. To achieve the proper low-rank approximation of the kernel matrix, we consider directly shrink the kernel matrix. Firstly, Nyström approximation technology is used to calculate the kernel matrix with approximate eigenvalues and eigenvectors (Williams and Seeger, 2000). Secondly, the best low-rank approximation of the matrix is obtained through cross-validation. Nyström technology is firstly proposed to solve integral equations. The

principle behind is to use a small number of sampling points to approximate convolution operators, and then to obtain the corresponding eigenvector. In recent years, because the feature space approximated by Nyströ m technology can reduce the computational complexity while retaining more data information, this method has become a helpful tool for processing the big data (Li et al., 2010). According to the analysis of Williams et al. (Williams and Seeger, 2000), Nyström technology can reduce the original computational complexity from $O(n^3)$ to $O(m^2n)$. Therefore, the Nyström technology is used in many different fields, such as image processing(Fowlkes et al., 2004), text clustering(Li et al., 2010), and fault detection(Ma et al., 2019), *etc*. Based on the above theory and application foundation,

M-NAKCCA uses the Nyström technology to obtain the low-rank approximation of the kernel matrix. M-NAKCCA method firstly considers taking the information processing for the raw data. The idea of information processing is based on the extenics theory proposed by Cai et al. (Yang and Wen, 2015) and discrete mathematics. It firstly uses raw data to establish a complete MEm, and then uses the matter-element indicators to split the data. Finally, the correlation degree (CRD) function is constructed to process the corresponding layer data. The segmented domain of MEm provides a larger external space for extracting the data statistical features, which makes the redundant information in the data matter-element compressed effectively.

The motivation of this work is to develop a novel nonlinear method, called M-NAKCCA, which aims to solve the problems as mentioned above. Besides, several typical process faults are also discussed. As we all know, industrial process faults can

be roughly divided into two categories (Chen et al., 2016; Juricek et al., 2004): (i) Additive faults. (ii) Multiplicative faults. In terms of the data statistical characteristics, additive fault means that the mean value changes over time, whereas multiplicative fault means that the variance changes accordingly. From the perspective of the detection object, additive fault refers to the sudden deviation of the sensor, whereas multiplicative fault means that the system parameter change or external random disturbance (Juricek et al., 2004). Chen et al. pointed out that multiplicative faults can be approximately transformed into additive faults (Chen et al., 2016). Therefore, if the proposed method has sufficiently excellent performance, it can effectively diagnose additive faults while detecting multiplicative faults. Moreover, drift fault and step fault are further studied. More details of step fault and drift fault definition can refer to (Jiang et al., 2017; Liu et al., 2016). Moreover, fault signal is often identified by T^2 and squared prediction error (SPE) control charts. In previous studies, it was found that the T^2 control chart is more insensitive than SPE (Cheng et al., 2019b). Therefore, T^2 usually cannot effectively identify the early stage of drifting faults. Larimore et al. used the residual-based control chart to test state-space change (Larimore, 1997). Chen et al. combined the local statistical approach and the residual of state-space to identify the incipient multiplicative fault (Chen et al., 2016). In this paper, the T^2 control chart is reconstructed by the residual of principal canonical space, and then uses it to store the system input-output information. Also, the corresponding upper control limit (UCL) is reconsidered. Recently, Ma et al. pointed out that the kernel density estimation(KDE) is used to set UCL can achieve a satisfactory classification effect (Ma et al., 2019). KDE

is different from the χ^2 statistic or F-distribution, which relaxes the restriction of data obeying Gaussian distribution.

Based on the above discussion, the main contribution of the paper is to develop a novel process monitoring framework for identifying quality-relevant multiplicative faults and additive faults. Meanwhile, the typical fault patterns, such as drifting faults and step faults, are fully considered under the proposed framework. The novelty and advantages of the proposed framework can be summarized as follows:

(1) In this paper, we firstly propose a novel process monitoring framework of M-NAKCCA, which is able to identify the different abnormal events of the system. Simultaneously, MEm technology is embedded in M-NAKCCA framework, aiming to improve the information extraction ability of the method.

(2) In the framework of M-NAKCCA, Nystro¨m technology is used to obtain the low dimensional approximation of the original high dimensional matrix. This alleviates the time-consuming problem of the traditional KCCA.

(3) To solve the problem that the T^2 control chart is not sensitive to the drift fault, the residual of input-output is used to reconstruct the T^2 control chart. Moreover, KDE is used to define and set up the corresponding UCL for fault detection.

The remainder of this paper is organized as follows. In Section 2, the paper presents the basic theory of CCA. In Section 3, The theoretical formula of M-NAKCCA is derived in detail. In Section 4， M-NAKCCA is used for monitoring some different type faults. In Section 5, the paper ends with some conclusions.

2. Preliminary work

CCA is a standard multivariate statistical method. In CCA, the internal structure between input data ($X \in R^{p\times n}$) and output data ($Y \in R^{q\times n}$) is described by extracting the correlation between two sets of variables, where p and q represent the number of monitoring variables resepectively. In(Ma et al., 2019), the correlation coefficients can be obtained by the following optimization function:

$$\rho(\alpha^T X, \beta^T Y) = \arg\max_{\alpha,\beta} \frac{\alpha^T \Sigma_{XY} \beta}{\sqrt{\alpha^T \Sigma_{XX} \alpha} * \sqrt{\beta^T \Sigma_{YY} \beta}}, \quad (1)$$

where non-zero constant vector α and β are the weight coefficients, the max correlation coefficient $\rho(*)$ can be learned by finding the best linear combination between $\alpha^T X$ and $\beta^T Y$. Σ_{XX}, Σ_{XY} and Σ_{YY} represent the covariance matrix. To ensure the uniqueness of the result, let $\alpha^T \Sigma_{XX} \alpha = 1$ and $\beta^T \Sigma_{YY} \beta = 1$. To solve Eq.(1), the following formula can be obtained by using the Lagrange multiplier:

$$L(\alpha, \beta, \gamma_X, \gamma_Y) = \alpha^T \Sigma_{XY} \beta - (\alpha^T \Sigma_{XX} \alpha - 1) * \left(\frac{\gamma_X}{2}\right) - (\beta^T \Sigma_{YY} \beta - 1) * \left(\frac{\gamma_Y}{2}\right), \quad (2)$$

γ_X and γ_Y represent the Lagrange multiplier. Eq.(1) can be simplified by transforming into the following general algebraic eigenvalue problem:

$$\begin{bmatrix} 0 & \Sigma_{XY} \\ \Sigma_{YX} & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda \begin{bmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (3)$$

The weight coefficients α and β can be obtained by solving the generalized eigenvalue of Eq.(3). λ is the corresponding eigenvalue.

3. Process monitoring framework of M-NAKCCA for quality-relevant faults

In this section, the theory of M-NAKCCA is derived in detail. Moreover, the entire procedure to implement the M-NAKCCA framework for the quality-relevant fault monitoring is presented in Figure 1.

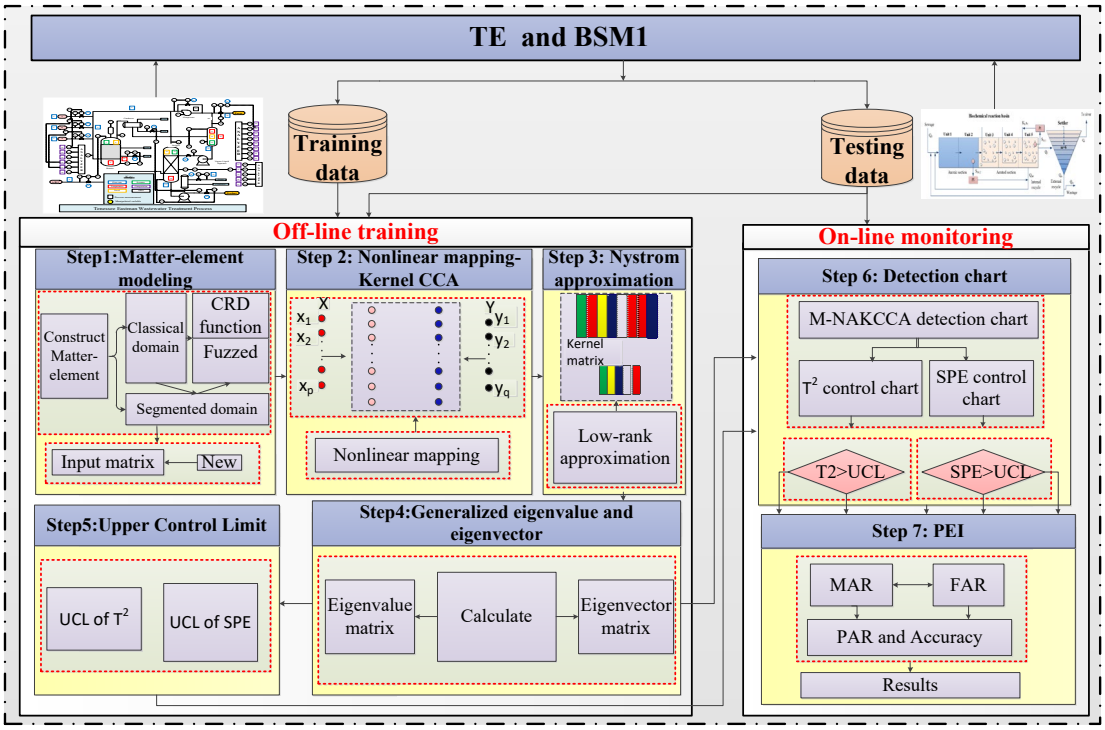

Figure 1. Procedures of implementing M-NAKCCA for process monitoring

3.1. The mixture kernel canonical correlation analysis

This work firstly embedded the matter-element model (MEm) into KCCA. Therefore, the MEm is established as following by the offline raw data:

$$M_t(\theta, x, \varphi(x))=\begin{pmatrix} \theta & x_{1t} & \varphi(x_{1t}) \\ & \vdots & \vdots \\ & x_{lt} & \varphi(x_{lt}) \end{pmatrix}, \tag{4}$$

where $\mathrm{M_t}(*)$ is matter-element containing the system information. θ is a system that generates raw data. $\varphi(x_{lt})$ represents the state value collected by the lth sensor at time t. The fault matter-element and non-fault matter-element can be established separately. Moreover, the unsupervised model only needs to get the non-fault state data during the offline training stage. According to (Gong et al., 2012; Yuan et al., 2013), the classical domain and segmented domain are established as follows:

$$M^C(\theta^{\mathrm{norm}}, x, \varphi(x))=\begin{pmatrix} \theta^{\mathrm{norm}} & x_1 & \varphi(x_1) \\ & \vdots & \vdots \\ & x_{p+q} & \varphi(x_{p+q}) \end{pmatrix}, \tag{5}$$

$$M^S(\theta^{\text{norm}}, x, \varphi'(x))=\begin{pmatrix} \theta^{\text{norm}} & x_1 & \varphi'(x_1) \\ & \vdots & \vdots \\ & x_{p+q} & \varphi'(x_{p+q}) \end{pmatrix}. \quad (6)$$

$M^C(*)$ is the classical domain, meaning the matter-element with respect to the system under normal condition. $\varphi(x_i)=\text{extremum}[x_{i1},\cdots,x_{in}]=(x_{i,\min}, x_{i,\max})$, where $x_{i,\min}$ and $x_{i,\max}$ represent the minimum and maximum value respectively, they are extracted from the sample set. i is the corresponding sensor ($1 \le i \le p+q$), n is the number of samples. $M^S(*)$ is the segmented domain, which represents the classical domain extended by the actual condition and engineer experience. In the actual industrial processes, the number of sampling is usually limited. Moreover, the existence of uncertainty disturbances makes it even difficult for the classical domain to describe the system state properly (Yong-xiu et al., 2011). Therefore, a new segmented domain is established, where $\varphi'(x_i)=(x'_{i,\min}, x'_{i,\max})$. $x'_{i,\min}$ and $x'_{i,\max}$ represent the maximum and minimum values of the segmented domain respectively. The corresponding extreme values are obtained by the following formula:

$$x'_{i,\min}=x_{i,\min}-\xi(x_{i,\max}-x_{i,\min}), \quad (7)$$

$$x'_{i,\max}=x_{i,\max}+\xi(x_{i,\max}-x_{i,\min}). \quad (8)$$

ξ is the penalty parameter, which is used to adjust the range of segmented domain. We standardize the offline matter-element model $M=(M_1 \quad \cdots \quad M_n)$. As profiled in Figure1, in the offline training stage, the proper correlation degree (CRD) function needs to be established to evaluate the core range of system transformation. Based on (Yong-xiu et al., 2011; Yuan et al., 2013), the corresponding CRD function can be derived as follows:

$$\mathrm{D}(x_{ij}) = \frac{\rho(x_{ij}, M^C)}{\delta(x_{ij}, M^C, M^S)}, \tag{9}$$

where $\rho(x_{ij}, M^C)$ is the distance between x_{ij} and classical domain M^C. x_{ij} is the value of the *j*th sensor at the *i*th sample. $\delta(x_{ij}, M^C, M^S)$ is the distance of x_{ij} between M^C and M^S. According to (Yuan et al., 2013). The formula of $\rho(x_{ij}, M^C)$ and $\delta(x_{ij}, M^C, M^S)$ can be constructed as follows:

$$\rho(x_{ij}, M^C) = \left| x_{ij} - \frac{x_{i,min} + x_{i,max}}{2} \right| - \frac{x_{i,max} - x_{i,min}}{2}, \tag{10}$$

$$\delta(x_{ij}, M^C, M^S) = \begin{cases} \rho(x_{ij}, M^S) - \rho(x_{ij}, M^C) & x_{ij} \notin M^C \\ \rho(x_{ij}, M^S) - \rho(x_{ij}, M^C) + (x_{i,min} - x_{i,\max}) & x_{ij} \in M^C \end{cases}. \tag{11}$$

The constructed distance function $\rho(*)$ and $\delta(*)$ combine with the CRD function, it can be deduced that CRD has the following three important properties: *i*). If $x_{ij} \in M^C$, $\mathrm{D}(x_{ij}) \geq 0$. It represents the test data under the normal working condition. *ii*). If $x_{ij} \notin M^C$ and $x_{ij} \in M^S$. Then, $-1 < D(x_{ij}) < 0$. In this case, it can be determined if the system is in a critical and emergent state. *iii*). If $x_{ij} \notin M^C$, $x_{ij} \notin M^S$. It represents if the system is out of work, so we can derive that $\mathrm{D}(x_{ij}) \leq -1$. When the offline CRD function is established, the online test data in the segmented domain can be modified accordingly. And the faulty data outside the segmented domain can sequently be enlarged. Suppose the test data $V = [V_{\mathrm{norm}}, V_{\mathrm{fault}}] \in R^{l*m}$. V_{norm} and V_{fault} represent the data collected by norm work condition and the abnormal condition respectively, m is the number of samples. According to the value of CRD, the fuzzy test set can be obtained by the following rules:

(1) If $\mathrm{D}(v_{ij}) \geq 0$, the test data is normed by mean and variance of training data, ${v_{ij}}^* = (v_{ij} - x_{i,mean})/\sqrt{var(x_i)}$.

(2) If $-1 < D(x_{ij}) < 0$, it represents the test data is out of the classical domain. In general, the data in the segmented domain is the system deviation caused by noise or human factors. Therefore, the test data can be returned to the classical domain by the following special processing. This can further mitigate the influence of corrupted data. If $x'_{i,\min} < v_{ij} < x_{i,min}$, $v_{ij}{}^{*} = (v_{i,min} - x_{i,mean})/\sqrt{var(x_i)}$. If $x_{i,\max} < v_{ij} < x'_{i,max}$, $v_{ij}{}^{*} = (v_{i,max} - x_{i,mean})/\sqrt{var(x_i)}$.

(3) If $\mathrm{D}(x_{ij}) \leq -1$, it represents the data is abnormal and the corresponding system is out of work. However, if the abnormal signal is weak, the monitoring method cannot identify it properly. Thus, we can let $v_{ij}{}^{*} = \xi'(v_{ij} - x_{i,mean})/\sqrt{var(x_i)}$. ξ' is the corresponding compensation coefficient, $\xi' = 1 - \xi D(x_{ij})$.

After the above MEm processes the data, it is assumed that the newly obtained training set and testing set are $U_{train} = [U_X,\ U_Y] \in R^{(p+q)*n}$ and $V_{test} = [V_X,\ V_Y] \in R^{(p+q)*n}$, respectively. $U_X \in R^{p*n}$ and $U_Y \in R^{q*n}$ represent the data set of input and output respectively. Since the processed data remains nonlinear, the performance of the linear model is severely affectedand could be deviated. To describe the internal structure of nonlinear data, the data are firstly mapped to high dimensional space:

$$\begin{cases} \phi: & U_X \in R^{p*n} \rightarrow \phi(U_X) = [\phi(x_1) \quad \cdots \quad \phi(x_n)] \\ & U_Y \in R^{q*n} \rightarrow \phi(U_Y) = [\phi(y_1) \quad \cdots \quad \phi(y_n)]' \end{cases} \tag{12}$$

where ϕ is the mapping function, H_X =span $\{\phi(x_1) \quad \cdots \quad \phi(x_n)\} \in R^{n*n}$ and H_Y=span$\{\phi(y_1) \quad \cdots \quad \phi(y_n)\} \in R^{n*n}$ represent the mapped feature space. Suppose H_X and H_Y are complete inner product spaces. So $\exists\ f_1 \in H_X$ and $f_2 \in H_Y$, let

$$f_1 = \sum_{i=1}^{n} \phi(x_i)\, w_{X_i}, \tag{13}$$

$$f_2 = \sum_{i=1}^{n} \phi(y_i)\, w_{Y_i}, \tag{14}$$

f_1 and f_2 are n-dimensional basis vectors of H_X and H_Y respectively. The core idea of CCA is to to find the optimal eigenvectors W_X and W_Y, in such a way that ${f_1}^T\phi(X)$ and ${f_2}^T\phi(Y)$ has the greatest correlation. The cost function can be constructed by combining Eq.(1):

$$\rho\left({f_1}^T\phi(X), {f_2}^T\phi(Y)\right) = \arg\max_{W_X, W_Y} \frac{cov({f_1}^T\phi(X), {f_2}^T\phi(Y))}{\sqrt{var({f_1}^T\phi(X))} * \sqrt{var({f_2}^T\phi(Y))}}, \tag{15}$$

$\phi(X)$ and $\phi(Y)$ represent the mapping function of input and output space respectively. In general, the corresponding kernel function is defined as $K_X = \langle\phi(U_X), \phi(U_X)\rangle = \phi(U_X)^T\phi(U_X)$, $K_Y = \phi(U_Y)^T\phi(U_Y)$. The covariance matrix of Eq. (15) can be derived as follows:

$$cov({f_1}^T\phi(X), {f_2}^T\phi(Y)) = \frac{1}{n-1}\sum_{i=1}^{n} {f_1}^T\phi(x_i)\,({f_1}^T\ \phi(y_i))^T =$$

$$\frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{d=1}^{n} {w_{X_j}}^T\,\phi(x_j)^T\phi(x_i)\ \ \phi(y_i)^T\phi(y_d)\ \ w_{Y_d} = {W_X}^T\Sigma_{K_XK_Y}W_Y, \tag{16}$$

where K_X and K_Y are the kernel matrix, $\Sigma_{K_XK_Y} = \frac{1}{n-1}K_X{K_Y}^T$ is the covariance matrix. So, $\left({f_1}^T\phi(X), {f_2}^T\phi(Y)\right) = \max_{W_X, W_Y} \frac{{W_X}^T\Sigma_{K_XK_Y}W_Y}{\sqrt{{W_X}^T\Sigma_{K_XK_X}W_X} * \sqrt{{W_Y}^T\Sigma_{K_YK_Y}W_Y}}$. When $\Sigma_{K_XK_X}$ and $\Sigma_{K_YK_Y}$ are invertible, Eq. (2) and Eq. (15) can be transformed into the following generalized eigenvalue problem:

$$\begin{bmatrix} 0 & K_X{K_Y}^T \\ K_Y{K_X}^T & 0 \end{bmatrix}\begin{bmatrix} W_X \\ W_Y \end{bmatrix} = \lambda\begin{bmatrix} K_X{K_X}^T & 0 \\ 0 & K_Y{K_Y}^T \end{bmatrix}\begin{bmatrix} W_X \\ W_Y \end{bmatrix}. \tag{17}$$

λ is the eigenvalue.

3.2 Nyström approximation of mixture KCCA for fault detection

3.2.1 Mixture Nyström approximation

Reformulation of CCA based on kernel function will lead to the increment of the

computational burden (Williams and Seeger, 2000; Yu et al., 2019). Therefore, how to retain more data information while compressing the high-dimensional space has become an important issue to be solved. The kernel matrix of high dimensional space is usually a sparse and non-full rank. Therefore, the assumption that $\Sigma_{K_X K_X}$ and $\Sigma_{K_Y K_Y}$ are both invertible cannot be guaranteed easily. Recently, the Nyström approximation technique is used to seek the low-rank approximation of the kernel matrix. the low-rank approximation can be derived as follows:

$$\widetilde{K} = \widetilde{U}\tilde{\Lambda}\widetilde{U}^T = \sum_{i=1}^{m} \tilde{\lambda}_i^{(n)} \tilde{u}_i^{(n)} (\tilde{u}_i^{(n)})^T, \tag{18}$$

$\widetilde{K}$ represents the Nyström approximations of the kernel matrix K, $\tilde{\lambda}_i^{(n)}$ and $\tilde{u}_i^{(n)}$ are the corresponding eigenvalue and eigenvector respectively. m is the dimension of the low-rank approximation matrix, the eigenvalue and eigenvector formulas can be derived as follows:

$$\tilde{\lambda}_i^{(n)} \stackrel{\text{def}}{=} \frac{n}{\tau} \tilde{\lambda}_i^{(\tau)} (i = 1, \cdots \tau), \tag{19}$$

$$\tilde{u}_i^{(n)} \stackrel{\text{def}}{=} \sqrt{\frac{\tau}{n}} \frac{1}{\tilde{\lambda}_i^{(\tau)}} K_{n,\tau} \tilde{u}_i^{(\tau)} (i = 1, \cdots \tau), \tag{20}$$

where τ is the dimension of the constructed subspace from the kernel space, and $m \leq \tau < n$. $K_{n,\tau}$ is the appropriate $n \times \tau$ submatrix of K, which can be obtained by the K-means clustering method(Ma et al., 2019). In summary, by embedding the Nyström approximations method, the computation complexity can be reduced from $O(n^3)$ to $O(\tau^2 n)$. The new eigenvalues and eigenvectors can be derived by introducing the low-rank matrix into Eq. (17).

Subsequently, the data space can be decomposed into quality relevant-subspace and irrelevant-subspace. To monitor the effluent quality relevant fault, the following quality

relevant T^2 can be obtained:

$$T^2 = Z^T \Sigma_r^{-1} Z, \tag{21}$$

according to reference (Chen et al., 2016), where $Z = W_Y{}^T K_Y - \Lambda_r W_X{}^T K_X$, $W_X = [w_{X_1} \quad \cdots \quad w_{X_r}] \in R^{n\times r}$, and $W_Y = [w_{Y_1} \quad \cdots \quad w_{Y_r}] \in R^{n\times r}$. $W_Y{}^T K_X$ is the score matrix of input space, $W_X{}^T K_Y$ is the score matrix of output space. Thus, residual matrix Z represents the system state changes between output space and input space. The matrix Z can be derived as followings:

$$Z = \left(W_Y{}^T K_Y - \Lambda_r W_X{}^T K_X\right), \tag{22}$$

where $W_X{}^T \Sigma_{K_X K_X} W_X \approx I_r$ and $W_Y{}^T \Sigma_{K_Y K_Y} W_Y \approx I_r$. Also, $\Sigma_{K_X K_Y} W_Y = \Lambda_r \Sigma_{K_X K_X} W_X$ and $\Sigma_{K_Y K_X} W_X = \Lambda_r \Sigma_{K_Y K_Y} W_Y$ can be easily derived by Eq. (3). Thus, Σ_r can be obtained as

$$\Sigma_r \approx \frac{1}{n-1} ZZ^T = \frac{1}{n-1}\left(W_Y{}^T K_Y - \Lambda_r W_X{}^T K_X\right)\left(W_Y{}^T K_Y - \Lambda_r W_X{}^T K_X\right)^T = W_Y{}^T \Sigma_{K_Y K_Y} W_Y$$

$$-\Lambda_r W_Y{}^T \Sigma_{K_Y K_X} W_X - \Lambda_r W_X{}^T \Sigma_{K_X K_Y} W_Y + \Lambda_r{}^2 W_X{}^T \Sigma_{K_X K_X} W_X = I_r - \Lambda_r{}^2, \tag{23}$$

where $\Lambda_r = diag(\lambda_1, \cdots, \lambda_r)$, it can be represented by the first r maximum eigenvalue. Moreover, according to reference (Zhu et al., 2016), the quality irrelevant fault can be monitored by the following SPE control chart:

$$\mathrm{SPE} = e^T e, \tag{24}$$

where $e = (I - W_X{}^T W_X) K_X$ or $e = (I - W_Y{}^T W_Y) K_Y$. Therefore, the SPE control chart can be divided into two parts (Zhu et al., 2016). The first SPE is more influential in the input space, whereas the second one is more reliable in the output space. In this paper, the SPE is calculated according to the actual working conditions.

3.2.2 Upper control limit

Kernel density estimation (KDE) is an essential nonparametric estimation method, which is often used for the setting of the upper control limit (UCL) for T^2 and SPE (Ma et al., 2019; Odiowei and Cao, 2010). Based on (Odiowei and Cao, 2010), the corresponding threshold can use the kernel density estimations shown as follows：

$$P(T^2 < UCL_{\mathrm{T}^2}) = \int_{-\infty}^{\mathrm{UCL}_{\mathrm{T}^2}} p(\mathrm{T}^2)\, d\mathrm{T}^2 = \tilde{\alpha}, \tag{25}$$

$$P(SPE < UCL_{\mathrm{SPE}}) = \int_{-\infty}^{UCL_{SPE}} p(\mathrm{SPE})\, d(\mathrm{SPE}) = \tilde{\alpha}. \tag{26}$$

$\tilde{\alpha}$ is the confidence level. UCL_{T^2} and UCL_{SPE} are upper control limit of T^2 and SPE respectively. If $T^2 > UCL_{T^2}$ or $\mathrm{SPE} > UCL_{\mathrm{SPE}}$, the system will declare that the working states are out of control. KDE is a method based on the data samples, which does not need to know the prior knowledge of data distribution.

3.2.3 Performance evaluation index(PEI)

The PEI is an important decision tool for evaluating the performance of a monitoring method. To comprehensively evaluate the performance of different methods, missed alarm rate (MAR), false alarm rate (FAR), fault diagnosis accuracy, and pre-alarm rate (PAR) are used for model evaluation. FAR means that the system false alarms proportion in the total number of alarms. Because the system engineers need to check the equipment after the alarm declaration, it will waste unnecessary workforce and material resources if suffering from the false alarms. Moreover, the missed alarms have more serious influence than false alarms. The PAR is constructed by mixing false alarm and missed alarm indicators (Cheng et al., 2019a). It is used as a comprehensive PEI together with fault diagnosis accuracy. The corresponding formulas are as follows:

$$\mathrm{MAR} = \mathrm{Fr}(\mathrm{Normal}|\mathrm{Fault}) = \frac{FN}{FN+TP}, \tag{27}$$

$$\mathrm{FAR} = \mathrm{Fr}(\mathrm{Fault}|\mathrm{Normal})\frac{FP}{FP+TN}, \tag{28}$$

$$P_{AR} = \varpi\mathrm{MAR} + (1-\varpi)\mathrm{FAR}, \tag{29}$$

$$\mathrm{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}. \tag{30}$$

Note that the "Normal" is the non-fault conditions. Fr(|) represents the conditional frequency (Cheng et al., 2019a). TP is true positive, TN represents true negative. FP is false positive. ϖ is the weight parameter ($0 \leq \varpi \leq 1$).

4. Case studies

In this studies, the proposed M-NAKCCA method is used to monitor the TE chemical plant and a wastewater treatment plant (BSM1). BSM1 is a water treatment platform developed by the international water association (IWA). TE is simulated a real chemical plant designed by the chemical company. Meanwhile, academics and industries widely recognize both platforms. Moreover, the parameter of PC is CPU intel core i7-6700HQ, 8GB RAM,1TB SSD.

4.1. Case study I: Tennessee Eastman chemical process

(1) Background: Tennessee Eastman process (TEP) is designed by the American Eastman chemical company, which is based on a real chemical plant. As shown in Figure 2, TEP consists of five core units: reactor, compressor, stripper, condenser, and separator. The whole process includes manipulated variables and measured variables. There are four gaseous reactants (A, B, C, D, E) and two liquid products (G and H). The reaction equation is as follows:

$$\begin{cases} \mathrm{A(g)} + \mathrm{C(g)} + \mathrm{D(g)} \rightarrow \mathrm{G(liq)} \\ \mathrm{A(g)} + \mathrm{C(g)} + \mathrm{E(g)} \rightarrow \mathrm{H(liq)} \\ \mathrm{A(g)} + \mathrm{E(g)} \rightarrow \mathrm{F(byproduct)'} \\ \mathrm{3D(g)} \rightarrow \mathrm{2F(byproduct)} \end{cases} \tag{31}$$

where F is the byproduct, in the reactor, the reaction rates are Arrhenius functions of temperature, and the process is irreversible. More detailed reaction information of the TEP can refer (Downs and Vogel, 1993). According to downs and vogel introduction (Downs and Vogel, 1993), we can know that the TEP may occur five type faults (step fault, random variation fault, slow drift fault, sticking fault, and unknown fault). Step fault and drift fault are the most noteworthy among the above five type faults. When the external disturbance is strong, it is easy to cause the step fault. Due to the harsh working environment, the equipment is easy to be corroded. The control accuracy of this corroded equipment will slowly degrade. Therefore, this paper further investigates the step fault and drift fault. On this basis, we have also studied the random variation fault. According to (Zhu et al., 2016), the process contains 52 observation variables. Before proceeding with the data, the platform started with a 25-hour steady state. It is noted that the test set and training set are have the same sampling interval (the interval is 3 minutes). The training set is generated from the 48 h (1–960) without fault working condition. The data from 961 to 1920 is used as the testing set, where the sample 961-1120 are normal data, and the rest of the data (1121-1920) is sampled from the faulty working condition.

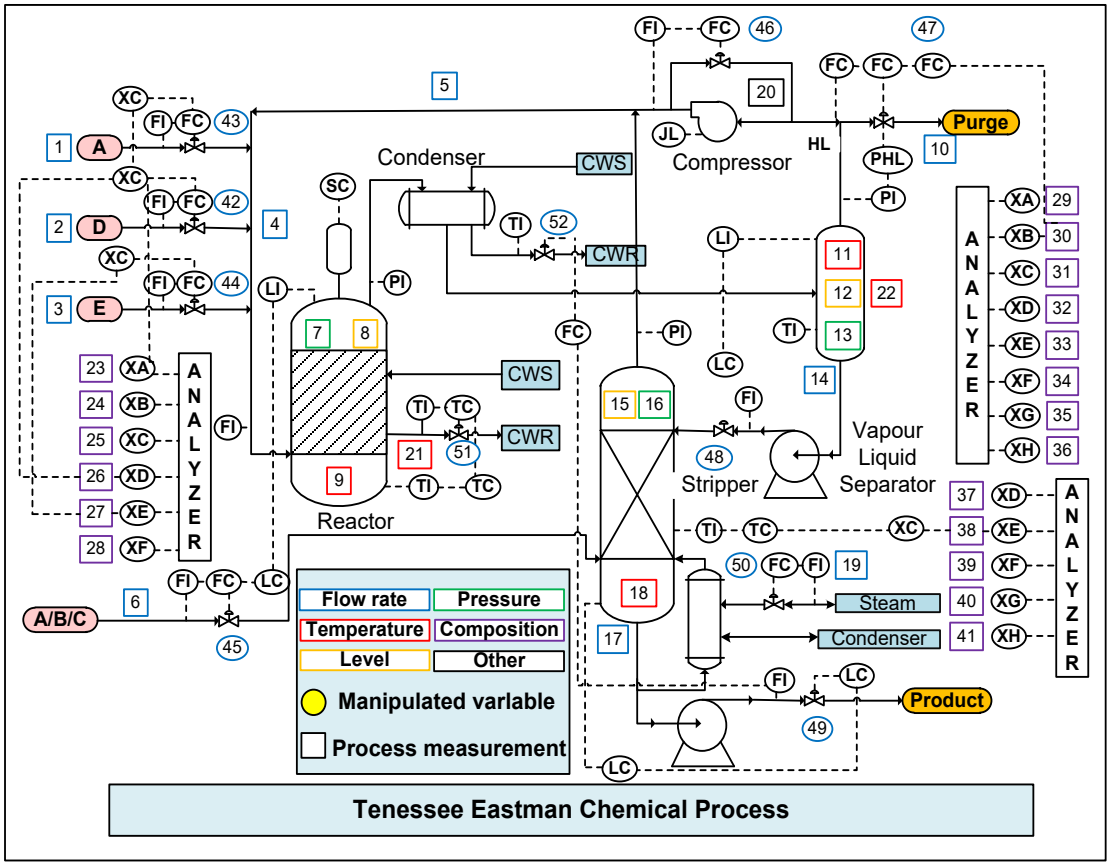

Figure 2. Schematic layout of the TE chemical process(Downs and Vogel, 1993).

(2) Discussion and analysis of monitoring result: M-NAKCCA is different from the previous overall modeling methods (PCA, SVM, ForeCA (ICA), *etc.*). It is used to analyze the correlation between two sets of variables. Therefore, the data set must be divided before training the model. Based on the study of Zhu et al.(Zhu et al., 2016). The study adds variables of the stream 9 (exhaust gas) as the quality-relevant variables based on the stream 11 (product). The rest variables are used as input (process) variables. As mentioned above, three typical faults (step fault, drift fault, and random variation fault) of TE have been further explored. According to the engineering experience, M-NAKCCA basic parameters are set as follows by trials and errors: the penalty parameter ξ of MEm is 0.01, and the kernel function is ' Gauss'. The confidence of UCL is 0.99. ϖ=0.6.

As profiled in Figure 3, the detection plot of M-NAKCCA/CCA method for the fault 1 is shown. The fault 1 is a step fault, which represents the component B step changed in the stream 4. Since the TE process involves chemical reactions, when an abnormality occurs in a specific component, it will inevitably cause a series of effects and faults will propagate around the entire process. Moreover, abnormal information will be stored in the testing data. As shown in Figure 3, the fault detection accuracy of the M-NAKCCA-T^2 is 98.33%. The accuracy of M-NAKCCA-SPE is 98.65%. Compared with the basic CCA method, the accuracy of fault detection is improved by 2.38% and 7.14% respectively. It is worth noting that the real chemical process has the self-adjusting function, so the system could be adjusted after the step fault. The M-NAKCCA method can capture the system state after the fault. On the contrary, the basic CCA method fails to deal with this issue. Fault 4 is a typical random variation fault, and it represents that the cooling water inlet temperature of the TEP reactor changes randomly. Because of the randomness of the disturbance, the system will have different degrees of chain reaction. Therefore, it is necessary to monitor the fault 4. The monitoring results are shown in Figure 4. The MAR of M-NAKCCA-T^2 is 0.126, the corresponding FAR equal to zero. The T^2 of M-NAKCCA has a fault diagnosis accuracy of 89.48%, while the CCA method has only 68.23%. The diagnosis accuracy of M-NAKCCA has increased by 31.15%.

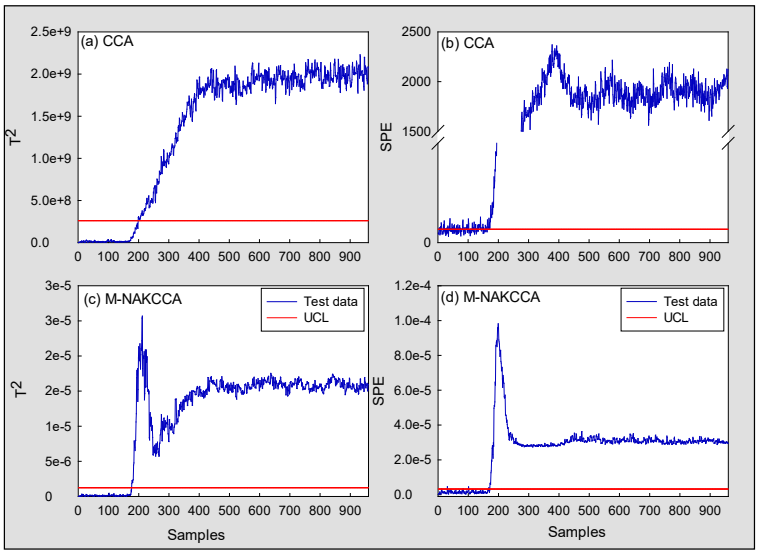

Figure 3. Monitoring results for a step fault (fault 1): (a) CCA-T^2, (b) CCA-SPE, (c) M-NAKCCA-T^2 and (d) M-NAKCCA-SPE

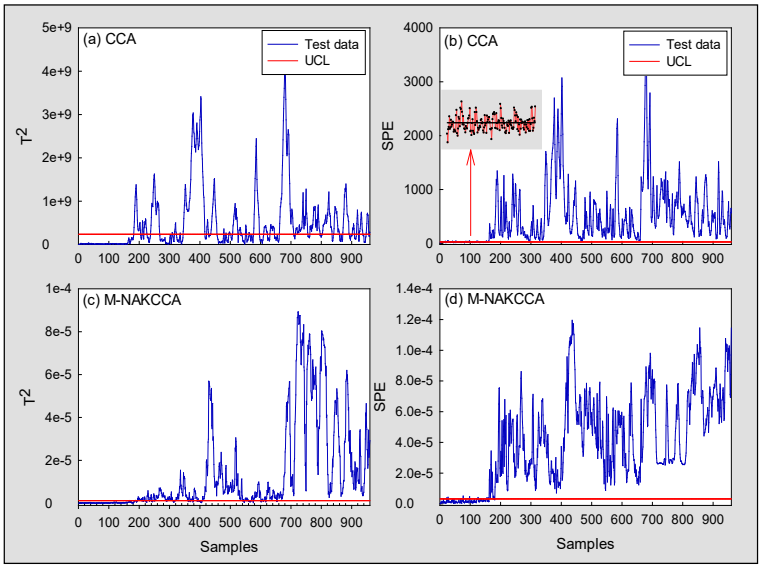

Figure 4. Monitoring results for random variation fault (fault 4): (a) CCA-T^2, (b) CCA-SPE, (c) M-NAKCCA-T^2, and (d) M-NAKCCA-SPE

Figure 5 presents the monitoring results of drift fault (fault 5). It is noteworthy that fault 5 is caused by the system reaction kinetic constants slight change, which is a slow drift fault. Although the M-NAKCCA method has fewer false alarms than CCA, the missed alarms are higher than CCA. Because the MEm is not well deal with the data,

the weak fault signal is not adequately compensated. Overall, the M-NAKCCA method diagnosis accuracy is higher than the CCA method. The accuracy of M-NAKCCA-SPE is 5.66% higher than that of CCA-SPE.

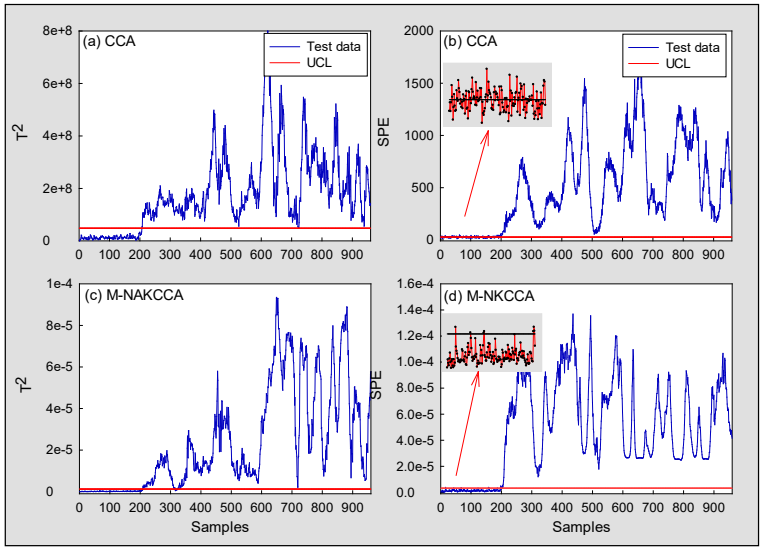

Figure 5. Monitoring results for a drift fault (fault 5): (a) CCA-T^2, (b) CCA-SPE, (c) M-NAKCCA-T^2 and (d) M-NAKCCA-SPE

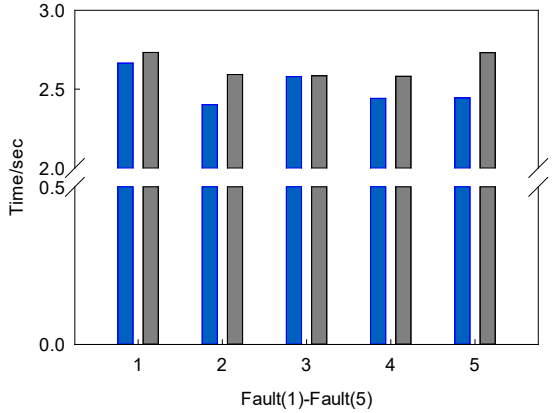

Figure 6. The consuming time of online monitoring (Note: Blue use Nystr ö m approximation, gray not use.)

Table 1. PEI values of different detection methods for the TE chemical process.

Average	PCA	KPCA	CCA	M-NAKCCA

(%)	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
FAR	0.25	0.25	23.5	-	6.625	44.875	0.0	4.0
MAR	34.925	11.225	20.475	-	18.1	0.7	5.5	1.7
PAR	21.055	6.835	21.685	-	13.51	18.37	3.3	2.62
Accuracy	70.854	90.604	79.021	-	83.813	91.938	95.417	97.917

As the formula derived in Section 3.2, we know that the Nyström approximation method can reduce the running time of the proposed method. Figure 6 presents the method running time for five different faults. Fault 1 and fault 2 are step faults. Fault 3 and fault 4 belong to random variable faults. Fault 5 classified as drift fault. Table 1 is the average value of detection results for the TE chemical process, which is mainly from the proposed M-NAKCCA method and three basic methods. As shown in Table 1, the average diagnosis accuracy of M-NAKCCA-T^2 method is 95.42%. The average accuracy of quality-irrelevant fault diagnosis is 97.92%. Moreover, the average PAR value of M-NAKCCA is low than that of the other three basic methods. This is mainly due to the M-NAKCCA method using the following essential technologies. Firstly, M-NAKCCA uses MEm and CRD to reduce and compensate the data subsequently. Secondly, the boosted nonlinear CCA technology is used to establish a reliable monitoring model, and this technology can extract more nonlinear data information while compressing high dimensional kernel space. Besides, KDE is used to set up the corresponding UCL，this technology can leap over the limitation of data must obeying a specific distribution. When the chemical data have some problems such as nonlinearity, noise, and multi-distribution, the M-NAKCCA method based on the above

technology can show superior performance.

4.2. Case study II: Wastewater treatment-benchmark simulation model 1

(1) **Background:** The international water association (IWA) design the benchmark simulation model 1 (BSM1) for removing the C and N, which is based on the real WWTP pre-denitrification process. Also, the BSM1 model is widely accepted by the most people. As shown in Figure 7, the core operating components of the wastewater treatment process, BSM1, interact between reactors and settlers: The first component is the biochemical reaction tank, which is based on the IAWQ activated sludge model No. 1 (ASM1). The second component is a secondary settler tank. It is designed by the Takács double exponential sedimentation velocity model. The detailed introduction can refer to the website (http://www.benchmarkWWTP.org). Firstly, BSM1 is simulated over 150 days with constant flow. Then, the sunny day's input data is used as the dynamic input of the WWTPs. In this study, the sampling interval is 15 min. And the simulation lasted for four weeks. During this period, the first two weeks' data as the training set. The training set has 1344 samples data, which is collected under the no-fault condition. In the testing set, different types of faults occur from 701 to 1344.

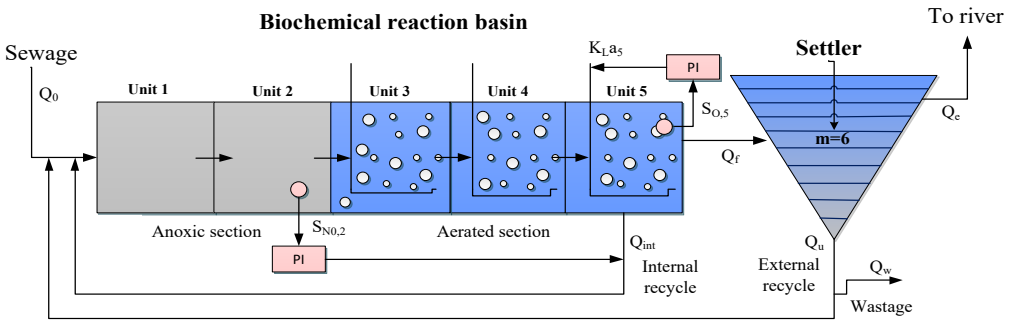

Figure 7. Schematic plot of BSM1(Liu et al., 2016).

(2) Discussion and analysis of monitoring results: In this section, according to the

knowledge about BSM1 and the engineering experience, thirty-two variables were selected, including twenty-eight process variables and four output variables. The monitoring variables covered all the processes of BSM1. Simultaneously, the severe abrupt faults and drift faults are considered by the proposed M-NAKCCA. M-NAKCCA basic parameters set as follows: the penalty parameter ξ is 0.5, and the kernel function is ' Gauss'. The confidence of UCL is 0.99. $\varpi = 0.6$.

Fault 1 is the oxygen sensor fault from the first unit of the BSM1. Oxygen is an essential element of biochemical reactions in activated sludge wastewater treatment, which directly affects the bacteria behaviours across the entire WWTP process. Therefore, M-NAKCCA is implemented for monitoring fault 1 herein. As shown in Figure 8. Missed alarm number of the M-NAKCCA-T^2 and M-NAKCCA-SPE is zero. The accuracy of M-NAKCCA-T^2 is 99.41%, while that of the CCA method is only 74.78%. This is mainly due to the traditional statistical method (PCA, CCA) can not effectively extract the nonlinear information of fault 1. Figure 9 is to profile the detection result of the drift fault (fault 3), it represents that autotrophic bacteria concentrations increase gradually. The fault diagnosis accuracy of M-NAKCCA-T^2 is 98.88%, and the accuracy of M-NAKCCA-SPE is 98.59%. It is important to notice that the fault signal during 700-800 is relatively weak. Fortunately, the MEm plays a proper compensation herein. However, CCA method has missed alarms at the beginning of the fault, and the MAR of CCA -T^2 is as high as 28.57%. By comparison, the M-NAKCCA-T^2 method's MAR is only 1.09%. Fault 3 is a typical drift fault, the corresponding data concurrently with nonlinear structure. The linear methods (PCA, CCA) cannot extract

the nonlinear structure information, which makes the behaviour of constructed linear model disordered.

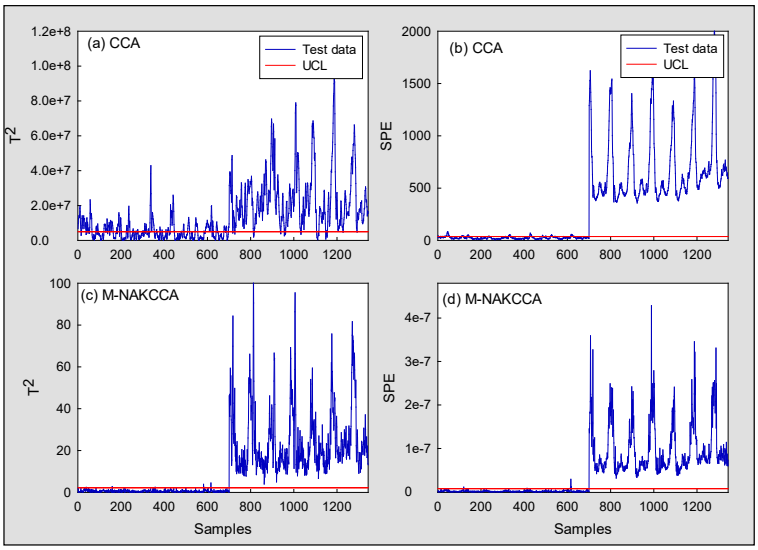

Figure 8. Monitoring results for a step fault (fault 1): (a) CCA-T^2, (b) CCA-SPE, (c) M-NAKCCA-T^2 and (d) M-NAKCCA-SPE

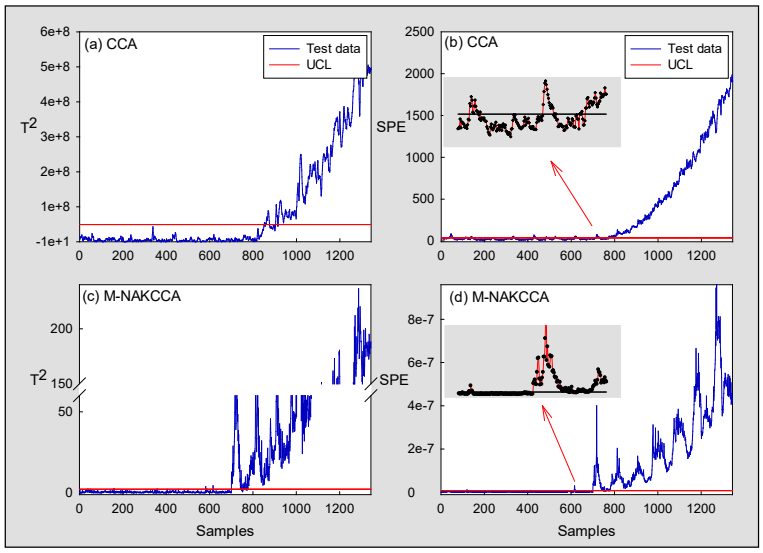

Figure 9. Monitoring results for a drift fault (fault 3): (a) CCA-T^2, (b) CCA-SPE, (c) M-NAKCCA-T^2 and (d) M-NAKCCA-SPE

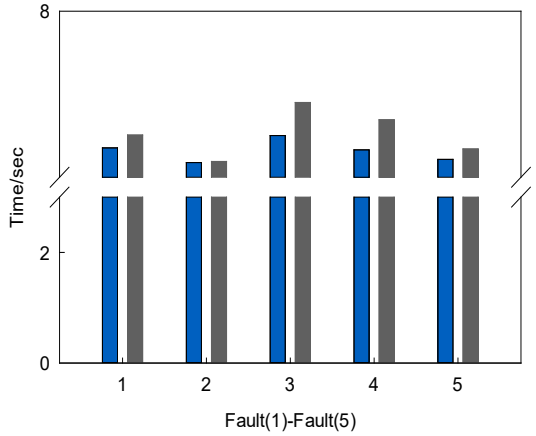

Figure 10. The consuming time of online monitoring (Note: Blue use Nystr ö m approximation, gray not use.)

Table 2. PEI values of different detection methods for the BSM1

Average	PCA		KPCA		CCA		M-NAKCCA	
(%)	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
FAR	2.143	1.143	3.6	-	8.629	13.0	7.429	6.971
MAR	60.093	10.621	52.888	-	19.068	36.646	1.957	2.143
PAR	36.913	6.830	33.173	-	14.892	27.188	4.145	4.074
Accuracy	70.089	94.316	72.783	-	86.369	75.670	95.194	95.342

Figure 10 shows the program running time for five different faults. Fault 1 and fault 2 are step faults. Fault 3-Fault 5 are classified as drift faults. Fault 3 is the most time-saving case by using the Nyström approximation technology. Fault 2 is the least time-consuming. Moreover, as summarized in Table 2, it presents the average value of detection results for the BSM1. The accuracy of M-NAKCCA method for detecting quality-relevant fault is 95.19%. The average accuracy for quality-irrelevant fault is 95.34%. By comparison with the basic CCA method, the accuracy of T^2 and SPE

improved by 10.218% and 25.997% respectively. Besides, The comprehensive index PAR of the M-NAKCCA method is also improved compared to three basic methods. The PAR value of M-NAKCCA-T^2 is only 0.04145, which is far below the CCA-T^2 method. Also, the PAR value of the basic PCA method is the highest, which is up to 0.36913. This is mainly because the wastewater treatment process contains complex biochemical reactions, and the corresponding process information usually exists in the data with nonlinear structure. CCA and PCA are linear methods, which can not extract the information of the nonlinear data effectively. Moreover, the nonlinear method of KPCA cannot show its talents. The accuracy of KPCA is low than CCA, and it is attributed to the KPCA method that cannot capture the structural information of the input space and output space. Besides, the M-NAKCCA uses MEm and CRD to compensate for the nonlinear data, and it further enhances the method performance. Therefore, the proposed M-NAKCCA method is promising for monitoring the wastewater treatment process.

5. Conclusion

Process monitoring and risk assessment are indispensable for the process industries, implicating accidence decrement, cost-saving and environment protection. To deal with this problem, a novel M-NAKCCA framework is proposed in this paper. The proposed method can identify diverse types of faults (drift fault, step fault, and random fault), with the ability to indicate the abnormal events. Under the M-NAKCCA monitoring framework, the raw dataset is firstly processed by the matter-element model. Secondly, the nonlinear CCA based on kernel function is developed. Meanwhile, Nyströ m

technology is effectively used to estimate the low-rank approximation of the Gram matrix, whose purpose is to reduce time consumption and improve the information retention capacity of the low-rank matrix. Finally, T^2 control chart is reconstructed by the residual of input-output to monitor the quality-relevant faults.

In this paper, the proposed M-NAKCCA framework is used to monitor a wastewater treatment process and a TE chemical process. The comparison study with the other basic methods (PCA, CCA and KPCA) demonstrates that the proposed method is more satisfactory in terms of FAR, MAR, PAR, and accuracy. Moreover, the result illustrates that the proposed method is effective for the detection of quality-related faults. However, since the external interference and internal processes behavior is too complicated, it is inevitable that the system failures will propagate across the entire process and to intertwine with each other. This can be resorted to incipient fault detection and fault locations. Unfortunately, the proposed M-NAKCCA framework only focuses on fault detection, while ignoring the location of the root cause of the fault. In future research, the M-NAKCCA framework will be further expanded to locate the root cause of faults and deal with more complex industrial problems.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. paper.

Acknowledgements:

This work was supported by the National Natural Science Foundation of

China (61873096, 61673181), Guangdong Basic and Applied Basic Research Foundation (2020A1515011057), Guangdong Technology International Cooperation Project Application (2020A0505100024). Fundamental Research Funds for the central Universities, SCUT (2020ZYGXZR034). Qilin Wang acknowledges Australian Research Council (ARC) Discovery Project DP200100933. The author thanks the anonymous referees and other people for their help to improve the article.

References:

Adedigba, S.A., Khan, F., Yang, M., 2017. Dynamic failure analysis of process systems using principal component analysis and Bayesian network. Industrial & Engineering Chemistry Research 56, 2094-2106.

Baklouti, I., Mansouri, M., Hamida, A.B., Nounou, H., Nounou, M., 2018. Monitoring of wastewater treatment plants using improved univariate statistical technique. Process Safety and Environmental Protection 116, 287-300.

Chen, Z., Zhang, K., Ding, S.X., Shardt, Y.A.W., Hu, Z., 2016. Improved canonical correlation analysis-based fault detection methods for industrial processes. Journal of Process Control 41, 26-34.

Cheng, H., Liu, Y., Huang, D., Liu, B., 2019a. Optimized Forecast Components-SVM-Based Fault Diagnosis With Applications for Wastewater Treatment. IEEE Access 7, 128534-128543.

Cheng, H., Liu, Y., Huang, D., Pan, Y., Wang, Q., 2020a. Adaptive Transfer Learning of Cross-Spatiotemporal Canonical Correlation Analysis for Plant-Wide Process Monitoring. Industrial & Engineering Chemistry Research 59, 21602-21614.

Cheng, H., Liu, Y., Huang, D., Xu, C., Wu, J., 2020b. A Novel Ensemble Adaptive Sparse Bayesian Transfer Learning Machine for Nonlinear Large-Scale Process Monitoring. Sensors 20, 6139.

Cheng, H., Wu, J., Liu, Y., Huang, D., 2019b. A novel fault identification and root-causality analysis of incipient faults with applications to wastewater treatment processes. Chemometrics and Intelligent Laboratory Systems 188, 24-36.

Dong, D., Mcavoy, T.J., 1996. Batch tracking via nonlinear principal component analysis. Aiche Journal 42.

Downs, J.J., Vogel, E.F., 1993. A plant-wide industrial process control problem. Computers & chemical engineering 17, 245-255.

Fowlkes, C., Belongie, S., Chung, F., Malik, J., 2004. Spectral grouping using the Nystrom method. IEEE transactions on pattern analysis and machine intelligence 26, 214-225.

Ge, Z., Zhang, M., Song, Z., 2010. Nonlinear process monitoring based on linear subspace and Bayesian inference. Journal of Process Control 20, 676-688.

Gong, J., Liu, Y., Chen, W., 2012. Land suitability evaluation for development using a matter-element model: A case study in Zengcheng, Guangzhou, China. Land Use Policy 29, 0-472.

Jiang, Q., Ding, S.X., Yang, W., Yan, X., 2017. Data-Driven Distributed Local Fault Detection for Large-Scale Processes Based on the GA-Regularized Canonical Correlation Analysis. IEEE Transactions on Industrial Electronics 64, 8148-8157.

Jiang, Q., Yan, X., 2015. Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA. Journal of Process Control 32, 38-50.
Juricek, B.C., Seborg, D.E., Larimore, W.E., 2004. Fault Detection Using Canonical Variate Analysis. Industrial & Engineering Chemistry Research 43, 458-474.
Kramer, M.A., 1991. Nonlinear principal component analysis using autoassociative neural networks. Aiche Journal 37.
Larimore, W.E., 1997. Optimal Reduced Rank Modeling, Prediction, Monitoring and Control using Canonical Variate Analysis. IFAC Proceedings Volumes 30, 61-66.
Li, M., Kwok, J.T., Lu, B., 2010. Making Large-Scale Nyström Approximation Possible, International Conference on International Conference on Machine Learning.
Liu, Y., Liu, B., Zhao, X., Xie, M., 2018. A Mixture of Variational Canonical Correlation Analysis for Nonlinear and Quality-Relevant Process Monitoring. IEEE Transactions on Industrial Electronics 65, 6478-6486.
Liu, Y., Xiao, H., Pan, Y., Huang, D., Wang, Q., 2016. Development of multiple-step soft-sensors using a Gaussian process model with application for fault prognosis. Chemometrics & Intelligent Laboratory Systems 157, 85-95.
Liu, Y., Xie, M., 2020. Rebooting data-driven soft-sensors in process industries: A review of kernel methods - ScienceDirect. Journal of Process Control 89, 58-73.
Ma, L., Dong, J., Peng, K., 2019. A Novel Hierarchical Detection and Isolation Framework for Quality-Related Multiple Faults in Large-Scale Processes. IEEE Transactions on Industrial Electronics, 1-1.
Odiowei, P.P., Cao, Y., 2010. State-space independent component analysis for nonlinear dynamic process monitoring. Chemometrics and Intelligent Laboratory Systems 103, 59-65.
Olsson, G., 2012. ICA and me – A subjective review. Water Research 46, 1585-1624.
Samuel, R.T., Cao, Y., 2015. Kernel Canonical Variate Analysis for Nonlinear Dynamic Process Monitoring. IFAC-PapersOnLine 48, 605-610.
Shen, Y., Ding, S.X., Haghani, A., Hao, H., Ping, Z., 2012. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. Journal of Process Control 22, 1567---1581.
Williams, C.K.I., Seeger, M., 2000. Using the Nyström Method to Speed Up Kernel Machines, Advances in Neural Information Processing Systems 13, Papers from Neural Information Processing Systems (NIPS) 2000, Denver, CO, USA.
Yang, C., Wen, C., 2015. Proceedings of the 2nd International Symposium on Extenics and Innovation Methods. Heredity 4, 313-325.
Yong-xiu, H., Ai-ying, D., Jiang, Z., Hai-ying, H., Li, F., 2011. Risk assessment of urban network planning in china based on the matter-element model and extension analysis. International Journal of Electrical Power & Energy Systems 33, p.775-782.
Yu, H., Khan, F., Garaniya, V., 2016. A sparse PCA for nonlinear fault diagnosis and robust feature discovery of industrial processes. AIChE Journal 62, 1494-1513.
Yu, J., Wang, K., Ye, L., Song, Z., 2019. Accelerated Kernel Canonical Correlation Analysis with Fault Relevance for Nonlinear Process Fault Isolation. Industrial & Engineering Chemistry Research 58, 18280-18291.
Yuan, X., Liu, X., Zhu, Q., 2013. Research and application of extension PCA fault diagnosis method for chemical process. Computers & Applied Chemistry.
Zhu, Q., Liu, Q., Qin, S.J., 2016. Concurrent canonical correlation analysis modeling for quality-relevant

monitoring. IFAC-PapersOnLine 49, 1044-1049.