Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

A Study of Complex Industrial Systems Based on Revised Kernel Principal Component Regression Method

Chengyuan Sun, * **Hongjun Ma** *,**

* *College of information science and engineering, Northeastern University, Shenyang, 110819, China*
** *Key Laboratory of Knowledge Automation for Process Industries, Ministry of Education, Beijing, 100083, China (e-mail: mahongjun@ise.neu.edu.cn).*

Abstract: As a data-driven process monitoring method, multivariable statistics techniques have special potentials and advantages to handle the increasingly prominent "Big data challenge" in the complex industrial systems. However, the standard partial least square (PLS) method and the principal component regression (PCR) method cannot maintain stable function in the nonlinear operating environment. In order to capture the precise relation of process variables and product variables, an approach called the revised kernel PCR (RKPCR) method is proposed in this thesis to resolve the problems encountered in the traditional PCR method. In addition, a brief and effective diagnosis logic is designed to decrease the difficulty of fault diagnosis. Finally, the effectiveness of the RKPCR algorithm is illustrated utilizing the Tennessee Eastman case (TEC) platform.

Keywords: Nonlinear process, the RKPCR method, quality-related, process monitoring.

1. INTRODUCTION

For the past quarter-century, fault detection methods have been applied to large scale plants and processes, which can effectively enhance the core competitiveness of the enterprises and dramatically reduce needless costs of the production procedures (Peng et al. (2013); Shao et al. (2019); Yin et al. (2012)). In the real industrial process, many types of equipment are working together to make the systems become more complex and undetectable. Without considering the precise mathematical model, data-driven approaches can directly derive the information of the systems from an ocean of data, which cope with the multivariable operation conditions of modern industries, for instance, chemical process and semiconductor manufacturing (Luo et al. (2019); Qin and Dong (2018)). To control these multivariable products, multivariate statistical process monitoring (MSPM) is leveraged to these increasingly complex processes. The principal component analysis (PCA) algorithm and the partial least square (PLS) algorithm, two common tools of MSPM, can find the abnormal changes in the process measurements. Not like the PCA method focusing only on process measurements, the PLS method can interpret the relationship between product and process variables, which fits the modern systems control (Zhang et al. (2019)).

However, the decomposition of PLS method is performed incompletely that further influences the subsequent fault detection results. Li et al. (2010) described a geometric expression for the structure of the PLS method to promote the monitoring results. In order to surmount the remaining unexpected alarms, Zhou et al. (2010) proposed the total PLS (T-PLS) method to enforce a follow-up operation on the separation of the PLS approach. Sooner, Yin et al. (2011) worked on the connection between quality variables and process variables to shape the new form, namely the modified PLS (M-PLS) technique. Combined with these mentioned two techniques, Qin and Zheng (2013) drafted a current PLS (C-PLS) approach to separate process variables into two sections by means of the prediction of output variables. These approaches open new avenues for the promotion of quality-related process monitoring. Thanks to the contributions of these scholars, the linear process monitoring aspects have obtained enough fruits. From the feedbacks of technicians, the PLS method cannot be very stable when the strength of fault rapidly enhances. What is worth mentioning, product quality occupying the first spot in the modern production process, which cannot be considered by the PCA algorithm like the PLS method (Dong et al. (2015); Zhang et al. (2015b); Wang et al. (2017)).

Motivated by improving the PCA algorithm, the principal component regression (PCR) is adopted to the quality-related process monitoring (Peng et al. (2015)). Compared with these existed PLSs' methods, the PCR approach is very suitable to handle the industrial process with a small number of quality variables. Nevertheless, in the real operating procedure, the feature of nonlinear always exists in these running data so that the aforementioned linear methods cannot guarantee the performance of detection results (Ge et al. (2009); Jiang et al. (2015)). Driven by this trend, Peng et al. (2013) and Jiao et al. (2017) developed the T-PLS method and the M-PLS method into nonlinear settings, respectively. Zhang et al. (2015a) analyzed and summed up these former techniques in the paper.

Hence, more researches about the nonlinear methods should be taken into account. What is more, from the knowledge of engineers and workers, a concise detection strategy is helpful and beneficial to operate (Wang et al. (2018)). These existing techniques may confront the d-ifficulty of incorrect results and complicated diagnosis regimes. Similar to the PLS method, the PCR algorithm also projects the process variables into two adverse meaning subspaces or parts. For two divided portions, every part may have some components which should belong to another one. Hence, this imperfect structure hugely influences the detection results so that the subsequent fault diagnosis and fault-tolerant control are contaminated.

In this paper, prompted by the aforementioned flaws, a novel method exploiting the connection between quality measurements and process measurements is proposed to induct an accurate division, which is called the revised kernel PCR (RKPCR) approach. The reliability of the proposed RKPCR is demonstrated on the Tennessee Eastman case (TEC) by comparing the monitored results of the T-KPLS method and the KPLS method.

The main contributions of this paper are listed as follows.

a. The RKPCR algorithm complements the monitoring approaches ground on the nonlinear circumstances, and it is more appropriate to the progress of the real complex industry.

b. The proposed method rewrites the separation of the process variables which can be employed to detect the fault met in the running process more exactly.

c. A concise diagnosis logic is provided that can effectively reduce the computation and difficulty of online manipulation.

The rest of this paper is organized as follows. In Section II, the intrinsic drawbacks of the PCR method is discussed. Section III reveals the detailed proving process of the proposed approach. In the TE case of Section IV, the availability of the RKPCR approach is illustrated with the comparisons of two methods. Eventually, the conclusions are drawn in Section V.

2. REVIEWS OF THE PCR METHOD

2.1 Presentation of the PCR method

Firstly, let's define the collected process variables as $\mathbf{X} \in \mathbb{R}^{N\times m}$ and quality variables as $\mathbf{Y} \in \mathbb{R}^{N\times j}$ (the definition of symbols is detailed as Appendix A).

Through projecting the variables into latent spaces with a lower dimension, the PCA method has ability to monitor the abnormal alterations in the running conditions. Combined with the defined $\mathbf{X}$ and $\mathbf{Y}$, and the PCA method can be expressed as:

$$\begin{cases} \mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}, \\ \mathbf{P} = \mathbf{X}^T\mathbf{T}, \end{cases} \tag{1}$$

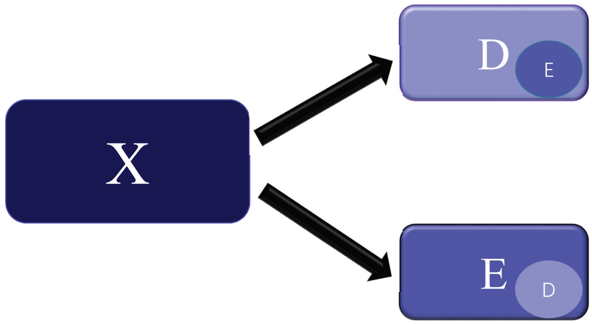

Fig. 1. Decomposition plot of $\mathbf{X}$ by the PCR method.

Algorithm 1 The procedure of the PCR method

Step 1: Normalize the data sets.
Step 2: Calculate the covariance matrix $\mathbf{S}$.
Step 3: Perform eigenvalue decomposition (EVD) on $\mathbf{S}$.
Step 4: Array the eigenvalue in the descending order.
Step 5: Select the first A eigenvalue.
Step 6: Group the corresponding eigenvector of A eigenvalue as $\mathbf{P}$.
Step 7: Calculate the score matrix $\mathbf{T}$, $\mathbf{T} = \mathbf{XP}$.
Step 8: Obtain the score matrix of $\mathbf{Y}$ based on (2).
Step 9: Acquire the further elaborated $\mathbf{X}$ as (4).

where $\mathbf{P} \in \mathbb{R}^{m\times A}$ and $\mathbf{T} \in \mathbb{R}^{N\times A}$ denote projected loading and score matrix of $\mathbf{X}$, respectively. A is the number of principal components, and $\mathbf{E}$ represents the residual part.

However, the PCA method only concerns fault and the affected process variables. Therefore, the PCR algorithm is improved to solve the problem of the PCA method that cannot obtain the link of product properties and faults (Jolliffe (1982)). Here, the PCR approach is applied to perform decomposition of $\mathbf{X}$ according to the prediction of quality variables. Before performing the PCR method, the important step is to normalize the data collected from the sensors of the real systems.

The least squares (LS) algorithm is used to construct the relationship between $\mathbf{T}$ and $\mathbf{Y}$.

$$\mathbf{Q}^T = (\mathbf{T}^T\mathbf{T})^{(-1)}\mathbf{T}^T\mathbf{Y}, \tag{2}$$

where $\mathbf{Q}$ means the score variables of $\mathbf{Y}$.

Then

$$\hat{\mathbf{Y}} = \mathbf{T}\mathbf{Q}^T. \tag{3}$$

The PCR algorithm can be constructed as follows:

$$\begin{cases} \mathbf{X} = \mathbf{D} + \mathbf{E} = \mathbf{T}\mathbf{P}^T + \mathbf{E}, \\ \hat{\mathbf{Y}} = \mathbf{T}\mathbf{Q}^T = \mathbf{XP}\mathbf{Q}^T, \end{cases} \tag{4}$$

where $\mathbf{D}$ and $\mathbf{E}$ represent quality-related part and quality-unrelated part of $\mathbf{X}$. The procedure of the PCR method is shown as **Algorithm** 1.

2.2 Drawbacks of the PCR method

Although $\mathbf{X}$ is decomposed by the relationship with $\mathbf{Y}$, the separated procedure is not very consummate that further compromises the detection results in the process monitoring. The so-called quality-related part may have the components of quality-unrelated. By the same token, the quality-unrelated part may not only contain itself. The inhered defects of the PCR method are as displayed in Fig.1. Besides, the PCR method and the PLS method are utilized under an implicit assumption that all operations are in the linear conditions. Consequently, these above techniques will lead to poor monitoring performance when they are employed in the nonlinear environment. For the purpose of solving these issues, in the next section, the revised kernel PCR approach is proposed.

3. THE REVISED KERNEL PCR METHOD

3.1 Kernel techniques

These above linear methods are only playing a part as they running near the operating points. Hence, there are inevitable that these above linear methods cannot work well in the nonlinear aspects. If the nonlinear features in the collected data sets cannot be solved that further limit the application of the PCR method and the PLS method. Rosipal and Trejo (2001) proposed the kernel PLS (KPLS) method to give a connection between quality variables and process variables under the nonlinear conditions, which introduced the kernel technique to the area of MSPM. Not only the KPLS method can excavate the detailed information of obtained measurements but also it can enhance the precision of monitoring performance. The core of the kernel techniques is to project the process data into the high dimensions to obtain an approximation of the linear model.

The three kernel techniques are:

1) **Polynomial kernel**

$$f_k(x,y) = < x, y >^d, \tag{5}$$

2) **Sigmoid kernel**

$$f_k(x,y) = \tanh(\beta_o < x,y > +\beta_1), \tag{6}$$

3) **Gaussian kernel**

$$f_k(x,y) = exp(-\frac{\parallel x-y \parallel^2}{c}). \tag{7}$$

3.2 The RKPCR method

According to the many applications of scholars and engineers, the Gaussian kernel is the most useful in these three kernel methods as it has the lower computation burden and the higher convenience. Therefore, the Gaussian kernel is adopted in this part. Firstly, the process variables are projected into feature space with high dimension,

$$\mathbf{X} \in \mathbb{R}^{N\times m} \to \mathbf{\Phi} \in \mathbb{R}^{N\times M}, \tag{8}$$

where $\mathbf{\Phi}$ is the infinite dimension, $M \in \infty$.

Then

$$\begin{cases} \mathbf{\Phi} = \mathbf{T}\mathbf{P}^T + \mathbf{\Phi}_e \\ \mathbf{P} = \mathbf{\Phi}^T\mathbf{T} \end{cases}. \tag{9}$$

From (4) and (9), $\mathbf{\Phi}$ and $\hat{\mathbf{Y}}$ can be deduced as

$$\hat{\mathbf{Y}} = \mathbf{T}\mathbf{Q}^T = \mathbf{\Phi}\mathbf{P}\mathbf{Q}^T = \mathbf{\Phi}\mathbf{H_k}\,, \tag{10}$$

where $\mathbf{H_k}$ denotes the coefficient matrix between $\mathbf{\Phi}$ and $\hat{\mathbf{Y}}$.

Theorem. Notice that before applying the PCR method, zero-mean step must be first considered.

$$\bar{\mathbf{\Phi}} = \begin{bmatrix} \phi(\mathbf{x}_1) - \bar{\phi} \\ \phi(\mathbf{x}_2) - \bar{\phi} \\ \vdots \\ \phi(\mathbf{x}_N) - \bar{\phi} \end{bmatrix} = \mathbf{\Phi} - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T\mathbf{\Phi}\,, \tag{11}$$

where $\mathbf{1}_N = \frac{1}{N}\begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}$, $\bar{\phi}$ is the mean value of ϕ.

However, the number of columns can be infinity, $\bar{\mathbf{\Phi}}$ cannot be calculated directly. So $\bar{\mathbf{f}}_k$ is utilized to substitute for $\bar{\mathbf{\Phi}}$ (Rosipal and Trejo (2001)).

$$\bar{\mathbf{f}}_k = \bar{\mathbf{\Phi}}\bar{\mathbf{\Phi}}^T\,. \tag{12}$$

Assumption. Based on (9), assuming a division as (13) that provides a detailed expression to make sense of process variables with quality.

$$\begin{cases} \bar{\mathbf{f}}_k = \bar{\mathbf{f}}_{ka} + \bar{\mathbf{f}}_{kb}, \\ \hat{\mathbf{Y}} = \bar{\mathbf{f}}_{ka}\mathbf{H}_k, \end{cases} \tag{13}$$

where $\bar{\mathbf{f}}_{ka}$ and $\bar{\mathbf{f}}_{kb}$ mean the correct quality-related projected subspace and correct quality-unrelated one, respectively.

Proof. On the basis of (9), (10), and (12), the link between $\hat{\mathbf{Y}}$ and $\bar{\mathbf{f}}_k$ can be calculated as

$$\begin{aligned} \hat{\mathbf{Y}} &= \bar{\mathbf{f}}_k\mathbf{P}\mathbf{Q}^T, \\ &= \bar{\mathbf{f}}_k\mathbf{H}_k, \end{aligned} \tag{14}$$

where $\mathbf{H}_k$ represents the new connection between $\hat{\mathbf{Y}}$ and $\bar{\mathbf{f}}_k$.

Firstly, the singular value decomposition (SVD) is performed on the matrix $\mathbf{H}_k\mathbf{H}_k^T$,

$$\mathbf{H}_k\mathbf{H}_k^T = \begin{bmatrix} \amalg_k & \tilde{\amalg}_k \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \amalg_k^T \\ \tilde{\amalg}_k^T \end{bmatrix}, \tag{15}$$

where $\amalg_k \in \mathbb{R}^{N\times j}$, $\tilde{\amalg}_k \in \mathbb{R}^{N\times(N-j)}$ and $\mathbf{\Lambda}_k \in \mathbb{R}^{j\times j}$.

Then the projection matrices of $\bar{\mathbf{\Phi}}$ is defined

$$\mathbf{B}_k = \amalg_k\amalg_k^T \in \mathbb{R}^{N\times N} \tag{16}$$

$$\mathbf{B}_k^\star = \tilde{\amalg}_k\tilde{\amalg}_k^T \in \mathbb{R}^{N\times N} \tag{17}$$

For these two parts, two score matrixes are constructed as

$$\mathbf{T}_a = \bar{\mathbf{f}}_k\amalg_k \in \mathbb{R}^{N\times j} \tag{18}$$

$$\mathbf{T}_b = \bar{\mathbf{f}}_k\tilde{\amalg}_k \in \mathbb{R}^{N\times(N-j)} \tag{19}$$

Hence, $\bar{\mathbf{f}}_k$ can be separated into two parts, $\bar{\mathbf{f}}_{ka}$ and $\bar{\mathbf{f}}_{kb}$, based on $\mathbf{B}_k^\star$ and $\mathbf{B}_k$.

$$\bar{\mathbf{f}}_{ka} = \bar{\mathbf{f}}\mathbf{B}_k = \mathbf{T}_a\amalg_k^T \in S_{\bar{\mathbf{\Phi}}_a} \tag{20}$$

$$\bar{\mathbf{f}}_{kb} = \bar{\mathbf{f}}\mathbf{B}_k^\star = \mathbf{T}_b\tilde{\amalg}_k^T \in S_{\bar{\mathbf{\Phi}}_b} \tag{21}$$

Then according to the feature of SVD, (22) and (23) can be obtained.

$$\mathbf{B}^{\star}_{k} + \mathbf{B}_k = \mathbf{I}_N \,, \tag{22}$$

$$\begin{aligned} &\amalg_k^T \tilde{\amalg}_k = \tilde{\amalg}_k^T \amalg_k = 0, \\ &\tilde{\amalg}_k^T \mathbf{H}_k = 0, \end{aligned} \tag{23}$$

then

$$\bar{\mathbf{f}}_k = \bar{\mathbf{f}}_{ka} + \bar{\mathbf{f}}_{kb} = \bar{\mathbf{f}}(\mathbf{B}^{\star}_{k} + \mathbf{B}_k) \,, \tag{24}$$

and then

$$\begin{aligned} \hat{\mathbf{Y}} &= \bar{\mathbf{f}}_k \mathbf{H}_k = \bar{\mathbf{f}}_k(\mathbf{B}^{\star}_{k} + \mathbf{B})\mathbf{H}_k, \\ &= \bar{\mathbf{f}}_k(\amalg_k \amalg_k^T + \tilde{\amalg}_k \tilde{\amalg}_k^T)\mathbf{H}_k, \\ &= \bar{\mathbf{f}}_k \amalg_k \amalg_k^T \mathbf{H}_k, \\ &= \bar{\mathbf{f}}_{k_a} \mathbf{H}_k. \end{aligned} \tag{25}$$

Accordingly, this assumption is proved. This Assumption will offer valuable detection results to the quality-related process monitoring.

3.3 Implement of the RKPCR method

When a new data $\mathbf{x}_{new}$ is going to perform the RKPCR method, data needs to extend into the high dimensional space as ϕ_{new}. Next, ϕ_{new} is calculated by zero-mean step described in Theorem. Similar to underlying properties of $\bar{\mathbf{\Phi}}$ that cannot be straightly measured, based on (12), it can be depicted as

$$\bar{\mathbf{f}}_{k_{new}} = \bar{\mathbf{\Phi}}_{new} \bar{\mathbf{\Phi}}_{new}^T \,. \tag{26}$$

For each online sample $\bar{\mathbf{f}}_{k_{new}}$, it can be divided into two mutually orthogonal portions by the RKPCR method, $\bar{\mathbf{f}}_{ka_{new}}$ and $\bar{\mathbf{f}}_{kb_{new}}$,

$$\begin{cases} \bar{\mathbf{f}}_{knew} &= \bar{\mathbf{f}}_{ka_{new}} + \bar{\mathbf{f}}_{kb_{new}}, \\ \hat{\mathbf{y}} &= \bar{\mathbf{f}}_{ka_{new}} \mathbf{H}_{knew}. \end{cases} \tag{27}$$

Then the score vectors of $\bar{\mathbf{f}}_{ka_{new}}$ and $\bar{\mathbf{f}}_{kb_{new}}$ are

$$\mathbf{t}_a = \bar{\mathbf{f}}_{knew} \amalg_{knew} \tag{28}$$

$$\mathbf{t}_b = \bar{\mathbf{f}}_{knew} \tilde{\amalg}_{knew} \tag{29}$$

The $\mathbf{T}^2$ statistics is chosen to depict the varies of $\bar{\mathbf{f}}_{ka_{new}}$ and $\bar{\mathbf{f}}_{kb_{new}}$, and they are listed as described below,

$$\mathbf{T}_a^2 = \mathbf{t}_a^T \left(\frac{\mathbf{T}_a^T \mathbf{T}_a}{N-1} \right)^{-1} \mathbf{t}_a, \tag{30}$$

$$\mathbf{T}_b^2 = \mathbf{t}_b^T \left(\frac{\mathbf{T}_b^T \mathbf{T}_b}{N-1} \right)^{-1} \mathbf{t}_b. \tag{31}$$

Then the thresholds of $\mathbf{T}_a^2$ and $\mathbf{T}_b^2$ can be worked out as the significance level α is given, and

$$\mathbf{J}_{th,T_a^2} = \frac{m\left(N^2-1\right)}{N\left(N-m\right)} \mathbf{F}_{m,N-m,\alpha}, \tag{32}$$

$$\mathbf{J}_{th,T_b^2} = \frac{\left(A-m\right)\left(N^2-1\right)}{N\left(N-A+m\right)} \mathbf{F}_{A-m,N-A+m,\alpha}. \tag{33}$$

The fault diagnosis strategy based on the RKPCR method is presented as

Algorithm 2 Process monitoring based on the RKPCR algorithm

Training phase:
Step 1: Calculate $\bar{\mathbf{f}}_k$ according to (11) and (12).
Step 2: Carry out Algorithm 1 on new data set ($\bar{\mathbf{f}}_k$, $\mathbf{Y}$) to acquire $\mathbf{H_k}$.
Step 3: Perform SVD on the matrix $\mathbf{H}_k \mathbf{H}_k^T$.
Step 4: Obtain the projection matrices $\mathbf{B}^{\star}_{k}$ and $\mathbf{B}_k$ on the basis of (16) and (17).
Step 5: Calculate score matrixes $\mathbf{T}_a$ and $\mathbf{T}_b$ based on (18) and (19).
Step 6: Set α and compute the control thresholds by (32) and (33).
***Testing phase*:**
Step 1: Form the collected new data $\mathbf{x}_{new}$ into $\bar{\mathbf{f}}_{k_{new}}$ by (11), (12) and (26).
Step 2: Calculate the score matrix vectors $\mathbf{t}_a$ and $\mathbf{t}_b$ according to (28) and (29).
Step 3: Obtain $\mathbf{T}_a^2$ and $\mathbf{T}_b^2$ based on (30) and (31).
Step 4: Distinguish the kind of fault condition based on the fault diagnosis strategy logic.

(1) $\mathbf{T}_a^2 \geq \mathbf{J}_{th,T_a^2}$

$\Longrightarrow$ A fault related with quality is detected and judged;

(2) $\mathbf{T}_a^2 < \mathbf{J}_{th,T_b^2}$ and $\mathbf{T}_b^2 \geq \mathbf{J}_{th,T_b^2}$

$\Longrightarrow$ A fault unrelated with quality is detected and judged;

(3) $\mathbf{T}_a^2 < \mathbf{J}_{th,T_a^2}$ and $\mathbf{T}_b^2 < \mathbf{J}_{th,T_b^2}$

$\Longrightarrow$ Detected no-fault;

The procedure of quality-related process monitoring based on the RKPCR method is shown as Algorithm 2.

4. APPLICATION ON THE TE CASE

The Tennessee Eastman Case (TEC) is a simulation benchmark created by the Eastman company in 1993 based on the real industrial process (Downs and Vogel (1993)). The TEC is a continuous process applied to test and verify the effectiveness of some monitoring and optimization algorithms. The reacted procedure of the TEC is as presented as (34), including 8 components (A-G), where 4 gas components, A, C, D, and E are reacted together to produce liquid products G and H, and also accompanying with liquid side stuff F and gas B. 21 faulty data sets and a no-fault one is supplied by the TEC, where every set consists of 500 training samples and 960 testing samples.

$$\begin{cases} A + C + D \rightarrow G \\ A + C + E \rightarrow H \\ A + E \rightarrow F \\ 3D \rightarrow 2F \end{cases} \tag{34}$$

From prior knowledge (Yin et al. (2012)), 21 faults can be divided into 2 classes, quality-related faults (IDV(1), IDV(2), IDV(5), IDV(6), IDV(10), etc), and quality-unrelated faults (IDV(3), IDV(4), IDV(9), IDV(11), IDV(14), and IDV(19)). The operating procedure of the TEC comprises 12 control measurements (XMV(1-12)) and 41 measured variables (XMEAS(1-41)), in which XMEAS(23-41) are components measurements. In this study, XMV(1-11) and XMEAS(1-22) are selected to compose the process variables $\mathbf{X}$. The components measurement XMEAS(35), the product G, is selected as $\mathbf{Y}$. Here, the training data of IDV(3), IDV(4), IDV(9), IDV(11), IDV(14), and IDV(19) is adopt to train the models of the T-KPLS method (Peng et al. (2013)), the KPLS method, and the RKPCR method, respectively. These three approaches are utilized to detect the preset fault in the testing data (faults are added in the 165th sample).

Table 1. FD rates of the KPLS method, the T-KPLS method, and the RKPCR method for 6 kinds of faults (%)

Faults	KPLS		T-KPLS		RKPCR	
	$\mathbf{T}^2_{kpls}$	$\mathbf{Q}_{kpls}$	$\mathbf{T}^2_{ko}$&$\mathbf{T}^2_{kr}$	$\mathbf{T}^2_{ky}$&$\mathbf{Q}_r$	$\mathbf{T}^2_a$	$\mathbf{T}^2_b$
IDV(3)	2.63	**0.88**	64.62	**1.88**	6.50	**24.5**
IDV(4)	99.87	**0**	58.47	**99.87**	0	**96.5**
IDV(9)	1.51	**0.25**	26.98	**1.38**	5.25	**21.5**
IDV(11)	73.9	**39.4**	97.49	**73.78**	9.50	**84.25**
IDV(14)	100	**97.74**	100	**100**	3.62	**100**
IDV(19)	54.58	**9.03**	91.84	**50.44**	1.00	**91.75**

The detection results of three methods under IDV(14) are shown as Fig. 2, Fig. 3, and Fig. 4. In Fig. 2, quality-related $\mathbf{T}^2_{kpls}$ and quality-unrelated $\mathbf{Q}_{kpls}$ all exceed the thresholds, indicating that the KPLS approach cannot depict running condition affected by fault IDV(14). As displayed in Fig. 3, the T-KPLS method provides false alarms that four statistical magnitudes are topper than the red line, where $\mathbf{T}^2_{ko}$&$\mathbf{T}^2_{kr}$ represents quality-related subspace and $\mathbf{T}^2_{ky}$&$\mathbf{Q}_r$ denotes quality-unrelated subspace. Beginning from the 165th sample in Fig. 4, it is clear that $\mathbf{T}^2_a$ is almost under the set threshold, and $\mathbf{T}^2_b$ exceeds the red line to raise an alarm. From the FD strategy of the RKPCR approach, it can come to a point that the fault IDV(14) is diagnosed as a quality-unrelated kind, which is accorded with the prior knowledge.

Fault detection (FD) rate, a common measured index, is employed to judge the detection property of methods as (35),

$$\textbf{FD rate} = \frac{\textbf{No.of detected samples}}{\textbf{total samples}} \times \mathbf{100}\%. \quad (35)$$

The FD rates of IDV(3), IDV(4), IDV(9), IDV(11), IDV(14), and IDV(19) are exhibited in Table 1. From the presented detection results of quality-unrelated faults, the higher FD rates are submitted by the corresponded subspace of the RKPCR method than the KPLS method and the T-KPLS method.

Consequently, based on the monitoring results, it can be clearly concluded that the RKPCR approach is more beneficial to monitor the faults in the running conditions compared with the KPLS approach and the T-KPLS approach. Different from four statistical variables of the T-KPLS technique, the RKPCR approach can depict the changing conditions with only two ones. Hence, the RKPCR technique also provides a more simple and better strategy than the KPLS method and the T-KPLS approach.

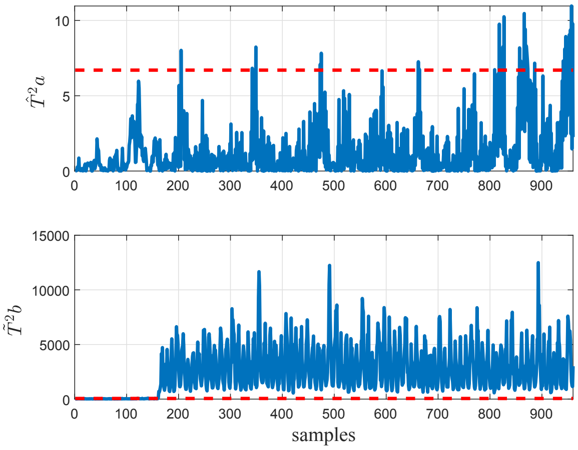

Fig. 2. Detection plot of the RKPCR method of IDV(14)

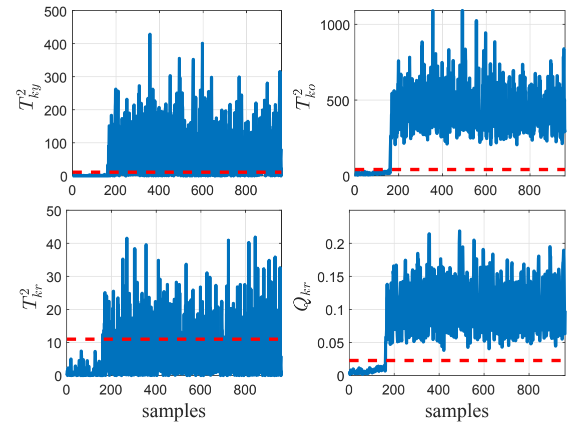

Fig. 3. Detection plot of the T-KPLS method of IDV(14)

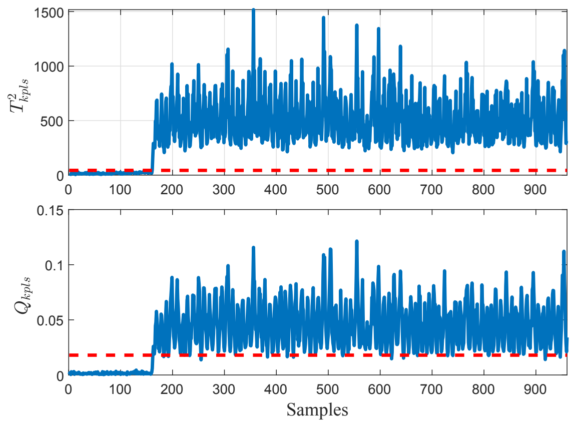

Fig. 4. Detection plot of the KPLS method of IDV(14)

5. CONCLUSION

In this thesis, the RKPCR technique has been devised to improve the quality-related monitoring performance of the standard PCR method by employing the reconstructed relationship between quality measurements and process measurements. Meanwhile, the designed diagnosis logic has been testified reliable and with less computation burden than the compared approaches. The proposed method has outperformed these compared approaches in light of the accorded experiment results of the TEC. Hence, the designed approach can be implemented to deal with the nonlinear behaviors in the running process.

ACKNOWLEDGEMENTS

This work was supported in part by the funds of National Science of Foundation China (Grant nos. 61873306, U1908213, 6162100, 61420106016), the Fundamental Research Funds for the Central Universities (Nos. N2004018, N170308028, N182608004, FRF-BD-19-002A), and the research fund of the State Key Laboratory of Synthetical Automation for Process Industries (Grant nos. SAPI2019-3, 2018ZCX19).

Appendix A. THE SYMBOLIC MEANING

$\mathbb{R}^j$	j response variables
$\mathbb{R}^{N\times M}$	N observations and m variables
$\mathbf{F}(a,b,\alpha)$	$\mathbf{F}$ distribution
$\mathbf{I}_N$	N dimensions unit matrix
S	span

REFERENCES

Dong, J., Zhang, K., Huang, Y., Li, G., and Peng, K. (2015). Adaptive total pls based quality-relevant process monitoring with application to the tennessee eastman process. *Neurocomputing*, 154, 77–85.

Downs, J.J. and Vogel, E.F. (1993). A plant-wide industrial process control problem. *Computers & chemical engineering*, 17(3), 245–255.

Ge, Z., Yang, C., and Song, Z. (2009). Improved kernel pca-based monitoring approach for nonlinear processes. *Chemical Engineering Science*, 64(9), 2245–2255.

Jiang, Q., Yan, X., and Huang, B. (2015). Performance-driven distributed pca process monitoring based on fault-relevant variable selection and bayesian inference. *IEEE Transactions on Industrial Electronics*, 63(1), 377–386.

Jiao, J., Zhao, N., Wang, G., and Yin, S. (2017). A nonlinear quality-related fault detection approach based on modified kernel partial least squares. *ISA transactions*, 66, 275–283.

Jolliffe, I.T. (1982). A note on the use of principal components in regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(3), 300–303.

Li, G., Qin, S.J., and Zhou, D. (2010). Geometric properties of partial least squares for process monitoring. *Automatica*, 46(1), 204–210.

Luo, H., Yin, S., Liu, T., and Khan, A.Q. (2019). A data-driven realization of the control-performance-oriented process monitoring system. *IEEE Transactions on Industrial Electronics*.

Peng, K., Zhang, K., and Li, G. (2013). Quality-related process monitoring based on total kernel pls model and its industrial application. *Mathematical Problems in Engineering*, 2013.

Peng, K., Zhang, K., You, B., and Dong, J. (2015). Quality-related prediction and monitoring of multi-mode processes using multiple pls with application to an industrial hot strip mill. *Neurocomputing*, 168, 1094–1103.

Qin, S.J. and Dong, Y. (2018). Hybrid latent variable modeling of high dimensional time series data. *IFAC-PapersOnLine*, 51(15), 563–568.

Qin, S.J. and Zheng, Y. (2013). Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures. *AIChE Journal*, 59(2), 496–504.

Rosipal, R. and Trejo, L.J. (2001). Kernel partial least squares regression in reproducing kernel hilbert space. *Journal of machine learning research*, 2(Dec), 97–123.

Shao, W., Ge, Z., and Song, Z. (2019). Quality variable prediction for chemical processes based on semisupervised dirichlet process mixture of gaussians. *Chemical Engineering Science*, 193, 394–410.

Wang, G., Jiao, J., and Yin, S. (2017). Quality-related fault detection approaches based on data preprocessing. *IFAC-PapersOnLine*, 50(1), 15740–15747.

Wang, G., Jiao, J., and Yin, S. (2018). Efficient nonlinear fault diagnosis based on kernel sample equivalent replacement. *IEEE Transactions on Industrial Informatics*, 15(5), 2682–2690.

Yin, S., Ding, S.X., Haghani, A., Hao, H., and Zhang, P. (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(9), 1567–1581.

Yin, S., Ding, S.X., Zhang, P., Hagahni, A., and Naik, A. (2011). Study on modifications of pls approach for process monitoring. *IFAC Proceedings Volumes*, 44(1), 12389–12394.

Zhang, K., Hao, H., Chen, Z., Ding, S.X., and Peng, K. (2015a). A comparison and evaluation of key performance indicator-based multivariate statistics process monitoring approaches. *Journal of Process Control*, 33, 112–126.

Zhang, K., Peng, K., Ding, S., Chen, Z., and Yang, X. (2019). A correlation-based distributed fault detection method and its application to a hot tandem rolling mill proces. *IEEE Transactions on Industrial Electronics*.

Zhang, Y., Sun, R., and Fan, Y. (2015b). Fault diagnosis of nonlinear process based on kcpls reconstruction. *Chemometrics and Intelligent Laboratory Systems*, 140, 49–60.

Zhou, D., Li, G., and Qin, S.J. (2010). Total projection to latent structures for process monitoring. *AIChE Journal*, 56(1), 168–178.