Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 707953, 14 pages
http://dx.doi.org/10.1155/2013/707953

Research Article

Quality-Related Process Monitoring Based on Total Kernel PLS Model and Its Industrial Application

Kaixiang Peng,[1] Kai Zhang,[1] and Gang Li[2]

[1] *Key Laboratory for Advanced Control of Iron and Steel Process, School of Automation and Electrical Engineering, University of Science and Technology of Beijing, Beijing 100083, China*
[2] *Department of Automation, TNList, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Kai Zhang; ustb.kaizhang@gmail.com

Received 5 October 2013; Accepted 31 October 2013

Academic Editor: Hui Zhang

Projection to latent structures (PLS) model has been widely used in quality-related process monitoring, as it can establish a mapping relationship between process variables and quality index variables. To enhance the adaptivity of PLS, kernel PLS (KPLS) as an advanced version has been proposed for nonlinear processes. In this paper, we discuss a new total kernel PLS (T-KPLS) for nonlinear quality-related process monitoring. The new model divides the input spaces into four parts instead of two parts in KPLS, where an individual subspace is responsible in predicting quality output, and two parts are utilized for monitoring the quality-related variations. In addition, fault detection policy is developed based on the T-KPLS model, which is more well suited for nonlinear quality-related process monitoring. In the case study, a nonlinear numerical case, the typical Tennessee Eastman Process (TEP) and a real industrial hot strip mill process (HSMP) are employed to access the utility of the present scheme.

1. Introduction

Multivariate statistic process monitoring (MSPM) is effective for detecting and diagnosing abnormal operating situations in many industrial processes, which helps by improve products' quality a lot. In MSPM, projection to latent structures (PLS) model pays more attention to quality-related faults while principal component analysis (PCA) considers all faults in a process [1–7]. The major advantage of PLS is its ability to capture the relations of a large number of highly correlated process variables and few quality variables. By building a PLS model on process variables and quality variables, the process data can be projected onto two low-dimension subspaces [1, 8]. Then some statistics can be calculated in these subspaces separately. It should be noted that PLS is a linear algorithm; thus, it performs well in linear or approximately linear data. However, when the process data have strong nonlinearity, PLS will give unsatisfactory results [8].

For many physical and chemical processes, the nonlinearity lying in the process data and quality data is too obvious to be neglected. To deal with this problem, many nonlinear PLS methods have been proposed [6, 9]. Generally, PLS can be improved by two ways for nonlinear cases, which are the modification of inner model and the modification of outer model, which reflects the relation between process variables and quality variables. A method called kernel projection to latent structures (KPLS) proposed by Rosipal and Trejo is developed successfully as a nonlinear PLS model [10]. In KPLS model, the original input data are transformed into a high-dimensional space via nonlinear mapping, and then a linear PLS model is created between the feature data and quality data [11–13]. KPLS takes the advantage over other nonlinear PLS approaches as it avoids the nonlinear optimization [14, 15]. In fact, it just uses the linear algorithm of PLS in the high-dimensional feature space.

In the aforementioned literature [16, 17], Li et al. revealed the geometric properties of PLS for process monitoring and compared monitoring policies based on various PLS, which indicates that the standard PLS model divides the measured space into two oblique subspaces. One includes the quality-related variations; another subspace contains the quality-unrelated variations. Two statistics are usually

utilized for fault detection separately [3, 18]. Although PLS-based methods work well in several cases, there are still some problems. In regular PLS, there are usually many components extracted from process variables $\mathbf{X}$ for predicting quality variables $\mathbf{Y}$. As a result, the PLS model is complex to interpret [16, 19–21]. These PLS components still include variations orthogonal to $\mathbf{Y}$ which have no contribution for predicting $\mathbf{Y}$. On the other hand, the $\mathbf{X}$-residuals from PLS model are not necessarily small in covariances. This makes the use of Q statistic on $\mathbf{X}$-residuals inappropriate. The KPLS model space decomposition is similar to PLS model, with the above-mentioned defects.

In order to improve the KPLS model, a new total kernel PLS (T-KPLS) is proposed for nonlinear quality-related process monitoring in this paper. First of all, we reveled and summarized the existing KPLS model and corresponding process monitoring techniques. Then T-KPLS is developed. The properties of the new model and the process monitoring strategies are discussed then. T-KPLS model can describe the nonlinear process according to quality data effectively and also give a further decomposition on the feature spaces in KPLS. Actually, besides nonlinearity, traditional MSPM approaches also possess the assumption that the processes operate under a Gaussian distribution and in a single mode. Also, increasing number of studies can be found in this area. However, due to the scope in this paper, these issues will be considered in the subsequent researches [14, 15, 22–25].

This paper is organized as follows. KPLS-related algorithm and process monitoring methods are introduced in Section 2. Section 3 proposes the algorithm of T-KPLS, discusses its properties, and constructs T-KPLS-based process monitoring policy. Section 4 provides a numerical simulation example and TEP benchmark to illustrate the feasibility of T-KPLS-based approaches. Furthermore, the new method is also implemented to a real industrial hot strip mill process in Section 5. Finally, this paper is concluded in Section 6.

Notation. The notation adopted in this paper is fairly standard. All vectors and matrices are presented in a bold fashion and written in a vector-matrix style. The symbols for scalars and functions are regularly formulated throughout this paper.

2. KPLS Model for Process Monitoring

2.1. KPLS Model. For a nonlinear process, the input matrix can be defined as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T \in \Re^{n\times m}$, which consists of n samples with m process variables, and output matrix with p quality variables can be denoted by $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^T \in \Re^{n\times p}$. Define ϕ as a nonlinear map which maps the input vector from the original space into the feature space F, in which they are related linearly approximately. After the nonlinear map, the original input matrix $\mathbf{X}$ is changed to $\mathbf{\Phi} = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \ldots, \phi(\mathbf{x}_n)]^T \in \Re^{n\times M}$. Note that the dimensionality of the feature space M can be very large and even infinite. Define $\mathbf{K} \in \Re^{n\times n}$ as the kernel matrix to represent $\mathbf{\Phi}\mathbf{\Phi}^T$, where $\mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle$, $i, j = 1, 2, \ldots, n$, where $K(\cdot)$ is an inner product operator in feature space. With the kernel trick, one can avoid performing explicit nonlinear mapping [10]. Similar to PLS, KPLS algorithm sequentially extracts the latent vectors $\mathbf{t}$, $\mathbf{u}$ and the weight vectors $\mathbf{w}$, $\mathbf{q}$ from the $\mathbf{\Phi}$ and $\mathbf{Y}$ matrices [12]. To eliminate the mean effect, mean centering in the high-dimensional space is performed. In order to center the feature data to zero mean, the following preprocessing for normal training data is necessary [10, 12, 13]: $\mathbf{\Phi} = \mathbf{\Phi}_{\text{raw}} - \mathbf{1}_n\overline{\mathbf{\Phi}}_{\text{raw}}$, where $\mathbf{\Phi}_{\text{raw}}$ is the directly mapped matrix, $\overline{\mathbf{\Phi}}_{\text{raw}}$ denotes the mean of $\mathbf{\Phi}_{\text{raw}}$, and $\mathbf{1}_n$ represents the n-dimension column vector whose elements are all one. So the centered $\mathbf{K}$ can be calculated as follows:

$$\mathbf{K} = \left(\mathbf{I}_n - \left(\frac{1}{n}\right)\mathbf{1}_n\mathbf{1}_n^T\right)\mathbf{K}_{\text{raw}}\left(\mathbf{I}_n - \left(\frac{1}{n}\right)\mathbf{1}_n\mathbf{1}_n^T\right). \quad (1)$$

For a test sample $\mathbf{x}_{\text{new}} \in \Re^m$, the directly mapped feature vector is $\phi(\mathbf{x}_{\text{new}})_{\text{raw}} \in \Re^M$; then the inner product is calculated by $(\mathbf{K}_{\text{raw}}^{\text{new}})_i = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle = K(\mathbf{x}_i, \mathbf{x}_{\text{new}})$. The centered vector $\phi(\mathbf{x}_{\text{new}})$ is $\phi(\mathbf{x}_{\text{new}}) = \phi(\mathbf{x}_{\text{new}})_{\text{raw}} - \overline{\mathbf{\Phi}}_{\text{raw}}^T$ and $\mathbf{K}_{\text{new}}$ are mean-centered as

$$\mathbf{K}_{\text{new}} = \left(\mathbf{I}_n - \left(\frac{1}{n}\right)\mathbf{1}_n\mathbf{1}_n^T\right)\left(\mathbf{K}_{\text{raw}}^{\text{new}} - \left(\frac{1}{n}\right)\mathbf{K}_{\text{raw}}\mathbf{1}_n\right). \quad (2)$$

The algorithm of KPLS modeling has been illustrated in Appendix A. After that, $\mathbf{\Phi}$ and $\mathbf{Y}$ can be represented as

$$\begin{aligned}\mathbf{\Phi} &= \widehat{\mathbf{\Phi}} + \mathbf{\Phi}_r = \mathbf{T}\mathbf{P}^T + \mathbf{\Phi}_r,\\ \mathbf{Y} &= \widehat{\mathbf{Y}} + \mathbf{Y}_r = \mathbf{T}\mathbf{Q}^T + \mathbf{Y}_r.\end{aligned} \quad (3)$$

Let $\mathbf{R} = \mathbf{\Phi}^T\mathbf{U}(\mathbf{T}^T\mathbf{K}\mathbf{U})^{-1} \in \Re^{M\times A}$; then

$$\mathbf{T} = \mathbf{\Phi}\mathbf{R}. \quad (4)$$

The derivation of (4) is presented in Appendix B.

The determination of kernel function $K(\cdot)$ is very important. According to Mercer's theorem, there exists a mapping into a space where a kernel function acts as a dot product if the kernel function is a continuous kernel of a positive integral operator. Hence, the necessary condition for the kernel function is to meet Mercer's theorem [10, 27]. A specific choice of kernel function implicitly determines the mapping $\mathbf{\Phi}$ and the feature space F. The most widely used kernel functions include Gaussian, polynomial, sigmoid function. In this study, the Gaussian kernel function is considered

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{c}\right), \quad (5)$$

where the parameter c is the width of a Gaussian function. It plays a crucial role in process monitoring. In general, when c becomes large, the robustness of this model increases whereas the sensitivity decreases. Namely, false alarms decrease while missing alarms increase. In [28], Mika et al. proposed a method for determining c, which is widely utilized for KPLS-based nonlinear regression [29]. In this paper, first of all, we choose an appropriate false alarm rate level for normal training data (10% in this paper). Then c can be searched along with the component number A until that the KPLS model with A components acquired by cross validation presents a false alarm rates below the predefined level.

2.2. KPLS-Based Fault Detection. Usually T^2 and Q statistics are used in KPLS-based monitoring, where T^2 is for quality-related faults and Q for quality-unrelated faults. Given a new sample, the score $\mathbf{t}_{\text{new}}$ of $\phi(\mathbf{x}_{\text{new}})$ can be calculated as

$$\mathbf{t}_{\text{new}} = \mathbf{R}^T\phi\left(\mathbf{x}_{\text{new}}\right) = \left(\mathbf{U}^T\mathbf{K}\mathbf{T}\right)^{-1}\mathbf{U}^T\mathbf{K}_{\text{new}} \in \Re^A. \quad (6)$$

The residuals of $\phi(\mathbf{x}_{\text{new}})$ are represented as $\phi_r(\mathbf{x}_{\text{new}}) = \phi(\mathbf{x}_{\text{new}}) - \mathbf{P}\mathbf{t}_{\text{new}}$, which cannot be calculated directly. Further, two statistics T^2 and Q can be calculated [3, 19] as follows:

$$\begin{aligned} T^2 &= \mathbf{t}_{\text{new}}^T\Lambda^{-1}\mathbf{t}_{\text{new}}, \\ Q &= \left\|\phi_r\left(\mathbf{x}_{\text{new}}\right)\right\|^2, \end{aligned} \quad (7)$$

where $\Lambda = (1/(n-1))\mathbf{T}^T\mathbf{T}$. Two kinds of control limits are given, respectively: $(A(n^2-1)/(n(n-A)))F_{A,n-A,\alpha}$ and $g\chi_{h^2,\alpha}$. $F_{A,n-A}$ is F-distribution with A and $n-A$ degrees of freedom. $g\chi_h^2$ is the χ^2-distribution with scaling factors g and h degrees of freedom [13]. Although $\phi(\mathbf{x}_{\text{new}})$ is unavailable, it is able to calculate Q by the kernel trick as follows:

$$Q = \phi^T\left(\mathbf{x}_{\text{new}}\right)\phi\left(\mathbf{x}_{\text{new}}\right) - 2\mathbf{t}_{\text{new}}^T\mathbf{T}^T\mathbf{K}_{\text{new}} + \mathbf{t}_{\text{new}}^T\mathbf{T}^T\mathbf{K}\mathbf{T}\mathbf{t}_{\text{new}}, \quad (8)$$

where

$$\begin{aligned} &\phi^T\left(\mathbf{x}_{\text{new}}\right)\phi\left(\mathbf{x}_{\text{new}}\right) \\ &\quad = 1 - \left(\frac{2}{n}\right)\sum_{i=1}^{n}\mathbf{K}_{\text{raw}}^{\text{new}}(i) + \left(\frac{1}{n^2}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{K}_{\text{raw}}(i,j). \end{aligned} \quad (9)$$

3. T-KPLS Model for Nonlinear Data

KPLS divides the feature space F into two subspaces. One is the principal space which is monitored by T^2, reflecting the major variation related to $\mathbf{Y}$. The other is the residual space which is monitored by Q, reflecting the variation unrelated to $\mathbf{Y}$. However, the principal part $\widehat{\mathbf{\Phi}}$ contains variations which do not affect output $\mathbf{Y}$ and is useless for predicting $\mathbf{Y}$. For the residual part $\mathbf{\Phi}_r$, as the objective of KPLS is to maximize the covariance between $\mathbf{\Phi}$ and $\mathbf{Y}$, it does not extract the variance of $\mathbf{\Phi}$ in a descending order. So the latter KPLS score may capture more variance in $\mathbf{\Phi}$ than the previous one. After the score vectors have been extracted, $\mathbf{Y}$ is best predicted, but the residual of $\mathbf{\Phi}$ may still contain the large variability. Therefore, it is not suitable to use Q statistic to monitor the residual part in KPLS. In this part, a T-KPLS model is proposed to improve the original KPLS model. Following that, the T-KPLS-based process monitoring strategy is established.

3.1. T-KPLS Model. The T-KPLS model is a further decomposition on the KPLS model. It can be thought as a postprocessing method to decompose the $\widehat{\mathbf{\Phi}}$ and $\mathbf{\Phi}_r$ further in KPLS. The detailed algorithm for T-KPLS can be found in Algorithm 1.

In step (4) of Algorithm 1, loading matrix $\mathbf{P}_o = \mathbf{\Phi}_o^T\mathbf{W}_o \in \Re^{M\times A_o}$, where $\mathbf{W}_o \in \Re^{n\times A_o}$ contains the scaled eigenvectors of $(1/n)\mathbf{\Phi}_o\mathbf{\Phi}_o^T$ corresponding to its A_o largest eigenvalues. In step (5), $\mathbf{P}_r = \mathbf{\Phi}_r^T\mathbf{W}_r \in \Re^{M\times A_r}$, where $\mathbf{W}_r \in \Re^{n\times A_r}$ are the scaled eigenvectors of $(1/n)\mathbf{\Phi}_r\mathbf{\Phi}_r^T$ corresponding to its A_r largest eigenvalues [27]. As $\phi(\cdot)$ is unknown, the algorithm in Algorithm 1 cannot be implemented intuitively, while the calculable steps are shown in Algorithm 2. In Algorithm 2,

$$\begin{aligned} \mathbf{K}_o &= Z_y\mathbf{T}\mathbf{T}^T\mathbf{K}\mathbf{T}\mathbf{T}^TZ_y, \\ \mathbf{K}_r &= \left(\mathbf{I}_n - \mathbf{T}\mathbf{T}^T\right)\mathbf{K}\left(\mathbf{I}_n - \mathbf{T}\mathbf{T}^T\right), \end{aligned} \quad (10)$$

where $Z_y = \mathbf{I}_n - \mathbf{T}_y(\mathbf{T}_y^T\mathbf{T}_y)^{-1}\mathbf{T}_y$.

In T-KPLS model, we can model $\mathbf{\Phi}$ and $\mathbf{Y}$ as follows:

$$\begin{aligned} \mathbf{\Phi} &= \mathbf{\Phi}_y + \mathbf{\Phi}_o + \mathbf{\Phi}_{rp} + \mathbf{\Phi}_{rr}, \\ \mathbf{Y} &= \mathbf{T}_y\mathbf{Q}_y^T + \mathbf{Y}_r. \end{aligned} \quad (11)$$

The meanings of different sections of $\mathbf{\Phi}$ are listed in Table 1. Compared with KPLS, T-KPLS is clearer for describing $\mathbf{\Phi}$ and more suitable for monitoring different parts of $\phi(\mathbf{x})$. T-KPLS does not change the prediction ability of $\mathbf{Y}$, but it decomposes $\mathbf{\Phi}$ thoroughly supervised by $\mathbf{Y}$. $\mathbf{T}_y$ is the score of $\mathbf{\Phi}_y$ and completely related to $\mathbf{Y}$ from the original $\mathbf{T}$, whereas $\mathbf{T}_o$ is the score of $\mathbf{\Phi}_o$ and orthogonal to $\mathbf{Y}$ in original $\mathbf{T}$. $\mathbf{T}_r$ is the main part of $\mathbf{\Phi}_r$. $\mathbf{\Phi}_{rr}$ represents the residual of $\mathbf{\Phi}$ and the noise. Note that in the T-KPLS model, all the scores $\mathbf{T}_y$, $\mathbf{T}_o$, and $\mathbf{T}_r$ have their definite values. However, the loadings $\mathbf{P}_y$, $\mathbf{P}_o$, and $\mathbf{P}_r$ are unknown because of the uncertain map function ϕ.

Table 1: Meaning of different sections of $\mathbf{\Phi}$.

Section	Description
$\mathbf{\Phi}_y$	The $\mathbf{Y}$-related part of $\widehat{\mathbf{\Phi}}$ which is responsible for predicting $\mathbf{Y}$
$\mathbf{\Phi}_o$	The part of $\widehat{\mathbf{\Phi}}$ that is orthogonal to $\mathbf{Y}$ in original $\mathbf{T}$ of KPLS
$\mathbf{\Phi}_{rp}$	The principal part of $\mathbf{\Phi}_r$ which represents a large variation in $\mathbf{\Phi}_r$
$\mathbf{\Phi}_{rr}$	The residual part which is not excited in $\mathbf{\Phi}$

In T-KPLS, the orthogonality among all score vectors holds. Meanwhile, $\mathbf{T}_o$ is orthogonal to output $\mathbf{Y}$. The proof is omitted, and one can refer to Zhou et al. [19].

3.2. T-KPLS-Based Quality-Related Process Monitoring. In multivariate statistical process monitoring, two types of statistics are widely used for fault detection. One is the D statistic which calculates the Mahalanobis distance between new scores and the normal scores. The other is the Q statistic which represents the square predict error of the sample. As for T-KPLS, the similar statistics are constructed. After T-KPLS model is built from normal historical data, the new scores and residuals are calculated from the new sample. Then, the statistics are constructed with corresponding control limits for fault detection.

(1) Perform KPLS algorithm on $\mathbf{X}$ and $\mathbf{Y}$ to get the model described in (3)
(2) Run PCA on $\widehat{\mathbf{Y}}$ with A_y components, where $\widehat{\mathbf{Y}} = \mathbf{T}_y\mathbf{Q}_y^T$, $A_y = \text{rank}(\mathbf{Q})$
(3) Define $\mathbf{\Phi}_y = \mathbf{T}_y\mathbf{P}_y^T$, where $\mathbf{P}_y^T = \left(\mathbf{T}_y^T\mathbf{T}_y\right)^{-1}\mathbf{T}_y^T\widehat{\mathbf{\Phi}}$, $\mathbf{P}_y \in \Re^{M\times A_y}$
(4) Run PCA on $\mathbf{\Phi}_o = \widehat{\mathbf{\Phi}} - \mathbf{\Phi}_y$, with A_o components, $A_o = A - A_y$, $\mathbf{\Phi}_o = \mathbf{T}_o\mathbf{P}_o^T$
(5) Perform PCA on $\mathbf{\Phi}_r$, with A_r components, where A_r is determined using PCA methods, $\mathbf{\Phi}_{rp} = \mathbf{T}_r\mathbf{P}_r^T$
(6) $\mathbf{\Phi}_{rr} = \mathbf{\Phi}_r - \mathbf{\Phi}_{rp} = \mathbf{\Phi}_r - \mathbf{T}_r\mathbf{P}_r^T$

Algorithm 1: T-KPLS algorithm for comprehension.

Obtain $\mathbf{K}$ and $\mathbf{Y}$

(1) After KPLS model: $\mathbf{T} = \mathbf{KU}(\mathbf{T}^T\mathbf{KU})^{-1}$
(2) Run eigenvector decomposition on $\widehat{\mathbf{Y}}$: $\mathbf{T}_y = \widehat{\mathbf{Y}}\mathbf{Q}_y = \mathbf{T}\mathbf{Q}^T\mathbf{Q}_y$
(3) Perform eigenvector decomposition on $(1/n)\,\mathbf{K}_o$ to get the eigenvectors $\mathbf{W}_o$ with regard to its largest A_o eigenvalues. $\mathbf{T}_o = \mathbf{K}_o\mathbf{W}_o$
(4) Perform eigenvector decomposition on $(1/n)\,\mathbf{K}_r$ to get the eigenvectors $\mathbf{W}_r$ with regard to its largest A_r eigenvalues. $\mathbf{T}_r = \mathbf{K}_r\mathbf{W}_r$

Algorithm 2: T-KPLS algorithm for calculation.

Table 2: Monitoring statistics and control limits.

Statistic	Calculation	Control limit
T_y^2	$\mathbf{t}_{y\text{new}}^T\mathbf{\Lambda}_y^{-1}\mathbf{t}_{y\text{new}}$	$\frac{A_y(n^2-1)}{n(n-A_y)}F_{A_y,n-A_y,\alpha}$
T_o^2	$\mathbf{t}_{o\text{new}}^T\mathbf{\Lambda}_o^{-1}\mathbf{t}_{o\text{new}}$	$\frac{A_o(n^2-1)}{n(n-A_o)}F_{A_o,n-A_o,\alpha}$
T_r^2	$\mathbf{t}_{r\text{new}}^T\mathbf{\Lambda}_r^{-1}\mathbf{t}_{r\text{new}}$	$\frac{A_r(n^2-1)}{n(n-A_r)}F_{A_r,n-A_r,\alpha}$
Q_r	$\left\|\phi_{rr}(\mathbf{x}_{\text{new}})\right\|^2$	$g\chi_{h,a}^2$

According to T-KPLS model, three score vectors can be calculated as follows:

$$\mathbf{t}_{y\text{new}} = \Theta_y\mathbf{K}_{\text{new}} \in \Re^{A_y},$$
$$\mathbf{t}_{o\text{new}} = \Theta_o\mathbf{K}_{\text{new}} \in \Re^{A_o}, \quad (12)$$
$$\mathbf{t}_{r\text{new}} = \Theta_r\mathbf{K}_{\text{new}} \in \Re^{A_r}.$$

Motivated by total PLS- (T-PLS-) based methods [19], four fault detection indices are constructed in Table 2. The expression of Q_r can be calculated as follows:

$$Q_r = \phi^T(\mathbf{x}_{\text{new}})\,\phi(\mathbf{x}_{\text{new}}) - \mathbf{K}_{\text{new}}^T\Omega_r\mathbf{K}_{\text{new}}. \quad (13)$$

The detailed expression of (12) and Q_r for calculation are shown in Appendix C.

3.3. Model Implementation. Implementation of the T-KPLS-based quality-related detection scheme involves offline training model and online testing model. As sketched in Figure 1, the training model aims to obtain the model parameters. When all parameters are available, the schematic plot for a testing sample is sketched in Figure 2. The whole procedure involves four steps: the acquisition of online measurement, the calculation of all scores for the new sample, the acquirement of four detection indices, and the result for quality-related detection.

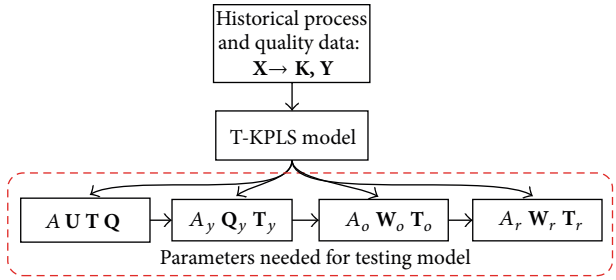

Figure 1: Training model T-KPLS-based monitoring.

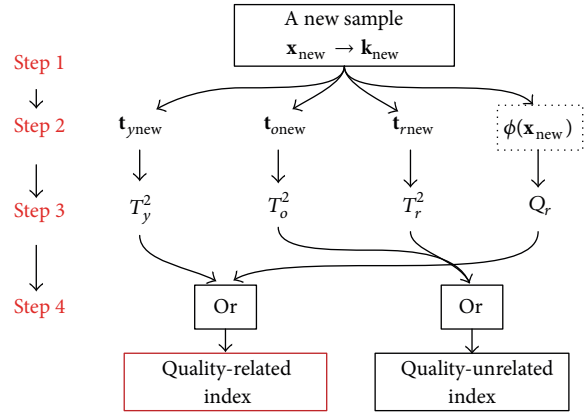

Figure 2: Flowchart of testing model for T-KPLS-based monitoring.

4. Case Study on Simulation Examples

In this section, two detailed simulation examples are carried out to demonstrate the advantage of T-KPLS.

4.1. Simulation on a Numerical Nonlinear Example. Firstly, a synthetic nonlinear numerical process without feedback is presented as follows:

$$\text{Process variable}: \begin{cases} \mathbf{x}_1 \sim \mathbf{N}(0,1), \ \mathbf{x}_2 \sim \mathbf{N}(0,1), \\ \mathbf{x}_3 = \sin(\mathbf{x}_1) + e_1, \\ \mathbf{x}_4 = \mathbf{x}_1^2 - 3\mathbf{x}_1 + 4 + e_2, \\ \mathbf{x}_5 = \mathbf{x}_2^2 + \cos(\mathbf{x}_2^2) + 1 + e_3, \end{cases} \tag{14}$$

$$\text{Quality variable}: \mathbf{y} = \mathbf{x}_3^2 + \mathbf{x}_3\mathbf{x}_4 + \mathbf{x}_1 + \nu,$$

where $e_i \sim \mathbf{N}(0, 0.01^2)$ $(i = 1, 2, 3)$, $\nu \sim \mathbf{N}(0, 0.05^2)$, $\mathbf{N}(\mu, \sigma^2)$ means the normal distribution with mean μ and variance σ^2. From (14), it is obvious that the abnormal variation in $\mathbf{x}_1$ can cause the disturbances in $\mathbf{x}_3$ and $\mathbf{x}_4$, while $\mathbf{x}_2$ just influences $\mathbf{x}_5$. As quality variable $\mathbf{y}$ merely relates to $\mathbf{x}_1$, $\mathbf{x}_3$, and $\mathbf{x}_4$, so the fault in $\mathbf{x}_1$ will affect $\mathbf{y}$, while the fault in $\mathbf{x}_2$ cannot.

We used 200 samples generated from the above process as a training dataset. The faulty dataset with 400 samples was also generated according to the following faults:

(i) Fault 1: a step bias in $\mathbf{x}_2$ at 201st sample, $\mathbf{x}_2 = \mathbf{x}_2^* + f$,

(ii) Fault 2: a ramp change in $\mathbf{x}_2$ at 201st sample, $\mathbf{x}_2 = \mathbf{x}_2^* + (k - 200)f$,

(iii) Fault 3: a step bias in $\mathbf{x}_1$ at 201st sample, $\mathbf{x}_1 = \mathbf{x}_1^* + f$,

(iv) Fault 4: a ramp change in $\mathbf{x}_1$ at 201st sample, $\mathbf{x}_1 = \mathbf{x}_1^* + (k - 200)f$,

where $\mathbf{x}_1^*$, $\mathbf{x}_2^*$ are the normal values of $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, f is the magnitude for step bias and slope for ramp change, and k is the sample number. Then the faulty measurements of variable $\mathbf{x}_3$, $\mathbf{x}_4$, and $\mathbf{x}_5$ are generated by (14).

Training samples are applied to perform a KPLS model on $(\mathbf{X}, \mathbf{y})$. The width of Gaussian kernel $c = 100$ is kept for this simulation. The components number $\mathbf{A} = 2$ is determined using cross validation, which provides a good prediction of $\mathbf{y}$. Then T-KPLS model is constructed based on KPLS, where $A_y = 1$ for the single output, and $A_r = 1$ is chosen as the principal component unrelated to $\mathbf{y}$.

According to the descriptions of Faults 1 and 2, they are quality-unrelated faults. Let $f = 1$; the monitoring results with KPLS model (T^2 and Q) are plotted in Figure 3. It is observed that Fault 1 causes significant alarms in both two detection indices of KPLS. However, the alarms in T^2 chart are false alarms for indicating a $\mathbf{y}$-related fault. Thus, KPLS-based monitoring causes false alarms for this disturbance. T-KPLS-based monitoring for Fault 1 is depicted in Figure 4. Among the four detection indices, T_y^2 is kept under the control line, which gives correct result. Also T_r^2 and Q_r alarm tinily. Compared with KPLS, T-KPLS provides lower false alarm rates for Fault 1. Similarly, the detection results of Fault 2 with $f = 0.005$ using KPLS and T-KPLS are shown in Figures 5 and 6, respectively. It is shown that the results for Fault 2 is similar to that of Fault 1. Table 3 lists the false alarm rates under different fault magnitudes f. In all simulations, we repeat 100 times and make use of the mean for conviction. From Table 3, it is clear that T-KPLS-based method gives lower false alarm rates.

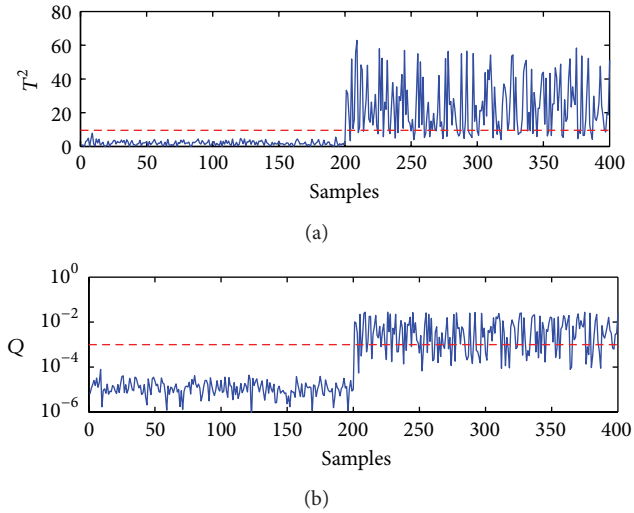

Figure 3: KPLS-based monitoring with 99% control limit when quality-unrelated Fault 1 occurs.

Table 3: False alarm rates of faults unrelated to $\mathbf{y}$ (%).

	Fault value (f)	KPLS (T^2)	T-KPLS (T_y^2)	T-KPLS (Q_r)	T-KPLS (T_y^2 or Q_r)
Fault 1	0.2	26.8	0	4.7	4.7
	0.4	31.7	0	11.6	11.6
	0.6	53.3	0	26.6	26.6
	0.8	77.2	0	43.3	43.3
Fault 2	0.002	24.5	0	8.3	8.3
	0.003	37.4	0	18.6	18.6
	0.004	44.5	0	27.8	27.8
	0.005	56.3	0	36.8	36.8

The predefined Faults 3 and 4 are quality-related. For Fault 3 with $f = 0.6$, KPLS-based method could detect this fault as shown in Figure 7. T-KPLS-based method performs sensitively in T_y^2, T_r^2, and Q_r in Figure 8. That is to say, the alarms in T^2 of KPLS are merely denoted by T_y^2 of T-KPLS. Thus, for this kind of fault, when the step magnitude is small enough, T-KPLS will work better than KPLS. For quality-related Fault 4 with $f = 0.005$, KPLS-based method cannot detect quality-related faults by T^2 as shown in Figure 9, while T-KPLS-based Q_r statistic detects the fault sensitively in Figure 10. It means that the variations leading $\mathbf{y}$ to abnormality occur in the residual space. The results of simulation on Faults 3 and 4 show that T-KPLS-based policy could improve the detection rates. Moreover, Table 4 lists the detection results which show that the quality-related fault can be detected by T-KPLS using T_y^2 and Q_r better.

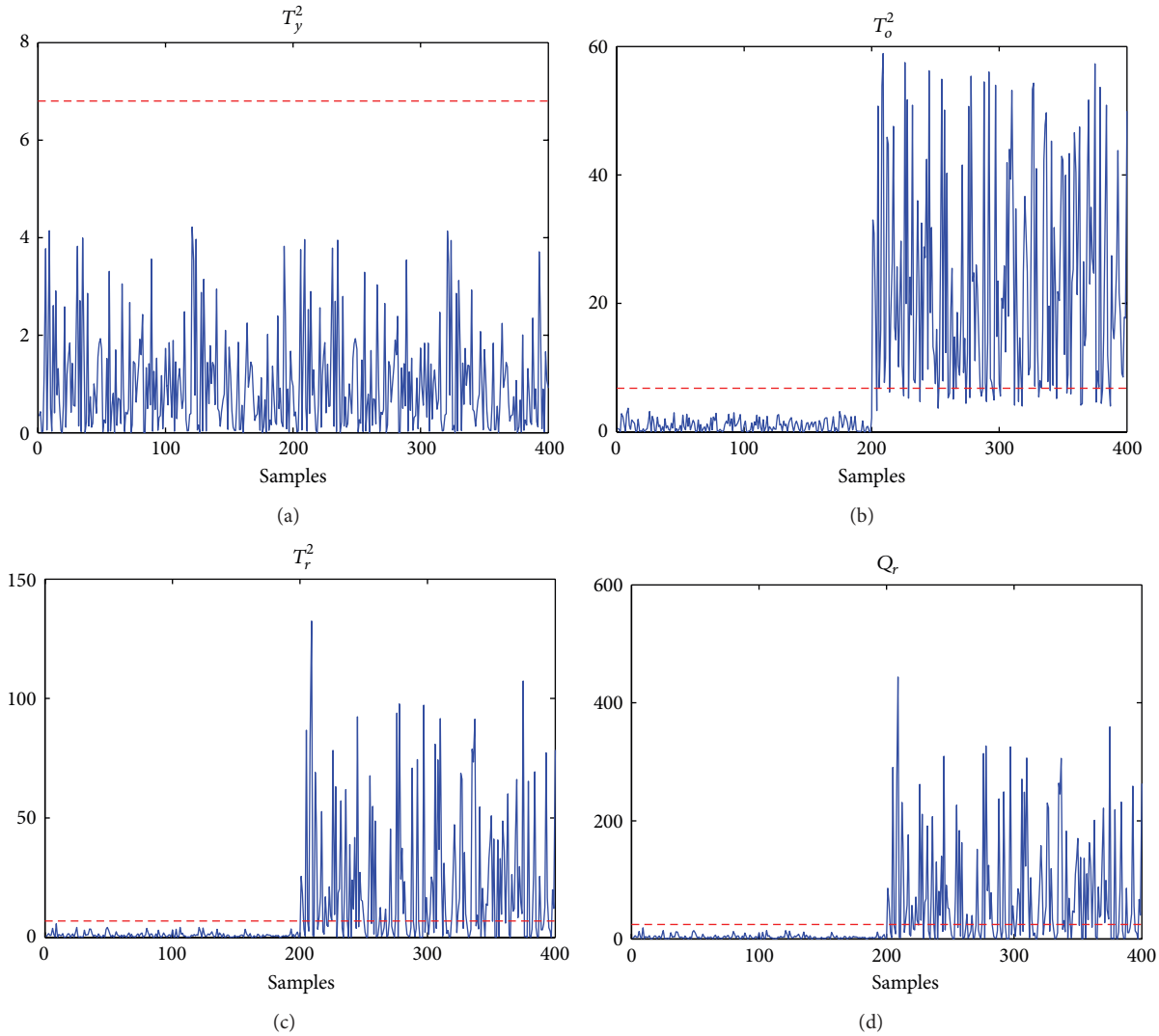

Figure 4: T-KPLS-based monitoring when quality-unrelated Fault 1 occurs.

Table 4: False detection rates of faults related to $\mathbf{y}$ (%).

	Fault value (f)	KPLS (T^2)	T-KPLS (T_y^2)	T-KPLS (Q_r)	T-KPLS (T_y^2 or Q_r)
Fault 3	0.2	63.6	80.5	57.3	83.6
	0.4	79.8	88.5	74.4	88.7
	0.6	90.3	99.2	86.3	99.2
	0.8	99.4	100	99.5	100
Fault 4	0.002	4.1	0	29.5	29.5
	0.003	4.3	0	43.1	43.1
	0.004	3.5	0	54.2	54.2
	0.005	4.6	0	62.1	62.1

4.2. Simulation on Tennessee Eastman Process

4.2.1. Tennessee Eastman Process. The Tennessee Eastman (TE) Process was provided by Eastman Chemical Company which is a realistic industrial process for evaluating different process control and monitoring technologies [16, 30]. The process has five major parts: a reactor, condenser, recycle compressor, liquid separator, and product stripper, and it involves eight components: A–H. The gaseous reactants A, C–E, and the inert B are fed to the reactor while the liquid products G and H are formed. The reactions in the reactor follow (15). The species F is a by-product. All reactions of this process are irreversible, exothermic, and approximately one-order with respect to the reactant concentrations. For detailed process description, one can refer to Lee et al. and Chiang et al. [30, 31]. The process used here is implemented under closed-loop control. All the training and testing datasets were generated by Chiang et al. and Lee et al., which can be openly downloaded in their website. The faults in the test dataset are introduced from the 160th sample. The TE process has been used as a benchmark process for evaluating process monitoring methods. Kano et al. applied PCA-based method for monitoring this process [32]. Russell et al. compared canonical vector analysis (CVA) and PCA-based technologies, while Lee et al. reviewed the results using both

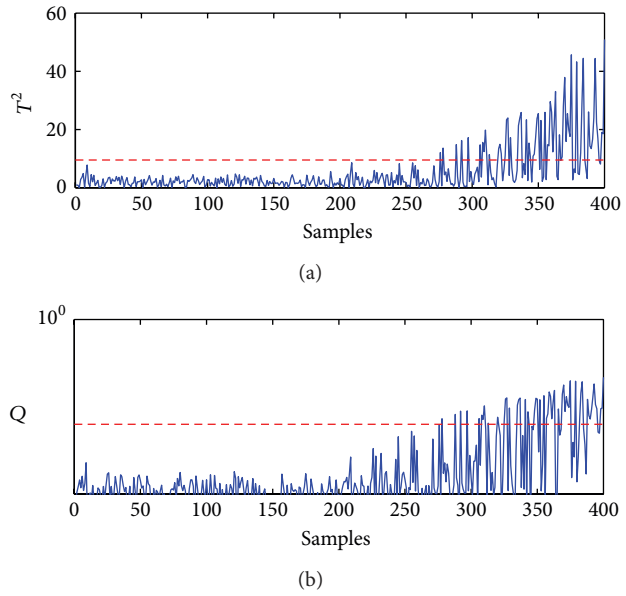

Figure 5: KPLS-based monitoring with 99% control limit when quality-unrelated Fault 2 occurs.

independent component analysis (ICA) and PCA for TEP [26, 30, 33]. Also, PLS-based monitoring policy has been utilized for quality-related fault detection [30]. In [31], Chiang et al. compared the fault detection and diagnosis method such as PCA, PLS, and Fisher discriminant analysis (FDA), according to the case study of TEP.

The TEP contains two blocks of variables: 12 manipulated variables and 41 measured variables. Process measurements are sampled with interval of 3 min, while nineteen composition measurements are sampled with time delays which vary from 6 min to 15 min. The time delay has a potentially critical impact on product quality control in this process, because the closed-loop control works when the next sample of quality variable is available [21]. Thus during this interval, the products are produced with uncontrolled quality. It also implies that the fault effect on product quality cannot be detected until next measurement sampled,

$$\begin{aligned} \mathrm{A}(\mathbf{g}) + \mathrm{C}(\mathbf{g}) + \mathrm{D}(\mathbf{g}) &\longrightarrow \mathrm{G}(\mathrm{liq}), \\ \mathrm{A}(\mathbf{g}) + \mathrm{C}(\mathbf{g}) + \mathrm{E}(\mathbf{g}) &\longrightarrow \mathrm{H}(\mathrm{liq}), \\ \mathrm{A}(\mathbf{g}) + \mathrm{E}(\mathbf{g}) &\longrightarrow \mathrm{F}(\mathrm{liq}), \\ 3\mathrm{D}(\mathbf{g}) &\longrightarrow 2\mathrm{F}(\mathrm{liq}); \end{aligned} \tag{15}$$

PLS and KPLS-based monitoring methods can detect the fault correlated to $\mathbf{Y}$, thus receiving wide applications in industrial cases. There are 21 predefined faults in TEP, in which 15 of them are known, denoted by IDV (1–15). IDV (1–7) are step changes in a process variable, for example, in the cooling water inlet temperature. IDV (8–12) are associated with an increase in the variability of some process variables. Fault 13 is a slow drift in the reaction kinetics. IDV (14–15) are associated with sticking valves [19, 20].

Table 5: Fault detection rate of TEP using T-PLS, KPLS, and T-KPLS (%).

Faults ID	Type	T-PLS	KPLS	T-KPLS
IDV(1)	Step	99.3	88.6	**99.7**
IDV(2)	Step	97.6	98.6	**99.6**
IDV(5)	Step	**99.5**	48.2	97.4
IDV(6)	Step	**99.8**	99.5	**99.8**
IDV(8)	Random variation	93.4	95.6	**97.3**
IDV(12)	Random variation	95.6	**99.8**	98.3
IDV(13)	Slow drift	95.3	96.4	**98.5**

Table 6: False alarm rates of TEP using T-PLS, KPLS, and T-KPLS (%).

Faults ID	Type	T-PLS	KPLS	T-KPLS
IDV(0)	—	**5.2**	8.6	5.9
IDV(3)	Step	**5.9**	9.8	**5.9**
IDV(4)	Step	33.5	25.3	**17.2**
IDV(9)	Step	5.3	8.2	**4.4**
IDV(11)	Random variation	32.3	32.7	**17.8**
IDV(14)	Random variation	12.4	22.7	**7.8**
IDV(15)	Slow drift	**5.3**	28.0	10.0

4.2.2. T-KPLS-Based Quality-Related Detection for TEP. In this case study, the component G in steam 9, that is, the 35th measured variable, is chosen as the output quality variable $\mathbf{y}$. The process variables $\mathbf{X}$ consist of measured variable 1–22 and manipulated variable 1–11. The detailed $\mathbf{X}$ and $\mathbf{y}$ are summarized by Li et al. [20]. We use 480 normal samples to build KPLS and T-KPLS model. The selection of kernel parameter c affects the detection results for this process significantly. According to the simulation results, the larger c is, the lower the false alarm rates and the higher the missing alarm rates will be. In this simulation, $c = 5000$ is chosen for the KPLS model. Eight principal components are kept according to cross validation. For T-KPLS, A_y is set to 1 because of the single quality variable, and $A_o = A - A_y = 7$, $A_r = 6$ are determined according to the KPCA-based method. TEP provides 21 faulty sample datasets, and each of them consists of 960 samples. Here, we apply 13 known fault sets to perform our simulation. First of all, these known faults should be divided into two groups including the quality-related faults and the quality-unrelated faults with the criteria proposed by Zhou et al. [19]. Here, the IDV (1, 2, 5, 6, 8, 12, 13) are related to quality variable $\mathbf{y}$; others are not. For comparison, the normal data set is also included in this simulation. As illustrated in Figures 11 and 12, the proposed approach with T_y^2 and Q_r can detect Fault 1 effectively, but show few false alarms for quality-unrelated Fault 3. The alarm for the quality-related fault is considered as an effective alarm, while the detection for quality-unrelated fault is thought to be a false alarm. Tables 5 and 6 list the fault detection rates and fault alarm rates of KPLS and T-KPLS. Also, the detection results by T-PLS [19] are cited in these two tables for comparison.

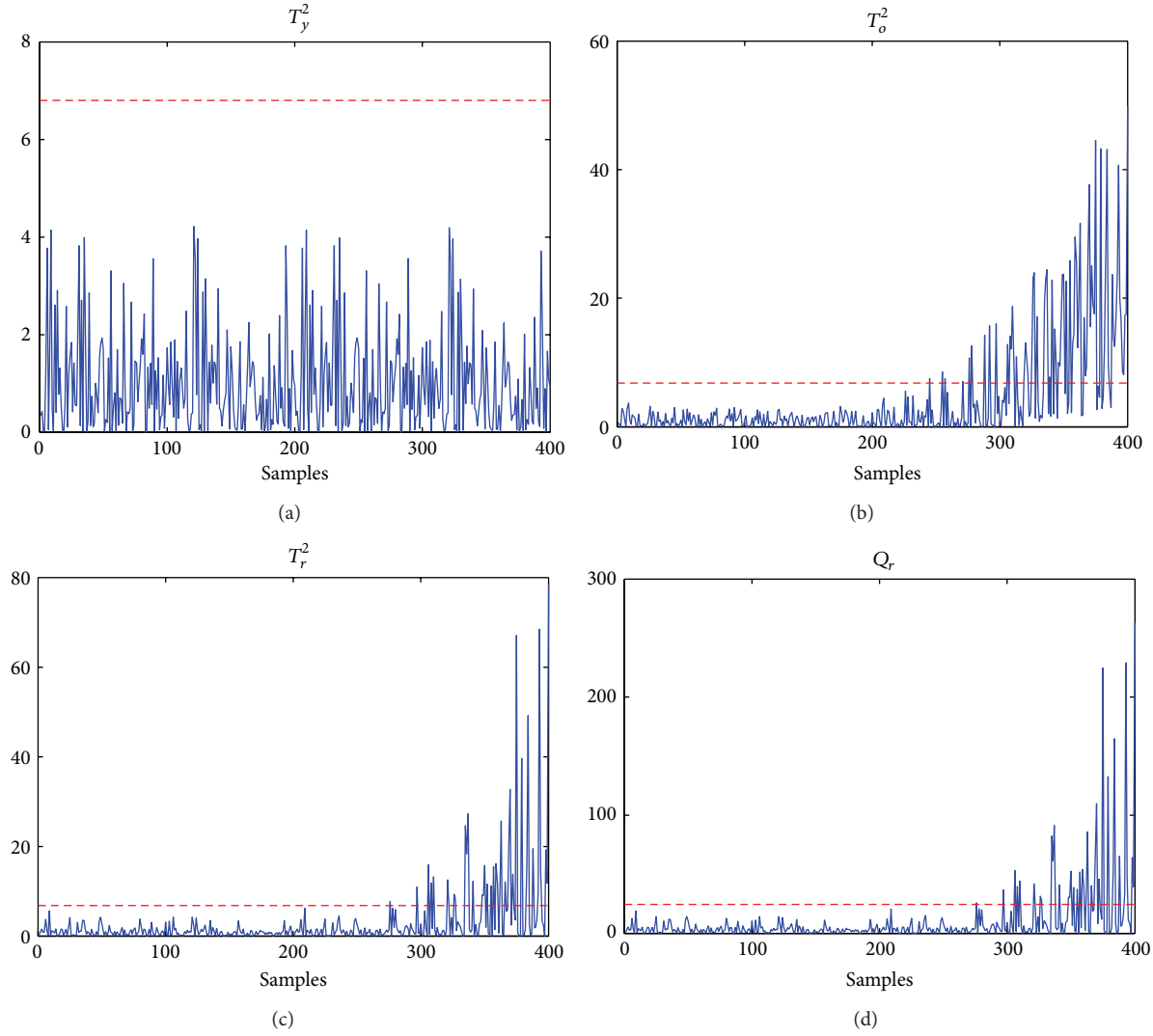

Figure 6: T-KPLS-based monitoring when quality-unrelated Fault 2 occurs.

From the detection results, it is observed that T-KPLS-based method gives a higher detection rate and lower false alarm rate than KPLS-based method. Compared with linear T-PLS, T-KPLS performs better in most cases. In Table 5, T-KPLS has higher detection rates in most cases. Meanwhile, T-KPLS gives lower false alarm rates in most cases as shown in Table 6. To sum up, T-KPLS is an improvement for KPLS, and it is effective to detect quality-related faults in nonlinear processes.

5. Application in Real Industrial Hot Strip Mill

Hot strip mill process (HSMP) is an extremely complex process in iron and steel industry. A schematic layout of the hot strip mill is illustrated in Figure 13 corresponding to the real industrial hot strip mill. According to Figure 13, the process generally consists of the following units: reheating furnaces, roughing mill, transfer table, crop shear, finishing mill, run-out table cooling, and coiler. The finishing mill has the most significant influence on the final thickness of steel strip, in which the controlled variables include average gap of the 7 finishing mill stands and work roll bending (WRB) force of the last 6 stands (WRB force of the first stand is not measured). The thickness and temperature of the strip after finishing rolling are around 850°C–950°C and 1.5–12.7 mm, respectively. As is well known from materials science, the kinetics of metallurgical transformations and the flow stress of the rolled steel strip are dominantly controlled by the temperature, which is mainly determined by the finishing temperature control (FTC).

The demand of dimensional precision, especially thickness precision of hot strip mill, has become stricter in recent years, which makes the improvement of thickness precision be a hot topic. In general, the thickness in exit of finishing mill is closely related to gap and rolling force and has little connection with bending force. In this paper, two classes of strips' manufacturing process are taken for this test with thicknesses, where their thickness targets are 3.95 mm and

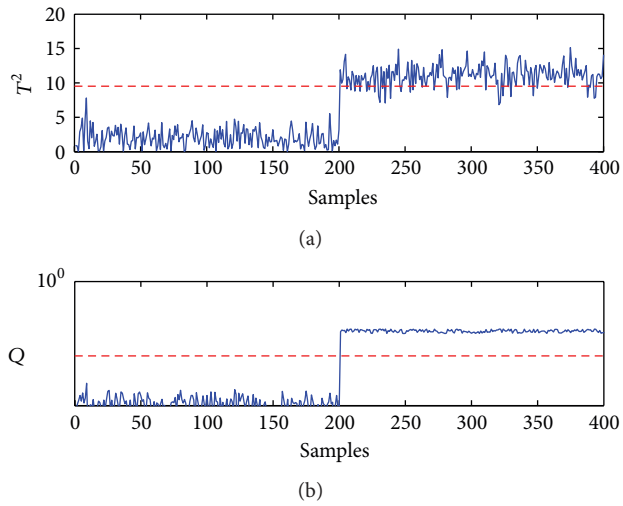

Figure 7: KPLS-based monitoring with 99% control limit when quality-related Fault 3 occurs.

Table 7: Process and quality variables in finishing mill.

Variable	Type	Description	Unit
1~7	Measured	F_i stand average gap, $i = 1, ..., 7$	mm
8~14	Measured	F_i stand total force, $i = 1, ..., 7$	MN
15~20	Measured	F_i stand work roll bending force, $i = 2, ..., 7$	MN
y	Quality	Finishing mill exit strip thickness	mm

2.70 mm, respectively. Based on historical dataset, the new proposed framework can be constructed with the measured process variables and quality variable which are listed in Table 7. In this case study, three kinds of frequently occurring faults are mainly studied, which are listed in Table 8, where all faults with the same duration time of 10 s are terminated artificially. In real circumstances, faults may occur in some driving units or sensors for measuring force, temperature, and gaps. Furthermore, malfunction of control loop in a single stand may also exist occasionally. To be summarized, three kinds of faults defined in control systems can all be found in finishing mill process. In this work, three typical faults separately selected from each type are chosen to support our study, which are tabulated in Table 8. Among all these faults, Fault 1 is a little quality-related; others are directly quality-related. Gaussian kernel parameter c affects detection results significantly. In this study, T-KPLS model is built, where $A = 8$ is determined according to cross validation, $A_y = 1$; because of the single output, $A_r = 10$ is obtained by KPCA-based method. In the model, $c_{\min} = 0$ and $c_{\max} = 10000$ are chosen, which yield an optimum $c = 7500$.

The results of thickness quality-related process monitoring are given by Table 9. As can be shown in Table 9, compared with PLS, KPLS, and T-PLS, T-KPLS-based method

Table 8: Typical faults in finishing mill.

No.	Description	Fault type	Quality related
1	Sensor fault of bending force measurement in F_5 stand	Sensor fault	No
2	Malfunction of hydraulic gap control loop in F_4 stand	Process fault	Yes
3	Actuator fault of cooling valve between F_2 and F_3 stands	Actuator fault	Yes

Table 9: Detection rate or false alarm rate for hot strip mill (%).

Fault No.	Type of detection	PLS (T^2)	KPLS (T^2)	T-PLS (T^2 or Q_r)	T-KPLS (T^2 or Q_r)
1	False alarm rate	0.104	0.117	0.366	**0.044**
2	Detection rate	0.998	**1.000**	**1.000**	**1.000**
3	Detection rate	0.656	0.870	0.900	**0.980**

just gives a little false alarm rate for quality-unrelated Fault 1, while for quality-related Fault 2 and 3, it presents higher detection rates, especially in Fault 3. In conclusion, T-KPLS is an appropriate enhancement for typical KPLS model, and it is effective to deal with the quality-related disturbances in real industrial processes.

Regarding HSMP, the following should be noted.

Remark 1. We clarify that the data considered about finishing mill process are acquired from real steel industrial field, namely, Ansteel Corporation, China. The faults occur occasionally and were eliminated manually.

Remark 2. In this implementation, only thickness has been concerned as the quality variable, whereas T-KPLS model can handle multioutput cases.

6. Conclusion

In this paper, the T-KPLS algorithm is proposed by further decomposing KPLS. The purpose of T-KPLS is to perform a further decomposition on the high dimension space induced by KPLS, which is more suitable for quality-related process monitoring. The process monitoring methods based on T-KPLS are developed to monitor the operating performance. Both theoretical analysis and simulation results show better performance of T-KPLS than KPLS. T-KPLS-based methods can give lower false alarm rates and missing alarm rates than KPLS-based methods in most simulated cases. However, there are still some problems needed to be considered in the modeling with T-KPLS, such as how to select an appropriate kernel function for a given process data and establish a framework for precisely choosing the kernel parameters. Due to the scope of this paper, further studies for these issues will be concerned in the future.

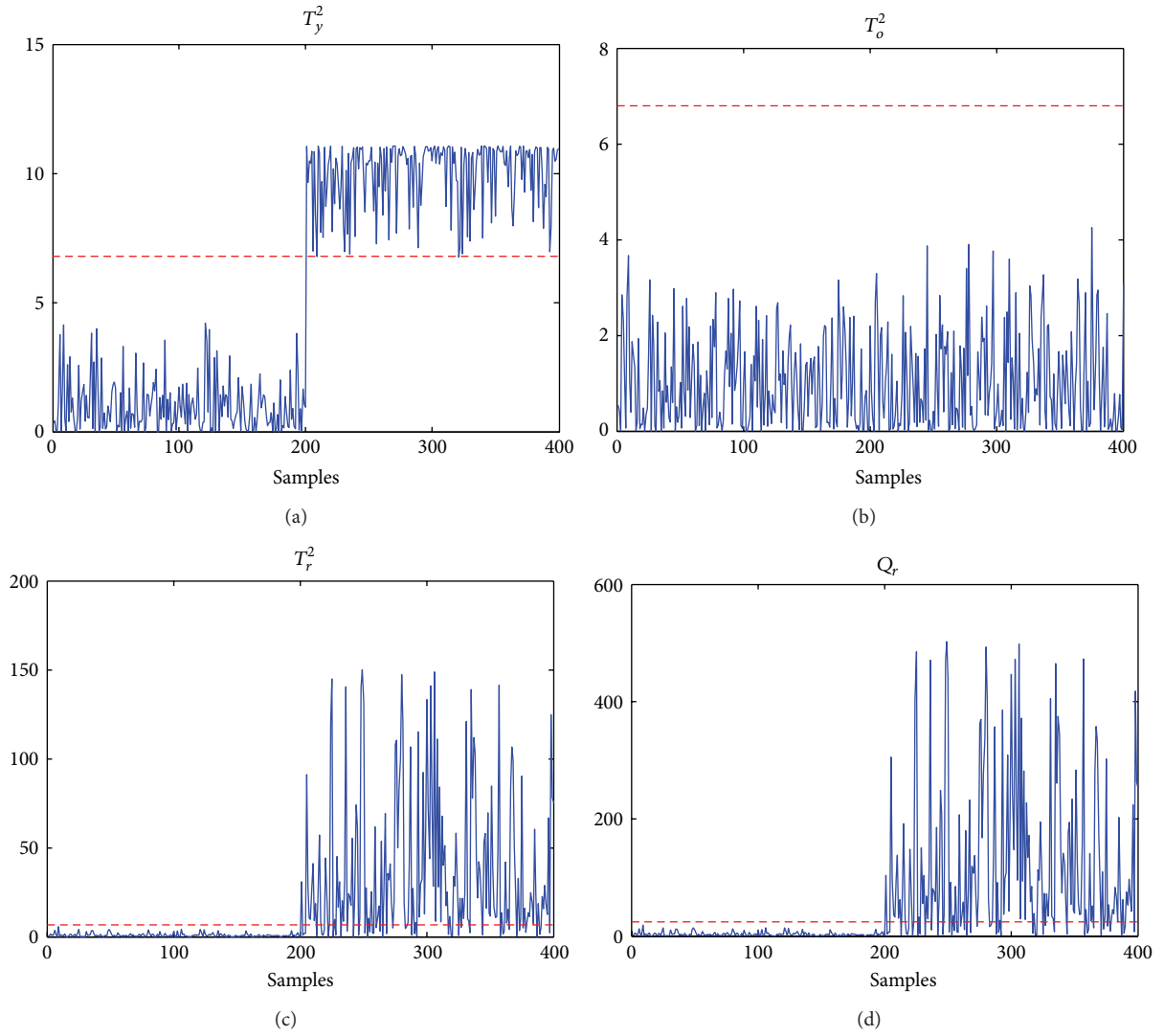

Figure 8: T-KPLS-based monitoring when quality-related Fault 3 occurs.

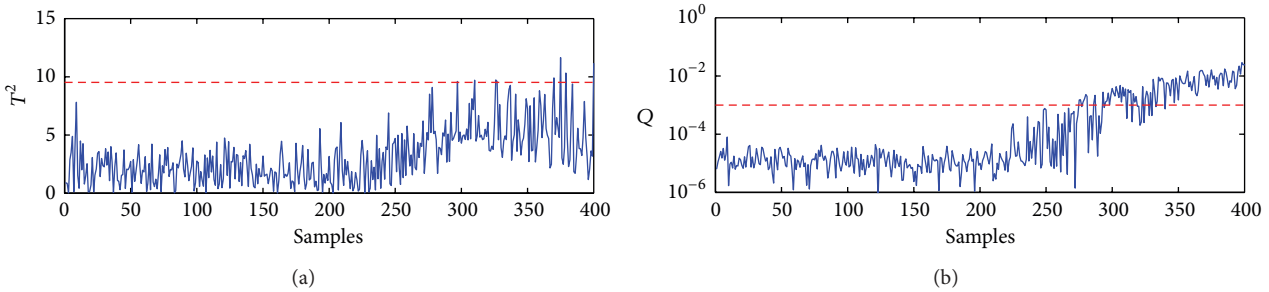

Figure 9: KPLS-based monitoring with 99% control limit when quality-related Fault 4 occurs.

Appendices

A. KPLS Algorithm

The nonlinear iterative KPLS algorithm is shown in Algorithm 3.

Based on Algorithm 3, the following equations hold:

$$\begin{aligned} \mathbf{P} &= \mathbf{\Phi}^T\mathbf{T}, \qquad \mathbf{Q} = \mathbf{Y}^T\mathbf{T}, \\ \mathbf{\Phi}_r &= \left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right)\mathbf{\Phi}, \\ \mathbf{Y}_r &= \left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right)\mathbf{Y}, \end{aligned} \tag{A.1}$$

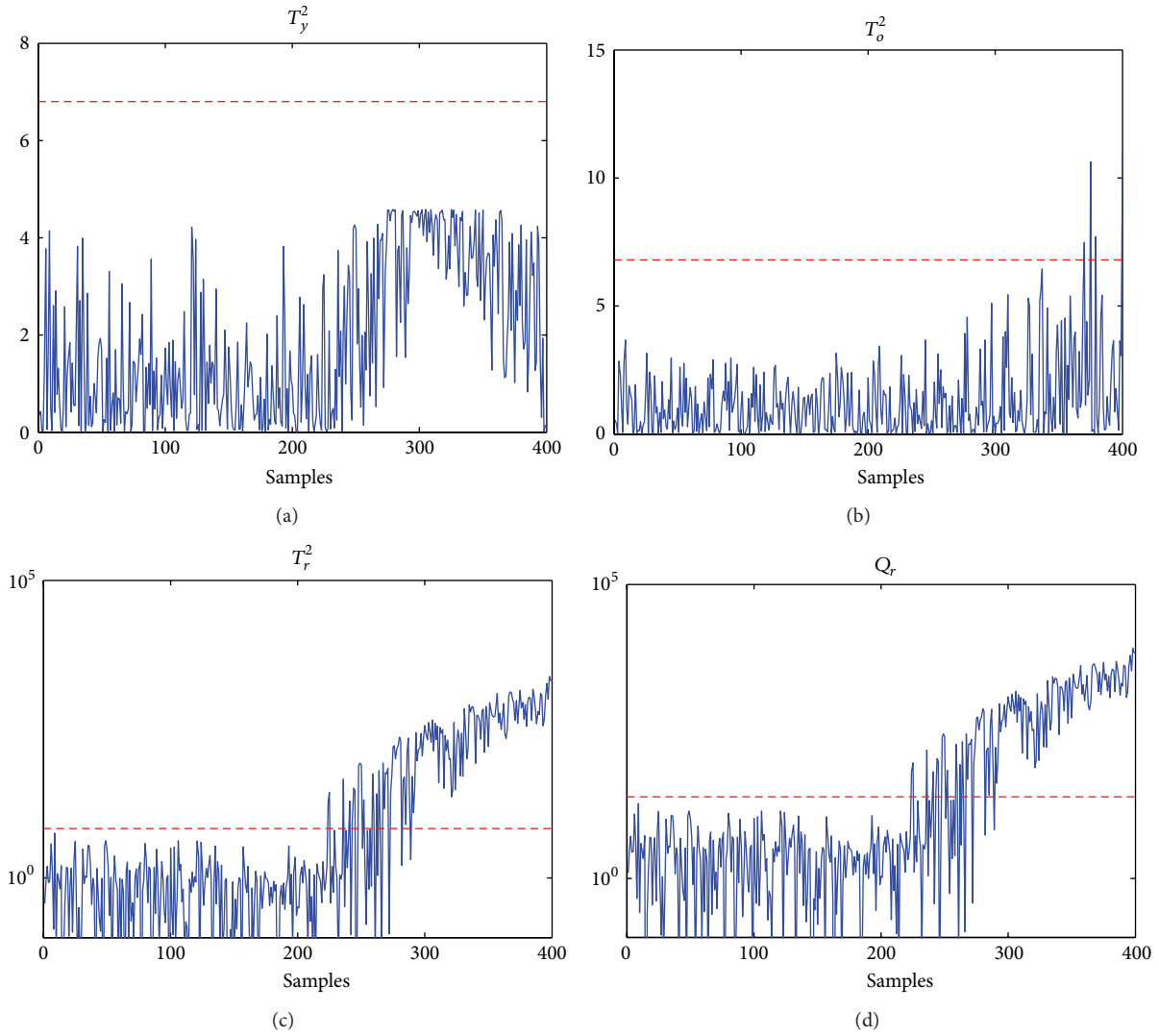

Figure 10: T-KPLS-based monitoring when quality-related Fault 4 occurs.

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_A] \in \Re^{n \times A}$ is the score matrix, and A is KPLS score number, obtained by cross validation [34]. $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_A] \in \Re^{m \times A}$, $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_A] \in \Re^{p \times A}$ are the loadings matrices, and $\mathbf{\Phi}_r$, $\mathbf{Y}_r$ are the residuals matrices.

B. The Proof of $\mathbf{T} = \mathbf{\Phi R}$

First of all, setting $\mathbf{U} = [\mathbf{u_1 u_2} \cdots \mathbf{u_A}]$, $\mathbf{K_1} = \mathbf{K} = \mathbf{\Phi \Phi}^T$. According to the KPLS algorithm in Algorithm 3, the following equations hold:

$$\mathbf{t}_1 = \frac{\mathbf{K_1 u_1}}{\left\|\mathbf{K_1 u_1}\right\|} \Longrightarrow \mathbf{K_1 u_1} = \mathbf{t}_1 \left\|\mathbf{K_1 u_1}\right\| = \mathbf{t}_1 C_{11},$$

$$\mathbf{t}_2 = \frac{\mathbf{K_2 u_2}}{\left\|\mathbf{K_2 u_2}\right\|} = \frac{\left(\mathbf{I} - \mathbf{t}_1 \mathbf{t}_1^T\right) \mathbf{K_1 u_2}}{\left\|\left(\mathbf{I} - \mathbf{t}_1 \mathbf{t}_1^T\right) \mathbf{K_1 u_2}\right\|}$$

$$\Longrightarrow \mathbf{K_1 u_2} = \mathbf{t}_1 \left(\mathbf{t}_1^T \mathbf{K_1 u_2}\right) + \mathbf{t}_2 \left\|\left(\mathbf{I} - \mathbf{t}_1 \mathbf{t}_1^T\right) \mathbf{K_1 u_2}\right\|$$

$$= \mathbf{t}_1 C_{12} + \mathbf{t}_2 C_{22},$$

$$\mathbf{t}_3 = \frac{\mathbf{K_3 u_3}}{\left\|\mathbf{K_3 u_3}\right\|} \Longrightarrow \mathbf{K_1 u_3} = \mathbf{t}_1 \left(\mathbf{t}_1^T \mathbf{K_1 u_3}\right) + \mathbf{t}_2 \left(\mathbf{t}_2^T \mathbf{K_1 u_3}\right)$$

$$+ \mathbf{t}_3 \left\|\left(\mathbf{I} - \mathbf{t}_1 \mathbf{t}_1^T\right) \left(\mathbf{I} - \mathbf{t}_2 \mathbf{t}_2^T\right) \mathbf{K_1 u_3}\right\|$$

$$= \mathbf{t}_1 C_{13} + \mathbf{t}_2 C_{23} + \mathbf{t}_3 C_{33}. \tag{B.1}$$

To sum up,

$$\mathrm{K}_1 \mathbf{u}_A = \mathbf{t}_1 C_{1A} + \mathbf{t}_2 C_{2A} + \mathbf{t}_3 C_{3A} \cdots \mathbf{t}_A C_{AA}. \tag{B.2}$$

Then,

$$\mathbf{K}_1 \mathbf{U} = \mathbf{TC}. \tag{B.3}$$

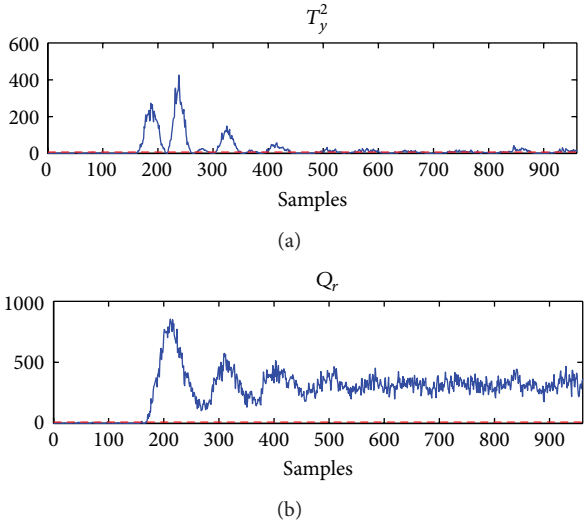

Figure 11: Detection of IDV (1) using T-KPLS (T_y^2 and Q_r). The dashed line represents the 99% control limit.

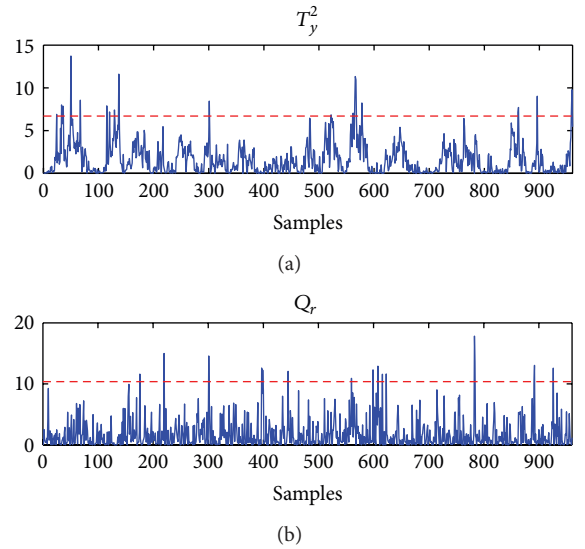

Figure 12: Detection of IDV (3) using T-KPLS (T_y^2 and Q_r). The dashed line represents the 99% control limit.

Here, $\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & \cdots & C_{1A} \\ & C_{22} & C_{23} & \cdots & \\ & & C_{33} & \cdots & \vdots \\ & & & \ddots & \\ & & & & C_{AA} \end{bmatrix}$ is a reversible upper triangle matrix. As $\mathbf{T}$ is a unit orthogonal matrix; namely, $\mathbf{T}^T\mathbf{T} = \mathbf{I}_A$, so $\mathbf{C} = \mathbf{T}^T\mathbf{K}_1\mathbf{U}$; then,

$$\mathbf{T} = \mathbf{K}_1\mathbf{U}\mathbf{C}^{-1} = \mathbf{K}_1\mathbf{U}\left(\mathbf{T}^T\mathbf{K}_1\mathbf{U}\right)^{-1} = \mathbf{\Phi}\mathbf{\Phi}^T\mathbf{U}\left(\mathbf{T}^T\mathbf{K}\mathbf{U}\right)^{-1}. \tag{B.4}$$

Thus $\mathbf{T} = \mathbf{\Phi}\mathbf{R}$ holds.

C. Calculations of Scores and Q_r

Motivated by the calculation in T-PLS model,

$$\begin{aligned}
\mathbf{t}_{y\text{new}} &= \mathbf{Q}_y^T\mathbf{Q}\mathbf{R}^T\phi\left(\mathbf{x}_{\text{new}}\right) \\
&= \mathbf{Q}_y^T\mathbf{Q}\left(\mathbf{U}^T\mathbf{K}\mathbf{T}\right)^{-1}\mathbf{U}^T\mathbf{K}_{\text{new}} = \Theta_y\mathbf{K}_{\text{new}}, \\
\mathbf{t}_{o\text{new}} &= \mathbf{P}_o^T\left(\mathbf{P}\mathbf{R}^T - \mathbf{P}_y\mathbf{Q}_y^T\mathbf{Q}\mathbf{R}^T\right)\phi\left(\mathbf{x}_{\text{new}}\right) \\
&= \mathbf{W}_o^T\mathbf{\Phi}_o\mathbf{\Phi}^T\mathbf{T}\mathbf{R}^T\phi\left(\mathbf{x}_{\text{new}}\right) \\
&\quad - \mathbf{W}_o^T\mathbf{\Phi}_o\mathbf{\Phi}^T\mathbf{T}\mathbf{T}^T\mathbf{T}_y\left(\mathbf{T}_y^T\mathbf{T}_y\right)^{-1}\mathbf{Q}_y^T\mathbf{Q}\mathbf{R}^T\phi\left(\mathbf{x}_{\text{new}}\right) \\
&= \mathbf{W}_o^TZ_y\mathbf{T}\mathbf{T}^T\mathbf{K}\mathbf{T}\left(\mathbf{U}^T\mathbf{K}\mathbf{T}\right)^{-1}\mathbf{U}^T\mathbf{K}_{\text{new}} \\
&\quad - \mathbf{W}_o^TZ_y\mathbf{T}\mathbf{T}^T\mathbf{K}\mathbf{T}\mathbf{T}^T\mathbf{T}_y\left(\mathbf{T}_y^T\mathbf{T}_y\right)^{-1} \\
&\quad \times\mathbf{Q}_y^T\mathbf{Q}\left(\mathbf{U}^T\mathbf{K}\mathbf{T}\right)^{-1}\mathbf{U}^T\mathbf{K}_{\text{new}} \\
&= \Theta_o\mathbf{K}_{\text{new}}, \\
\mathbf{t}_{r\text{new}} &= \mathbf{P}_r^T\left(\mathbf{I} - \mathbf{P}\mathbf{R}^T\right)\phi\left(\mathbf{x}_{\text{new}}\right) \\
&= \mathbf{W}_r^T\mathbf{\Phi}_r\phi\left(\mathbf{x}_{\text{new}}\right) - \mathbf{W}_r^T\mathbf{\Phi}_r\mathbf{P}\mathbf{R}^T\phi\left(\mathbf{x}_{\text{new}}\right) \\
&= \mathbf{W}_r^T\left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right)\mathbf{K}_{\text{new}} \\
&\quad - \mathbf{W}_r^T\left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right)\mathbf{K}\mathbf{T}\left(\mathbf{U}^T\mathbf{K}\mathbf{T}\right)^{-1}\mathbf{U}^T\mathbf{K}_{\text{new}} \\
&= \Theta_r\mathbf{K}_{\text{new}}.
\end{aligned} \tag{C.1}$$

The Q statistic for T-KPLS is as follows:

$$\begin{aligned}
Q_r &= \left\|\phi_{rr}\left(\mathbf{x}_{\text{new}}\right)\right\|^2 = \left\|\phi_r\left(\mathbf{x}_{\text{new}}\right) - \mathbf{P}_r\mathbf{t}_{r\text{new}}\right\|^2 \\
&= \phi_r^T\left(\mathbf{x}_{\text{new}}\right)\phi_r\left(\mathbf{x}_{\text{new}}\right) - 2\phi_r^T\left(\mathbf{x}_{\text{new}}\right)\mathbf{P}_r\mathbf{t}_{r\text{new}} \\
&\quad + \mathbf{t}_{r\text{new}}^T\mathbf{P}_r^T\mathbf{P}_r\mathbf{t}_{r\text{new}}.
\end{aligned} \tag{C.2}$$

The first part of Q_r is detailed in (8). And the second part is

$$\begin{aligned}
&\phi_r^T\left(\mathbf{x}_{\text{new}}\right)\mathbf{P}_r\mathbf{t}_{r\text{new}} \\
&\quad = \left(\phi\left(\mathbf{x}_{\text{new}}\right) - \mathbf{P}\mathbf{t}_{\text{new}}\right)^T\mathbf{\Phi}_r^T\mathbf{W}_r\mathbf{t}_{r\text{new}} \\
&\quad = \phi^T\left(\mathbf{x}_{\text{new}}\right)\mathbf{\Phi}_r^T\mathbf{W}_r\mathbf{t}_{r\text{new}} - \mathbf{t}_{\text{new}}^T\mathbf{P}^T\mathbf{\Phi}_r^T\mathbf{W}_r\mathbf{t}_{r\text{new}} \\
&\quad = \phi^T\left(\mathbf{x}_{\text{new}}\right)\mathbf{\Phi}^T\left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right)\mathbf{W}_r\mathbf{t}_{r\text{new}} \\
&\quad\quad - \mathbf{t}_{\text{new}}^T\mathbf{T}^T\mathbf{\Phi}\mathbf{\Phi}^T\left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right)\mathbf{W}_r\mathbf{t}_{r\text{new}} \\
&\quad = \mathbf{K}_{\text{new}}^T\left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right)\mathbf{W}_r\mathbf{t}_{r\text{new}} \\
&\quad\quad - \mathbf{t}_{\text{new}}^T\mathbf{T}^T\mathbf{K}\left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right)\mathbf{W}_r\mathbf{t}_{r\text{new}}.
\end{aligned} \tag{C.3}$$

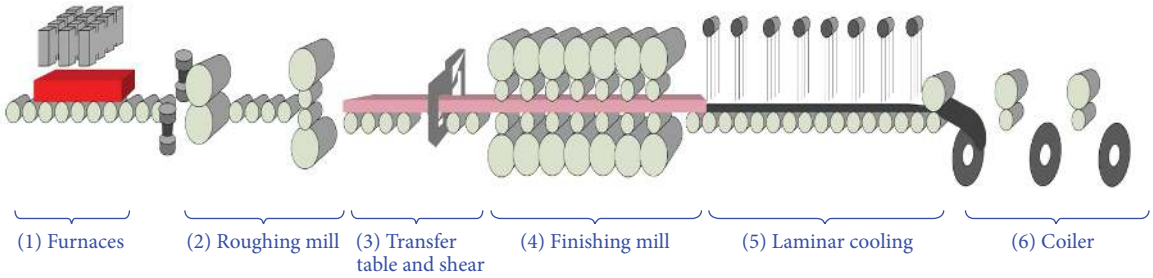

Figure 13: Schematic layout of the hot strip mill.

(1) Set $i = 1$, initialize $\mathbf{u}_i$ as the first column of $\mathbf{Y}_i$.
(2) $\mathbf{t}_i = \mathbf{\Phi}_i \mathbf{w}_i = \mathbf{K}_i \mathbf{u}_i$, where $\mathbf{w}_i = \mathbf{\Phi}_i^T \mathbf{u}_i$.
(3) Scale $\mathbf{t}_i$ to unit length, $\mathbf{t}_i = \mathbf{t}_i / \|\mathbf{t}_i\|$.
(4) $\mathbf{u}_i = \mathbf{Y}_i \mathbf{q}_i$, where $\mathbf{q}_i = \mathbf{Y}_i^T \mathbf{t}_i$.
(5) Scale $\mathbf{u}_i$ to unit length, $\mathbf{u}_i = \mathbf{u}_i / \|\mathbf{u}_i\|$.
Repeat (2)–(5) until $\mathbf{t}_i$ convergence.
(6) Deflate matrices $\mathbf{K}$, $\mathbf{Y}$ and $\mathbf{\Phi}$:

$$\mathbf{\Phi}_{i+1} = \left(\mathbf{I} - \mathbf{t}_i \mathbf{t}_i^T\right) \mathbf{\Phi}_i, \quad \mathbf{Y}_{i+1} = \left(\mathbf{I} - \mathbf{t}_i \mathbf{t}_i^T\right) \mathbf{Y}_i$$
$$\mathbf{K}_{i+1} = \left(\mathbf{I} - \mathbf{t}_i \mathbf{t}_i^T\right) \mathbf{K}_i \left(\mathbf{I} - \mathbf{t}_i \mathbf{t}_i^T\right).$$

(7) Set $i = i + 1$, loop to step (1), until $i > A$.

Algorithm 3: KPLS algorithm.

The last one is

$$\begin{aligned} &\mathbf{t}_{r\text{new}}^T \mathbf{P}_r^T \mathbf{P}_r \mathbf{t}_{r\text{new}} \\ &= \mathbf{t}_{r\text{new}}^T \mathbf{W}_r^T \mathbf{\Phi}_r \mathbf{\Phi}_r^T \mathbf{W}_r \mathbf{t}_{r\text{new}} \\ &= \mathbf{t}_{r\text{new}}^T \mathbf{W}_r^T \left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right) \mathbf{K} \left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right) \mathbf{W}_r \mathbf{t}_{r\text{new}}. \end{aligned} \tag{C.4}$$

By substituting $\mathbf{t}_{r\text{new}}$ with $\Theta_r \mathbf{K}_{\text{new}}$ and combining relevant parts, Q_r can be expressed as

$$Q_r = \phi^T(\mathbf{x}_{\text{new}}) \phi(\mathbf{x}_{\text{new}}) - \mathbf{K}_{\text{new}}^T \Omega_r \mathbf{K}_{\text{new}}. \tag{C.5}$$

Acknowledgements

This work was supported by national 973 projects under Grants 2010CB731800 and 2009CB32602 and by NSFC under Grants (61074084 and 61074085), China, and Beijing Key Discipline Development Program (no. XK100080537). We also appreciate the data support from Ansteel Corporation in Liaoning Province, China.

References

[1] S. J. Qin, "Statistical process monitoring: basics and beyond," *Journal of Chemometrics*, vol. 17, no. 8-9, pp. 480–502, 2003.

[2] B. M. Wise and N. B. Gallagher, "The process chemometrics approach to process monitoring and fault detection," *Journal of Process Control*, vol. 6, no. 6, pp. 329–348, 1996.

[3] J. F. MacGregor, C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process monitoring and diagnosis by multiblock PLS methods," *AIChE Journal*, vol. 40, no. 5, pp. 826–838, 1994.

[4] J. F. MacGregor and T. Kourti, "Statistical process control of multivariate processes," *Control Engineering Practice*, vol. 3, no. 3, pp. 403–414, 1995.

[5] T. Kourti, P. Nomikos, and J. F. MacGregor, "Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS," *Journal of Process Control*, vol. 5, no. 4, pp. 277–284, 1995.

[6] G. Baffi, E. B. Martin, and A. J. Morris, "Non-linear projection to latent structures revisited (the neural network PLS algorithm)," *Computers and Chemical Engineering*, vol. 23, no. 9, pp. 1293–1307, 1999.

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

[8] S. Wold, N. Kettaneh-Wold, and B. Skagerberg, "Nonlinear PLS modeling," *Chemometrics and Intelligent Laboratory Systems*, vol. 7, no. 1-2, pp. 53–65, 1989.

[9] G. Baffi, E. B. Martin, and A. J. Morris, "Non-linear projection to latent structures revisited: the quadratic PLS algorithm," *Computers and Chemical Engineering*, vol. 23, no. 3, pp. 395–411, 1999.

[10] R. Rosipal and L. J. Trejo, "Kernel partial least squares regression in reproducing kernel Hillbert space," *Journal of Machine Learning Research*, vol. 2, pp. 97–123, 2001.

[11] Y. W. Zhang and Z. Y. Hu, "On-line batch process monitoring using hierarchical kernel partial least squares," *Chemical Engineering Research and Design*, vol. 89, no. 10, pp. 2078–2084, 2011.

[12] K. Kim, J. M. Lee, and I. B. Lee, "A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction," *Chemometrics and Intelligent Laboratory Systems*, vol. 79, no. 1-2, pp. 22–30, 2005.

[13] Y. W. Zhang, H. Zhou, S. J. Qin, and T. Y. Chai, "Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares," *IEEE Transactions on Industrial Informatics*, vol. 6, no. 1, pp. 3–10, 2010.

[14] H. Zhang, Y. Shi, and M. Liu, "H1 step tracking control for networked discrete-time nonlinear systems with integral and predictive actions," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 337–345, 2013.

[15] H. Zhang, Y. Shi, and B. Mu, "Optimal H1-based linear-quadratic regulator tracking control for discrete-time Takagi-Sugeno fuzzy systems with preview actions," *ASME Transactions, Journal of Dynamic Systems, Measurement, and Control*, vol. 135, Article ID 044501, 5 pages, 2013.

[16] G. Li, S. J. Qin, and D. H. Zhou, "Geometric properties of partial least squares for process monitoring," *Automatica*, vol. 46, no. 1, pp. 204–210, 2010.

[17] G. Li, S. J. Qin, and D. H. Zhou, "Output relevant fault reconstruction and fault subspace extraction in total projection to latent structures models," *Industrial and Engineering Chemistry Research*, vol. 49, no. 19, pp. 9175–9183, 2010.

[18] S. W. Choi and I. B. Lee, "Multiblock PLS-based localized process diagnosis," *Journal of Process Control*, vol. 15, no. 3, pp. 295–306, 2005.

[19] D. H. Zhou, G. Li, and S. J. Qin, "Total projection to latent structures for process monitoring," *AIChE Journal*, vol. 56, pp. 168–178, 2010.

[20] G. Li, C. F. Alcala, S. J. Qin, and D. Zhou, "Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the Tennessee Eastman process," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 5, pp. 1114–1127, 2011.

[21] G. Li, S. J. Qin, Y. D. Ji, and D. H. Zhou, "Total PLS based contribution plots for fault diagnosis," *Acta Automatica Sinica*, vol. 35, no. 6, pp. 759–765, 2009.

[22] J. Yu and S. J. Qin, "Multimode process monitoring with bayesian inference-based finite Gaussian mixture models," *AIChE Journal*, vol. 54, no. 7, pp. 1811–1829, 2008.

[23] J. Yu, "Localized Fisher discriminant analysis based complex chemical process monitoring," *AIChE Journal*, vol. 57, no. 7, pp. 1817–1828, 2011.

[24] J. Yu, "A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes," *Chemical Engineering Science*, vol. 68, no. 1, pp. 506–519, 2012.

[25] H. Zhang, Y. Shi, and A. Saadat Mehr, "On H∞ filtering for discrete-time takagi-sugeno fuzzy systems," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 2, pp. 396–401, 2012.

[26] E. L. Russell, L. H. Chiang, and R. D. Braatz, "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 81–93, 2000.

[27] H. J. Cho, J. Lee, S. W. Choi, D. Lee, and I. Lee, "Fault identification for process monitoring using kernel principal component analysis," *Chemical Engineering Science*, vol. 60, no. 1, pp. 279–288, 2005.

[28] S. Mika, B. Schölkopf, A. Smola, and K. R. Müller, "Kernel PCA and de-noising in feature spaces," *Advances in Neural Information Processing Systems*, vol. 11, pp. 536–542, 1999.

[29] R. D. Jia, Z. Z. Mao, Y. Q. Chang, and S. N. Zhang, "Kernel partial robust M-regression as a flexible robust nonlinear modeling technique," *Chemometrics and Intelligent Laboratory Systems*, vol. 100, no. 2, pp. 91–98, 2010.

[30] G. Lee, C. H. Han, and E. S. Yoon, "Multiple-fault diagnosis of the Tennessee Eastman process based on system decomposition and dynamic PLS," *Industrial and Engineering Chemistry Research*, vol. 43, no. 25, pp. 8037–8048, 2004.

[31] L. H. Chiang, E. L. Russell, and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 2, pp. 243–252, 2000.

[32] M. Kano, S. J. Hasebe, I. Hashimoto, and H. Ohno, "A new multivariate statistical process monitoring method using principal component analysis," *Computers and Chemical Engineering*, vol. 25, no. 7-8, pp. 1103–1113, 2001.

[33] J. M. Lee, C. K. Yoo, and I. B. Lee, "Statistical monitoring of dynamic processes based on dynamic independent component analysis," *Chemical Engineering Science*, vol. 59, no. 14, pp. 2995–3006, 2004.

[34] S. Wold, "Cross-validatory estimation of the number of components In factor and principal components models," *Technometrics*, vol. 20, pp. 397–405, 1975.