Contribution rate plot for nonlinear quality-related fault diagnosis with application to the hot strip mill process

Kaixiang Peng a, Kai Zhang a,*, Gang Li b,*, Donghua Zhou b

a 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, PR China
b Department of Automation, TNList, Tsinghua University, Beijing 100084, PR China

ARTICLE INFO

Article history:
Received 14 April 2012
Accepted 30 November 2012

Keywords:
Total kernel PLS
Nonlinear process
Fault diagnosis
Contribution rate
Hot strip mill process

ABSTRACT

In this paper, a nonlinear fault diagnosis scheme is established for the hot strip mill process (HSMP). In HSMP, the faults affecting quality index are denoted as quality-related faults, which should be taken care of as soon as possible. Projection to latent structures (PLS) is a basic model for quality-related fault detection in linear processes. In the presented work, a total kernel PLS (T-KPLS) model is utilized for modeling and monitoring HSMP, which is a typical nonlinear process. However, diagnosis tools have not been developed aiming at the nonlinear case based on T-KPLS model. Motivated by the successful use of contribution plot for the linear case, a contribution rate plot is proposed to extend contribution plots to the nonlinear case. In the end of this paper, the proposed method is applied to the hot strip mill process effectively.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Hot strip mill process (HSMP) is a competitive business in which quality and efficiency are of paramount importance. The key performance indicators of the products in HSMP are thickness, crown and width, of which thickness is most critical. Nowadays, single-loop control of traditional automatic control technologies is not the core issue. Process monitoring as well as scheduling, optimization problems have become the important contents of industrial process control. The factors such as high temperature, high pressure and high speed lead to severe faults or abnormal circumstances in HSMP, which may result in significant economic losses. Thus, a real-time monitoring of operating conditions in the rolling process is quite desired. Based on the availability of analytical process model, model-based techniques have been well established in the framework of modern control theory (Ding, 2008; Patton, Frank, & Clark, 2000). However, model-based technique is not practical for HSMP, as it is hard to acquire accurate physical models. In the past two decades, multivariable statistical process monitoring (MSPM) technologies have attracted more and more attentions as a cluster of popular data-driven approaches (Chiang, Russell, & Braatz, 2001; Qin, 2003).

For an industrial process like HSMP, all collected variables can be roughly divided into two groups: process variables \( \mathbf{x}_i \) (\( i = 1, \ldots, m \)) and quality variables \( \mathbf{y}_i \) (\( i = 1, \ldots, p \)), where \( \mathbf{x}_i \) reflects process state or environment such as gap, rolling force or bending force, and \( \mathbf{y}_i \) indicates the products’ quality or key performance of processes such as thickness or flatness. Generally, process data \( \mathbf{x}_i \) can be obtained instantaneously, whereas quality data \( \mathbf{y}_i \) are often difficult to be sampled on line, and usually come with significant sampling delays. In practice, quality variables are affected and determined by the process variables as well as possibly additional unmeasured factors. However, not all of faults in process variables will affect quality variables. Some of faults affect processes but have no impact on products’ quality. Others change the environment which leads to disturbance on quality index. The faults occurring in the process variables that have impacts on quality variables ultimately are labeled as quality-related faults, which should be dealt with urgently. For instance, in HSMP, the exit thickness in finishing mill is closely relevant with rolling force or gaps between rollers in all stands. Once faults occur in sensors or actuators associated with these items, thickness is bound to be violated. Therefore, this kind of faults deserve more attentions (Li, Qin, & Zhou, 2010b, 2009).

In MSPM, multivariate projection techniques such as PCA and PLS have been employed. Both methods are able to project process variable \( \mathbf{x} \) onto a low-dimensional space which is usually called principal space, in which multivariate control chart and contribution plot are utilized for fault detection and diagnosis, respectively. PCA models only make use of process data \( \mathbf{x} \) for monitoring...
Morris (1999a, 1999b, 2000), Zhao, Zhang, Xu, and Xiong (2006), fault diagnosis with the proposed method. In the end, conclusions index. Section 4 describes hot strip mill process, and performs method is proposed and discussed based on a combined detection variables in hot strip mill process.

alarms, and proposed total PLS models to improve monitoring related process monitoring led to more false alarms and missing alarms, and concluded that the standard PLS was appropriate (Li, Qin, & Zhou, 2010a). Zhou et al. recognized that applying PLS for quality-related fault detection (Peng, Zhang, Li, & Zhou, 2009) to transform into a feature space via a nonlinear mapping, and other nonlinear PLS models, in which original input data are transformed into a feature space via a nonlinear mapping, and gives a further decomposing scheme is established based on T-KPLS model. At first, T-KPLS outcome of introducing kernel technique into T-PLS model.

T-KPLS possesses the characteristics of both KPLS and T-PLS. It under review). In nature, one can interpret T-KPLS structure with two different aspects: (1) it is a postprocess of KPLS, (2) it is the contribution rate for fault diagnosis. Choi et al. extended it to the multi-mode and non-Gaussian, and the whole process operates under a universal mode. However, the nonlinearity of HSMP is focused only. Multi-mode and non-Gaussian approaches also contain the assumptions that the process variables are normal distributed, and the process operates under a universal mode. However, the variables’ correlations in most industrial processes such as HSMP are non-Gaussian or approximately Gaussian, and the whole processes may be switched in multiple conditions. In recent years, many improvements have been made to handle the shifting modes (e.g. Yu, 2011, 2012; Yu & Qin, 2008). While for non-Gaussian property, various effective methods have also been reported without Gaussian assumption (e.g. Ge, Gao, & Song, 2011; Rashid & Yu, 2012a, 2012b). However, owing to the scope of paper, the nonlinearity of HSMP is focused only. Multi-mode and non-Gaussian cases will be considered in the further research.

For nonlinear processes, Peng et al. proposed the total kernel PLS for quality-related fault detection (Peng, Zhang, Li, & Zhou, under review). In nature, one can interpret T-KPLS structure with two different aspects: (1) it is a postprocess of KPLS, (2) it is the outcome of introducing kernel technique into T-PLS model. T-KPLS possesses the characteristics of both KPLS and T-PLS. It can model nonlinear input data, and gives a further decomposition of KPLS.

It is necessary to isolate the faulty variables after a fault is detected. Contribution plot is one of the most popular method, as it can tell faulty variables without known fault information. However, it is difficult to construct traditional contribution plot for kernel based models (Alcala & Qin, 2011; Alcala, Qin, & Zhou (2011)). For kernel PCA models, Cho, Lee, and Choi (2005) utilized the concept of contribution rate for fault diagnosis. Choi et al. extended it to the multi-scale kernel PCA (Choi, Julian, & Lee, 2008). Nevertheless, this contribution rate predefined in Rakotomamonjy (2003) does not have explicit physical meanings which makes it difficult to explain the origination of a fault. In this paper, a new contribution rate plot based on the T-KPLS model is proposed for identifying faulty variables in hot strip mill process.

The rest of this paper is organized as follows. Section 2 presents a brief introduction to T-KPLS model and T-KPLS based fault detection policy. In Section 3, a new contribution rate method is proposed and discussed based on a combined detection index. Section 4 describes hot strip mill process, and performs fault diagnosis with the proposed method. In the end, conclusions are presented in Section 5.
model is built with historical process and quality data under normal operation. Then the new scores and residuals are calculated from the new sample. Four statistic plots are constructed with their corresponding control limits, which are used for monitoring variations in process data \( X \) according to quality data \( Y \).

For a new process sample, three new score vectors are calculated as shown in A.2. Assuming that the training data be multivariate normal in the feature space, Hotelling's \( T^2 \) statistic is utilized to measure systematic variations. Taking \( T^2 \) for example, \( A_2 = (1/(n-1))T^2 \), \( T^2 \) calculates the Mahalanobis distance between the new score \( t_{new} \) and the original scores \( T \). \( Q \) is used to detect the abnormality occurring in the residuals. The control limits of \( T^2 \)-statistic and Q-statistic are obtained under the assumption that the measurements in the feature space are multivariate normal. Noted that \( Q \) is calculated in detail as shown in Appendix A.3. According to the property of T-KPLS model is built with historical process and quality data under normal condition, it is able to conclude that the choice of \( \Omega \), \( A \), and \( c \) are optimal and fixed. Otherwise, bisection and cross-validation mentioned in Rosipal and Trejo (2001) is applied to search for the optimal \( c_{max} \).

\[
\begin{align*}
\phi &= \frac{T^2_y}{\delta_y} + \frac{Q_r}{\delta_{Q_r}} = \frac{t_{new}^T A_2^{-1} t_{new}}{\delta_y} + \frac{\|\phi_{Q_r}(X_{new})\|^2}{\delta_{Q_r}} \\
&= \frac{\phi^T(X_{new})RQ^1r_{Q_r}Q^1r_{Q_r} \phi(X_{new})}{\delta_y} \\
&\quad + \frac{\phi^T(Y_{new})(I-R_P)^{T}(I-P_P)^{T}(I-R_P)^{T}\phi(X_{new})}{\delta_{Y}} \\
&= K_{new}^TQ_{new} + K_{new}(X_{new},X_{new}) \\
\end{align*}
\]

where \( Q \in \mathbb{R}^{n \times n} \) can be obtained in Appendix B.

The control limits of combined index are estimated with nonparametric approaches. Among various probability density function estimation methods, kernel density estimation (KDE) is a simple technique without imposition of parametric model, which is well established for univariate random process. Chen, Kruger, Meronk, and Leung (2004) proposed a synthesis KDE method to estimate the control limit for a synthesis index. Odiowei and Cao (2010) utilized KDE to estimate the control limit for CVA, PCA and PLS, and compared the detecting results with the traditional methods. In this work, KDE is used to calculate the control limit of \( \phi \). KDE gives the estimation of PDF using Gaussian kernel function: \( \hat{f}(\phi) = (1/\sqrt{n}) \sum_{i=1}^{n} k_c((\phi - \phi_i)/\delta) \), where \( k_c(x) = (1/ \sqrt{2\pi} \sigma^{1/2}) e^{-x^2/(2\sigma^2)} \), \( \delta \) is the width of kernel. The optimal width is given by: \( \delta_{opt} = 1.06\sigma N^{-0.25} \) according to Odiowei and Cao (2010), \( \sigma \) is the standard variance of samples. For more information related to KDE, one can also refer to Statistical Pattern Recognition Toolbox for Matlab (ftp://cmp.felk.cvut.cz/pub/cmp/articles/Franc-TR-2004-08.pdf).

### Table 1

<table>
<thead>
<tr>
<th>Index</th>
<th>Calculation</th>
<th>Control limit Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T^2_y )</td>
<td>( t_{new}^T A_2^{-1} t_{new} )</td>
<td>( \delta_y )</td>
</tr>
<tr>
<td>( Q_r )</td>
<td>( |Q_r|^2 )</td>
<td>( \delta_{Q_r} )</td>
</tr>
<tr>
<td>( T^2 )</td>
<td>( t_{new}^T A_2^{-1} t_{new} )</td>
<td>( \delta_y )</td>
</tr>
<tr>
<td>( T^2 )</td>
<td>( t_{new}^T A_2^{-1} t_{new} )</td>
<td>( \delta_y )</td>
</tr>
</tbody>
</table>

\( F \)-distribution with \( A_2 \) and \( n-A_2 \) degrees of freedom.
\( g \cdot h = \text{mean}(Q), \) \( 2g^2 \cdot h = \text{var}(Q) \).

#### 3.2. Fault diagnosis based on T-KPLS

HSMP operates under extremely complicated atmosphere where the variables are correlated nonlinearly. Thus, it is hard to locate the faulty variables accurately. For linear models, people often use contribution plots to reveal the variables which are responsible for the fault (Choi & Lee, 2005; Lee, Han, & Yoon, 2004). However, it is hard to calculate contributions for each variable in nonlinear models (Alcala & Qin, 2011). Cho et al. (2005) and Rakotomamonjy (2003) proposed a method to measure contribution of each variable using the gradient of kernel function in kernel PCA models and Choi et al. (2008) expanded it to multi-scale kernel PCA based methodology. However, the physical meanings of their approaches are not demonstrated clearly. In this paper, a contribution rate plot is proposed to measure the contribution to index \( \phi \), to find out which should be in charge of the disturbance in quality data.

#### 3.2.1. Definition of contribution rate

Let \( \psi = [\psi_1, \ldots, \psi_m]^T \), where \( \psi_i > 0 \), \( i = 1, \ldots, m \), and \( X \otimes \psi \) represents the vector of \( [X_1\psi_1, \ldots, X_i\psi_i, \ldots, X_m\psi_m] \), where \( X_i \psi_i \) represents the variation of variable \( X_i \) which can be described piecewise: if \( \psi_i > 1 \), \( |X_i\psi_i| > |X_i| \); if \( \psi_i = 1 \), \( X_i\psi_i = X_i \); if \( \psi_i < 1 \), \( |X_i\psi_i| < |X_i| \). Thus, \( \phi(X \otimes \psi) = \phi(X) \) holds.

Then the first-order Taylor series expansion can be performed on \( \phi(X \otimes \psi) \) around \( \psi = 1m \):

\[
\phi(X \otimes \psi) \approx \phi(X) + \sum_{i=1}^{m} \frac{\partial \phi(X \otimes \psi)}{\partial \psi_i} |_{\psi = 1m} (\psi_i - 1) \\
\]

In the end, based on the above conclusions, the definition of contribution rate is illustrated:

**Definition 1.** Given a measured sample \( X \). \( C(x) \doteq \| \phi(X \otimes \psi) / \psi \| \) is defined as the contribution rate of the \( j \)th variable to index \( \phi \).
As described above, contribution rates means variable's contribution gradient to the index \( \varphi \). If all the variables are brought back to the normal level with a same little percent, the increase provided by the variables with large contribution rate to index \( \varphi \) will be large. Thus those variables which have large contribution rates are able to affect \( \varphi \) and quality variables significantly.

3.2.2. Calculation of contribution rate

In this part, the contribution rates of all variables are given in computable forms. Eq. (4) is firstly expanded to

\[
\varphi_{\text{new}} = \frac{K(x_{\text{new}}, x_{\text{new}}) + K^T_{\text{new}} \Omega K_{\text{new}}}{\delta r} + K^T_{\text{new}} \Omega K_{\text{new}}
\]

Moreover, there is a remark to make use of contribution rate Calculation of

Further more, there is a remark to make use of contribution rate calculation of \( K_{\text{PCA}} \), etc.

\[
\text{In the end, the ultimate expression for calculation of contribution rate can be illustrated in the following:}
\]

\[
C(x_{\text{new}, i}) = \frac{\text{trace} \left( \Omega \hat{x}_{\text{new}, i}^{\Sigma} \right)}{m_{\epsilon}(i)}
\]

where \( \text{trace}(\cdot) \) means the trace of a matrix, the detailed derivation of this expression is demonstrated in Appendix C.

3.2.3. Implementation issues

Before practical implementing, it needs to be noted that the variables' contribution rates under normal condition are different from each other, since the feature mapping may affect variables unevenly. Thus it is meaningful to normalize the raw contribution rates to make sure that all variables give the same contribution rates under normal operating situations statistically. Namely, the ultimate contribution rates for both training and testing samples are divided by its normal mean:

\[
C(x, i) = \frac{C_{\text{raw}}(x, i)}{m_{\epsilon}(i)}
\]

(8)

\[
C(x_{\text{new}, i}) = \frac{C_{\text{raw}}(x_{\text{new}, i})}{m_{\epsilon}(i)}
\]

(9)

where \( m_{\epsilon}(i) = \frac{1}{n} \sum_{j=1}^{n} C_{\text{raw}}(x, i) \). The variables with relevant large contribution rate are identified as latent faulty variables. Further more, there is a remark to make use of contribution rate in practice as follows:

**Remark 1.** This diagnosis method also can be introduced into other kernel models, such as KPCA, etc.

3.3. Model implementation

Implementation of the proposed diagnosis scheme involves off-line training model and real-time testing model. As shown in Fig. 1, the training model consists of three steps. The first one is the establishment of T-KPLS model with normal dataset. Then the calculation of \( \varphi \) from \( \varphi \) is followed later. Subsequently, the mean contribution rates \( m_{\epsilon} \) are derived based on \( C_{\text{raw}}(x) \). Meanwhile, the schematic plot for the testing model is sketched in Fig. 2. The whole procedure involves the acquisition of online measurement, the calculation of \( \varphi_{\text{new}} \) and the contribution rates for the new sample. The contribution rates are used for fault diagnosis in the following part.

4. Case study on hot strip mill

4.1. Hot strip mill Process

Hot strip mill process (HSMP) is an extremely complex process in steel manufacturing industry. Due to the increasing quality requirements of the steel products over the worldwide market, e.g. thickness, width or flatness, diagnosis of the hot strip mill process could bring significant economic and social benefits and thus becomes one of the major interesting topics in steel manufacturing industry. Fig. 3 lays out a real HSMP photo in a steel plant (Ansteel corporation in Liaoning Province, China), which serves as the background plant in this paper. A schematic layout of the hot strip mill is illustrated in Fig. 4 corresponding to the real industrial hot strip mill. According to Fig. 4, the process generally consists of the following units: reheating furnaces, roughing mill, transfer table, crop shear, finishing mill, run out table cooling and coiler. In the roughing mill, the thickness of hot steel slabs is reduced to set gauge appropriately and its length proportionally increases with the thickness reduction, since the width of hot steel slab remains unchanged. Then, the finishing mill performs further and more precise gauge reduction, which serves as a background process in this part. This rolling process is characterized by the high efficiency and production rate. The annual output of 1700 mm width hot rolling production line is about 350 million tons with strip rolling speed up to 20 m/s.

As shown in Fig. 4, there are 7-stand finishing mills in the underlying process. Each mill stand consists of a mill housing containing two cylindrical work rolls and a pair of larger backup
rolls to support them. Each stand has its own drive unit offering major drive power to the work rolls. The distance between the work rolls, or the roll gap, is adjusted by positioning the backup rolls through fast hydraulic actuators. The thickness is measured by an X-ray device in finishing mill exit. Automatic gauge control (AGC) controls the gaps of finishing mill stands to produce the desired exit strip thickness. During thickness reduction of the strip its shape undergoes mechanical deformation. The side effects of the gauge control are variations of the strip width and crown. The crown can be influenced by changing the bending actuators forces which are applied between the two work rolls and between work roll and back up roll as shown in Fig. 5. The change of bending force little affects the exit thickness of the stand mill. The rolling speed is set to allow the last stand to perform the final reduction at the finishing temperature, between 850 °C and 950 °C, specified to reach certain mechanical properties. Finishing temperature control (FTC) can control the temperature of the strip at the exit of the finishing mill. Its primary function is to calculate and adjust the acceleration of the finishing mill to the desired temperature at the exit of the finishing mill. Simultaneously, the purpose of the FTC based on the sprays model is to calculate and adjust the conditions of the mill, including spray flow references and so on, achieving the desired temperature of the strip finishing. The hot steel is quite fragile as it is rolled and the tension between the finishing mill stands must be closely controlled at very low levels in order to avoid stretching or tearing the strip. The mill stands are separated by low-tension loopers which ensure accurate speed tracking of the product through the mill. The strip emerging from the finishing mill is generally much longer than the run out table cooling. As a result, the cooling starts before the tail end leaves the finishing mill. The strip is cooled by the water from coolant headers along the run out table cooling and then wrapped into a coil by the coiler unit, which is essentially a rotating mandrel. The major elements of the quality of the products in a hot strip mill are thickness, crown, width, temperature of the strip at the exit of finishing mill, etc. Specially, the thickness is important to determine the quality of the strip and the productivity of the steel company. Meanwhile, strips in aspect of thickness have various scales to meet different demands.

4.2. Thickness related fault detection and diagnosis

In the steel industry many manufacturers have built data base to store measurable information like gap, rolling force and bending force of each stand. 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 simulation with thicknesses shown in Fig. 7, where their thickness targets are 3.95 mm and 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 3. Gaussian kernel parameter \( c \) affects detection results for this process significantly. Generally speaking, the larger the \( c \) is, the lower the false alarm rates and the higher the missing alarm rates will be. In this simulation, two T-KPLS models are built as shown in Table 2, where \( A \) is determined according to cross validation, \( A_1 = 1 \) because of the single output, \( A_2 \) is obtained by KPCA based method. For both models, \( c_{\text{min}} = 0 \) and \( c_{\text{max}} = 10,000 \) are chosen to search for the optimal \( c \).

Before implementing contribution rate plot for HSMP, it is necessary to determine the threshold of the contribution rate for each variable. Because of smearing effect, it is not efficient to use the absolute control limit of contribution rate for diagnosis. Here, the relative contribution rate: 

\[
C_r(x_{\text{new}}, i) = \frac{C(x_{\text{new}}, i)}{\sum_{i=1}^{m} C(x_{\text{new}}, i)}
\]

is utilized, which guarantees \( \sum_{i=1}^{m} C_r(x_{\text{new}}, i) = 1 \). This is because the use of the relative contribution is more convenient to identify
faulty variables. Note that the relative contribution rates of all variables are equivalent statistically under normal condition, namely around the level of \(1/m\), while faulty variables related to this fault will become larger than the level. Therefore, it is reasonable to determine the threshold of relative contribution rate plot as \(\theta_C = 1/m\).

To sum up, the flowchart of utilizing the new proposed approach for HSMP is given in Fig. 6, which includes the preliminary steps and realtime monitoring steps. Note that it is merely designed for one class of strip steel, while for different strips, the training model should be built according to corresponding historical dataset.

The environment in a hot strip mill is very harsh. Heat, vibrations and large forces are significantly affecting equipments like actuators and sensors. Furthermore, some man-made abnormal operations are also familiar in HSMP. Among all these faults, some of them can cause significant influence on the final thickness, while others will not. In this case study, three kinds of frequently occurring faults are mainly studied, which are listed in Table 4, where all faults with the same duration time of 10 s are terminated artificially. The exit thicknesses under three faulty circumstances are shown in Fig. 8. It is obvious that Fault 1 is quality-unrelated, others are quality-related.

### 4.3. Application discussion

The results of quality-related fault detection with \(\varphi\) are given in Figs. 9, 10 and 13, respectively. As shown in Fig. 9, T-KPLS based method just gives a little false alarm rate, while for quality-related Faults 2 and 3, \(\varphi\) delivers good performances. Furthermore, the results of detection rates or false alarm rates with \(\varphi\) are summarized in Table 5, where T-KPLS is more appropriate.

Fault 1 involves a step change in the measurement of bending force on \(F_5\) stand. When the fault occurs, there is a sudden increase in \(x_{18}\). Subsequently the close loop control reacts to drop the bending force in \(F_6\) \((x_{19})\) as well as \(F_7\) \((x_{20})\). However, this fault only affects the flatness of the strip rather than the exit thickness. According to Fig. 9, the fault is quality-unrelated and undetectable with \(\varphi\) which is consistent with Fig. 8.

Fault 2 induces a malfunction of hydraulic gap control loop in \(F_4\) stand, which will influence on the gap of \(F_4\) stand immediately \((x_4)\), then the total force of this stand \((x_{11})\) varies correspondingly. With the function of close loop control, both gaps and force in the subsequent stands are also violated in response to this fault.

### Table 2

<table>
<thead>
<tr>
<th>Models by T-KPLS for HSMP.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 (c=5000)</td>
<td>(A=8, A_x=1, A_y=A_y=7, A_t=10)</td>
</tr>
<tr>
<td>Model 2 (c=7500)</td>
<td>(A=10, A_y=1, A_y=A_y=9, A_t=16)</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Variable Type</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_1 - x_7)</td>
<td>Measured (F_i) stand average gap, (i=1,\ldots,7)</td>
<td>mm</td>
</tr>
<tr>
<td>(x_8 - x_{14})</td>
<td>Measured (F_i) stand total force, (i=1,\ldots,7)</td>
<td>MN</td>
</tr>
<tr>
<td>(x_{15} - x_{20})</td>
<td>Measured (F_i) stand work roll bending force, (i=2,\ldots,7)</td>
<td>MN</td>
</tr>
<tr>
<td>(y) Quality</td>
<td>Finishing mill exit strip thickness</td>
<td>mm</td>
</tr>
</tbody>
</table>

Fig. 6. Schematic plot for HSMP with new proposed approach.
According to the exit thickness as shown in Fig. 8, the quality variable is shown to be faulty with a time delay. However, the monitoring result sketched in Fig. 10 gives a timely detection without delay. To this point, the detection scheme is significant to provide a reference for workers to initialize a maintenance strategy in time. From Fig. 11, it can be seen that the contribution rates are severely inflected from the time of being detected. Furthermore, a real time diagnosis for this fault is given in Fig. 12, where the mark $J$ is labeled as an effective diagnosis at the corresponding time. In the end, $x_4$ is first identified, thus, it can be judged that variable 4 is the faulty resources, and variables 5, 11, 6, 12 are running with fault then. The diagnosis results with relative contribution rate do not only determine the faulty variables but also point out their time sequence, which shows the causality of these variables and helps to find the root cause ultimately.

Fault 3 is an actuator fault of cooling valve between $F_2$ and $F_3$, which is prevalent in finishing mill process. Its occurrence causes severe variations both in stand total force and gap ($x_2$, $x_7$ and $x_10$) from $F_3$ to $F_7$ stand immediately. Exit thickness is influenced ultimately as shown in Fig. 8. $\varphi$ based detection plotted in Fig. 13 is able to detect this fault effectively and timely. Contribution rates for this fault are presented in Fig. 14. Real-time diagnosis results are shown in Fig. 15, where $x_3$ and $x_{10}$ are identified successively from the outset of Fault 3, then, with the propagation of this fault, other faulty variables are also picked up. The diagnosis results reveal the fact clearly. When Fault 3 occurs, $x_3$ is firstly influenced, so is $x_{10}$ next, later, another relevant variables are affected. From above discussion, it can be concluded that the contribution rate based nonlinear diagnosis method works well for thickness-related fault detection and isolation.

![Fig. 8. Faulty exit thickness in finishing mill sampled with a time interval of 10 ms.](image)

![Fig. 9. Detection result for Fault 1 with $\varphi$.](image)

![Fig. 10. Detection result for Fault 2 with $\varphi$.](image)

![Fig. 11. Relative contribution rate values of Fault 2 sampled from 15 s to 25 s.](image)

According to the exit thickness as shown in Fig. 8, the quality variable is shown to be faulty with a time delay. However, the monitoring result sketched in Fig. 10 gives a timely detection without delay. To this point, the detection scheme is significant to provide a reference for workers to initialize a maintenance strategy in time. From Fig. 11, it can be seen that the contribution rates are severely inflected from the time of being detected. Furthermore, a real time diagnosis for this fault is given in Fig. 12, where the mark $\circ$ is labeled as an effective diagnosis at the corresponding time. In the end, $x_4$ is first identified, thus, it can be judged that variable 4 is the faulty resources, and variables 5, 11, 6, 12 are running with fault then. The diagnosis results with relative contribution rate do not only determine the faulty variables but also point out their time sequence, which shows the causality of these variables and helps to find the root cause ultimately.

Fault 3 is an actuator fault of cooling valve between $F_2$ and $F_3$, which is prevalent in finishing mill process. Its occurrence causes severe variations both in stand total force and gap ($x_2$, $x_7$ and $x_10$) from $F_3$ to $F_7$ stand immediately. Exit thickness is influenced ultimately as shown in Fig. 8. $\varphi$ based detection plotted in Fig. 13 is able to detect this fault effectively and timely. Contribution rates for this fault are presented in Fig. 14. Real-time diagnosis results are shown in Fig. 15, where $x_3$ and $x_{10}$ are identified successively from the outset of Fault 3, then, with the propagation of this fault, other faulty variables are also picked up. The diagnosis results reveal the fact clearly. When Fault 3 occurs, $x_3$ is firstly influenced, so is $x_{10}$ next, later, another relevant variables are affected. From above discussion, it can be concluded that the contribution rate based nonlinear diagnosis method works well for thickness-related fault detection and isolation.

### Table 4

<table>
<thead>
<tr>
<th>Fault no.</th>
<th>Description</th>
<th>Occurrence time (s)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sensor fault of bending force in $F_5$ stand of strip 1</td>
<td>10</td>
<td>Quality-unrelated</td>
</tr>
<tr>
<td>2</td>
<td>Malfunction of gap control loop in $F_4$ stand of strip 2</td>
<td>20</td>
<td>Quality-related</td>
</tr>
<tr>
<td>3</td>
<td>Fault of cooling valve between $F_2$ and $F_3$ in strip 1</td>
<td>10</td>
<td>Quality-related</td>
</tr>
</tbody>
</table>

### Table 5

<table>
<thead>
<tr>
<th>Fault no.</th>
<th>Type of detection</th>
<th>$PLS(T^2)$</th>
<th>$KPLS(T^2)$</th>
<th>$T-PLS(T^2)$ or $Q_r$</th>
<th>$T-KPLS(\varphi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>False alarm rate</td>
<td>0.104</td>
<td>0.117</td>
<td>0.366</td>
<td>0.032</td>
</tr>
<tr>
<td>2</td>
<td>Detection rate</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>3</td>
<td>Detection rate</td>
<td>0.656</td>
<td>0.870</td>
<td>0.900</td>
<td>0.989</td>
</tr>
</tbody>
</table>

![Fig. 12. Real-time diagnosis results are shown in Fig. 15, where $x_3$ and $x_{10}$ are identified successively from the outset of Fault 3, then, with the propagation of this fault, other faulty variables are also picked up.](image)
5. Conclusions

In this paper, the purpose is to deal with the nonlinear process diagnosis for hot strip mill process based on T-KPLS model. Firstly, the quality-related detection indices $T^2_y$ and $Q_r$ are incorporated into a combined index $j$. And a KDE based method is utilized to calculate the control limit of the new index. Then, contribution rate plots are developed for quality-related fault diagnosis based on $j$. To show the effectiveness, the hot strip mill process was taken as the demonstration. It can be concluded from analysis and simulation that contribution rate plot is effective in nonlinear quality-related fault diagnosis, and the combined index is also proper for fault detection.

Due to the limitation of the scope in this paper, the focus are just concentrated on the nonlinearity reflected in hot strip mill process. In the future, more tools will be explored to handle the behaviors such as non-Gaussian, multi-mode or dynamics, etc.

Acknowledgment

This work was supported by the National 973 Project under Grants (2010CB731800, 2009CB32602) and by the NSFC under Grants (61074084, 61074085 and 61273173), PR China. Also thanks the Beijing Key Discipline Development Program (no. XK100080537) and Beijing Natural Science Foundation (4122029).
Appendix A. T-KPLS model

A.1. T-KPLS algorithm

For T-KPLS algorithm, one can refer to Table A1. More details can be found in Peng et al. (under review).

A.2. Calculation for new scores

Based on T-KPLS algorithm above, three kinds of scores for a new sample can be calculated as follows:

\[ t_{n\text{ew}} = Q^T_k (U^T_k T)^{-1} U_k^T K_{n\text{ew}} \in \mathbb{R}^k \]
\[ s_{n\text{ew}} = W_s^T (I - T R P) K (I - T T^T)^{-1} U^T_k \]
\[ a_{n\text{ew}} = \frac{1}{n} \sum_{i=1}^{n} K(x_i, x_{n\text{ew}}) \]

(A.1)

A.3. Details of \( Q \)

The Q-statistic for T-KPLS based detection is expressed as

\[ Q_t = \phi^T(x_{n\text{ew}}) \phi(x_{n\text{ew}}) - 2(K_{n\text{ew}}(I-T T^T)^{-1} U_k^T W_t x_{n\text{ew}}) + t_{n\text{ew}}^T W_t (I-T T^T)^{-1} U_k^T K_{n\text{ew}} \]

where \( \phi^T(x_{n\text{ew}}) \phi(x_{n\text{ew}}) = 1 - (2/n) \sum_{i=1}^{n} 1 \mathbf{1}_{K_{n\text{ew}}(i,i)} + (1/n^2) \sum_{i=1}^{n} \sum_{j=1}^{n} K_{n\text{ew}}(i,j) \).

Appendix B. The derivation of \( \Omega \)

First of all, \( \phi \) is given in the form

\[ \phi = \frac{\phi^T(x_{n\text{ew}}) R C_k (A_{C_k} - Q^T_k A_{C_k} Q^T_k)}{\delta_{p^T}} + \frac{\phi^T(x_{n\text{ew}}) (I - R P^T) (I - P P^T) (I - P R^T)}{\delta_{p^T}} \phi(x_{n\text{ew}}) \]

\[ + \frac{\phi^T(x_{n\text{ew}}) R C_k Q^T_k A_{C_k} Q^T_k R C_k}{\delta_{p^T}} + \frac{-2 R P^T - P P^T + P P^T P R^T + R P^T P R^T + R P^T P R^T - P P^T P R^T R P^T}{\delta_{p^T}} \]

Then substitute expressions of \( R, P, \) and \( P_r \) into Eq. (B.1), in this way \( \Omega \) is obtained as follows:

\[ \Omega = \frac{U^T (C_k U)^{-1} Q^T k (A_{C_k} - Q^T_k A_{C_k} Q^T_k) U^T}{\delta_{p^T}} - 2 \frac{(I-T T^T)^{-1} U_k^T - (I-T T^T)^{-1} U_k^T}{\delta_{p^T}} - \frac{(I-T T^T)^{-1} U_k^T - (I-T T^T)^{-1} U_k^T}{\delta_{p^T}} \]

\[ \]}

\[ + \frac{(W_t^2 (I-T T^T)^{-1} U_k^T - W_t^2 (I-T T^T)^{-1} U_k^T) W_t (I-T T^T)^{-1} U_k^T - W_t (I-T T^T)^{-1} U_k^T W_t (I-T T^T)^{-1} U_k^T}{\delta_{p^T}} \]}

(B.2)

Appendix C. Calculations for contribution rate

Firstly, to obtain \( C(x_{n\text{ew}}, i) \), it is necessary to describe kernel function as follows:

\[ K(V \odot x_{n\text{ew}}, x_i) = \exp \left( -\frac{||V \odot x_{n\text{ew}} - x_i||^2}{c} \right) \]

(C.1)

Then the partial derivative of kernel function with respect to the scale factor is defined as

\[ \frac{\partial K(V \odot x_{n\text{ew}}, x_i)}{\partial v_j} \bigg|_{v = 1} = \frac{2}{c} K(x_{n\text{ew}}, x_i) x_{n\text{ew}, i} \]

(C.2)

The sth element of \( K_{n\text{ew}} \) is

\[ K_{n\text{ew}}(s) = K_{n\text{ew}}(s) - \frac{1}{n} \sum_{i=1}^{n} K(x_i, x_{n\text{ew}}) - \frac{1}{n} \sum_{j=1}^{n} K(x_i, x_{n\text{ew}}) + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) = K_{n\text{ew}}(s) - \frac{1}{n} \sum_{i=1}^{n} K(x_i, x_{n\text{ew}}) + a_i \]

(C.3)

where \( K_{n\text{ew}}(s) = K(x_{n\text{ew}}, x_i) \), \( a_i = (1/n^2) \sum_{j=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) = \frac{1}{n} \sum_{j=1}^{n} K(x_i, x_j) \), while the term of \( a_i \) in the partial derivative process is negligible since it is merely the summation of kernel function of training dataset and not give any contribution to evaluate the effect of \( x_{n\text{ew}} \). According to Eq. (C.2), the partial derivative for Eq. (C.3) is shown as

\[ \frac{\partial K_{n\text{ew}}(s)}{\partial v_j} \bigg|_{v = 1} = -\frac{2}{c} \left( x_{n\text{ew}, i} (x_{n\text{ew}, i} - x_i) K(x_{n\text{ew}, i}) - \sum_{j=1}^{n} x_{n\text{ew}, j} (x_{n\text{ew}, i} - x_j) K(x_{n\text{ew}, j}) \right) \]

(C.4)
Based on above analysis, the contribution rate of \(i\)th variable to \(\varphi_{\text{new}}\) is obtained:

\[
C(\varphi_{\text{new}}|i) = \frac{(K(\varphi_{\text{new}} \otimes \Psi)|_{v=1})}{\delta v_j} = \frac{-2}{n} \sum_{j=1}^{n} \frac{\partial (K(\varphi_{\text{new}}|v))}{\partial v_j} \bigg|_{v=1} + \frac{\partial (K(\varphi_{\text{new}}^T \Omega_{\text{new}}|v))}{\partial v_j} \bigg|_{v=1} = \frac{-2}{n} \sum_{j=1}^{n} \frac{\partial (K(\varphi_{\text{new}}|v))}{\partial v_j} + \frac{\partial \left(\frac{1}{n} \sum_{j=1}^{n} (K_{\text{new}}(p)|v)K_{\text{new}}(q|x_{\text{new}}) - x_{\text{new}})\right)}{\partial v_j} + \frac{\partial (K(\varphi_{\text{new}}^T \Omega_{\text{new}}|v))}{\partial v_j} \bigg|_{v=1} = \frac{-2}{n} \sum_{j=1}^{n} \frac{\partial (K(\varphi_{\text{new}}|v))}{\partial v_j} + \text{trace} \left(\frac{\partial (K(\varphi_{\text{new}}^T \Omega_{\text{new}})|v)}{\partial v_j} \bigg|_{v=1} \right)
\]

(C.6)

References


