

# A quality-based nonlinear fault diagnosis framework focusing on industrial multimode batch processes

**Abstract**—This paper proposes a framework for quality-based fault detection and diagnosis for nonlinear batch processes with multimode operating environment. The framework seeks to address (1) the mode partition problem using a kernel fuzzy C-clustering method, and the optimal cluster number will be guaranteed by a between-within proportion index; (2) the diagnosis problem using a contribution rate method based on an improved kernel partial least squares model, by which better detection and diagnosis performances are provided; (3) the classification of online measurements using a hybrid kernel partial least squares regression and the Bayes inference theory, where the new coming measurement can be correctly assigned to its constituent mode. The whole framework is developed for batch processes, and applied to the hot strip mill rolling process. It is shown using the real industrial data that for faults affecting the thickness and flatness of the strip steel in this process, the detection and diagnosis abilities of the present methods are better compared with the existing methods.

**Index Terms**—Multimode, fault diagnosis, kernel PLS, contribution rate, batch hot strip mill

## NOMENCLATURE

$n$	Total number of data.
$N_i$	Number of data within the $i^{\text{th}}$ mode.
$c$	Number of cluster.
$\mathbf{z}, \mathbf{Z}$	Process measurement vector, matrix.
$\mathbf{y}, \mathbf{Y}$	Quality data vector, matrix.
$\hat{\mathbf{y}}$	Prediction of the quality variable.
$\bar{\mathbf{X}}$	Multibatch process and quality data.
$\mathbf{x}, \mathbf{X}$	Combined process and quality vector, matrix.
$\phi(\cdot)$	Nonlinear projection function.
$\phi, \Phi$	high-dimensional projected vector, matrix.
$K_g$	Gaussian kernel function.
$\mathbf{k}$	Kernel vector.
$\mathbf{K}_{raw}$	Raw kernel matrix.
$\mathbf{K}$	Normalized kernel matrix.
$\mathbf{w}$	Weight vector in KPLS.
$\mathbf{t}$	Score vector of $\mathbf{z}$ .
$\mathbf{T}$	Score matrix of $\mathbf{Z}$ .
$\mathbf{u}$	Score vector of $\mathbf{y}$ .
$\mathbf{U}$	Score matrix of $\mathbf{Y}$ .
$\mathbf{U}$	Weight matrix in clustering methods.
$\mathbf{q}$	Loading vector of $\mathbf{y}$ .
$\mathbf{Q}$	Loading matrix of $\mathbf{Y}$ .
$\mathbf{I}_m$	$m$ -dimensional identity matrix.
$\mathbf{1}_m$	$m$ -dimensional unity matrix.
$p(\cdot)$	Probability of a stochastic event.
$\psi_y$	Regression matrix.

## I. INTRODUCTION

WITH the rapid development of the scale, degree of automation and integration of industrial processes, challenges have emerged for practitioners with problems of frequently switched operating points, different operating batches and strong nonlinearities [1]. The complexity also makes it even harder to guarantee the product quality, the most core profitable indicator for an industrial process. For example in the hot strip mill rolling (HSMR) process, the increasing demands on product quality subject to the requirement on different grades of products and lower energy consumption have made it so complicated that it is involving different kinds of challenges in optimal control and process monitoring areas. Table I shows an overview of products in an HSMR plant, where it can be found the most concerned quality indices by the plant engineers: the final thickness and width. Different from the qualitative modelling of product quality [2], [3], this work use the quantitative one, which is famously quantified as quality variables. According to different uses, grades of strip steels are different and the annual output of them are also different. It is known that different grades of steel underly that the rolling system works in different operating modes, and also well-known that the strip steel are produced coil by coil with each coil representing a batch. Noted that the multimode and multibatch properties in HSMR process are slightly different from those in chemical processes [4]. Another handicap for ensuring the product quality is the strong nonlinearity in HSMR, in this case the traditional linear theory will not be sensible. In recent years, to solve such kinds of problem, quality-related fault detection and diagnosis (QFDD) methods have been raised and proved to be useful for quality control [1], [5], [6], [7], [8].

In general, QFDD methods can be divided into model-based ones and data-driven ones. Of them, the later developed using the process measurements has drawn more attentions over past decades [9], [10]. Its advantage is evident when dealing with some large-scale processes, as the first-principles models can not accurately identifiable for this kind of process. The commonly used data-driven methods are multivariate statistics methods [11], [12], subspace aided data-driven methods [10], machine-learning and signal process-based methods [7], *etc.* However, for the nonlinear processes with multiple operating modes, the traditional approaches, e.g. principal component analysis (PCA) and partial least squares (PLS) that are for linear steady state processes cannot be usable. For these processes, Zhao *et al.* used multiway PCA and PLS approaches [13], [14]. In [14], the multiway PLS was developed using the angles between different operating models, where the angle measures the similarities of two PLS models. In [15], Hwang and Han proposed a hierarchical clustering-based super PCA model for realtime use. Doan and Srinivasan [16] resolved this

TABLE I  
DIFFERENT TYPES OF STRIP STEEL PRODUCTS

Steel grade	Usage description	Thickness (mm)	Width (mm)	Annual Output ( $\times 10^6$ kg)	Coil (Batch)
SS330-SS540	General carbon steel	1.21-12.7	900-1550	700	800
SPHC-SPHE	High-quality carbon steel	1.21-12.7	900-1550	700	600
SAPH310-SAPH400	Alloy steel	1.51-12.7	1050-1550	700	1000
SPA-H	Corrosion resistant steel	1.51-12.7	900-1550	350	500
X42-X60	Oil pipeline steel	5.1-12.7	1050-1550	700	300
Others	—	1.21-15.9	900-1550	350	200

problem by dividing the continuous multimode processes into different stages, and built models for each stage. Since then, a large number of stage division approaches have emerged to promote the development of multimode QFDD. Yu *et al.* combined the finite Gaussian mixture models (FGMM) and Bayesian inference (BI) technique in [4], [17] to identify the constituent modes. By FGMM method, different process modes can be represented using different Gaussian distribution with unique means and covariances. The method was introduced to dynamic processes by Haghani *et al.* [18]. The above mentioned methods are all linear, thus, cannot well address the nonlinear multimode problem. To divide the nonlinear data, this paper would like to introduce a kernel fuzzy C-clustering method (KFCM) [19]. Furthermore, traditional approaches have not well understood how to obtain the optimal mode number. This is still an open topic, but can significantly affect the QFDD results. Thus, in this paper, a validity index called Between-Within Proportion (BWP) which considers the geometric property of the dataset will be introduced to solve this problem [20].

The QFDD methods based on kernel PCA (KPCA) [21] and kernel PLS (KPLS) [22] models were widely applied, in which original process data are transformed into a feature space via a nonlinear mapping, and then a linear model is built in the feature space. With the aid of the kernel trick, the explicit nonlinear mapping function can be avoided. Recently, such kind of models have been extended to batch processes for QFDD [23], [24], [25]. However, KPLS has drawbacks such as low fault detection rate (FDR) and high false alarm rate (FAR) [5]. Thus, some improvements should be derived based on KPLS model to let it more efficient.

Once a quality-related fault is successfully detected, the root cause should be identified as soon as possible [7]. Many fault diagnosis approaches were developed, e.g. discriminant analysis [26], pattern matching using dissimilarity factors [27], contribution plot [28] *etc.* Of them, contribution plot method was broadly used, as it is free of *a priori* process knowledge [29]. Considering KPLS model with the kernel function involved, a contribution rate plot approach was proposed by Peng *et al.* [5] It was also proved that this approach can be promoted to any kernel methods [5].

Therefore, objectives of this paper are:

- to cluster the nonlinear multimode data using KFCM-based method, and optimize the mode number using a new BWP index,
- to develop an improved KPLS model and contribution rate-based diagnosis for nonlinear QFDD,
- to realize the online classification based on KPLS regres-

sion and Bayes inference, and

- to apply the proposed approaches to a real HSMR process and compare the performance between the present and the existing ones.

Two probability distributions are used in this work. Let  $\chi^2(l)$  be the  $\chi^2$ -distribution with  $l$  degrees of freedom;  $F(a, b)$  be the  $F$ -distribution with  $a$  and  $b$  degrees of freedom. Let  $prob(\chi^2 > \chi_{1-\alpha}^2(l)) = \alpha$  represent the probability that  $\chi^2 > \chi_{1-\alpha}^2(l)$  equals  $\alpha$  and  $prob(F > F_{1-\alpha}(a, b)) = \alpha$  represent the probability that  $F > F_{1-\alpha}(a, b)$  equals  $\alpha$ .

## II. METHODOLOGY

In modern industry, emerging techniques have made the processes more integrated, for example, in HSMR process, the same set of equipment can batch-wise produce different grades of strip steels. In this case, the fault diagnosis that designed for guaranteeing qualities of products becomes more and more challenging. Given the multibatch and multimode process and quality data, this part resolve the problems using three methods: (1) KFCM clusters the different modes, (2) the improved KPLS model fulfils the QFDD, and (3) the combination of KPLS regression and Bayes inference makes the whole framework efficiently online applicable.

### A. KFCM-based mode partition and BWP index

This section gives a detailed view on KFCM algorithm. Let  $\bar{\mathbf{X}}$  be the three-dimensional multibatch data. After unfolding it to be two-dimensional [17], the data turn to be the form of  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{k \times n}$  that contain the mixed process data collected from different operating modes. It includes the process measurements  $\mathbf{Z} \in \mathbb{R}^{m \times n}$  and quality measurements  $\mathbf{Y} \in \mathbb{R}^{l \times n}$  in a way of  $\mathbf{X} = \begin{bmatrix} \mathbf{Z} \\ \mathbf{Y} \end{bmatrix}$ . It is assumed in this part that no *a priori* process knowledge is available to partition them, hence, an unsupervised clustering method is needed. Among the exiting methods, the fuzzy C-clustering method (FCM) has been intensively used, as it is a soft fuzzy clustering which uses the membership matrix  $U$  to reduce the false clustering rate (FCR). In sense of FCR, the method is better than K-means clustering (KMC) method, a hard clustering method. Let  $U \in \mathbb{R}^{c \times n}$  and it is constrained by  $\sum_{i=1}^c u_{ij} = 1, \forall j = 1, \dots, n$ , where  $c$  is the pre-specified clustering number. For nonlinear clustering problem, it has been frequently encountered that the process nonlinearity can be linearly approximated after transferring  $\mathbf{x}$  with a function  $\phi(\cdot)$ .  $\phi(\cdot)$  may not have an explicit form or even project  $\mathbf{x}$

onto a infinite-dimensional space by  $\phi(\mathbf{x})$ . Then, the KFCM seeks to minimize the objective function:

$$\begin{aligned} J_{u_{i,j}} &= \sum_{i=1}^c \sum_{j=1}^n u_{i,j}^\delta d_{i,j}^2 \\ \text{s.t. } \sum_{i=1}^c u_{i,j} &= 1, (j = 1, \dots, n) \end{aligned} \quad (1)$$

with  $d_{i,j} = \|\phi(\mathbf{x}_j) - \mathcal{C}_i\|_E$  denoting the Euclid distance between  $\phi(\mathbf{x}_j)$  and the  $i^{\text{th}}$  clustering center, namely  $\mathcal{C}_i$ , in the  $\hat{m}$ -dimensional linear space and  $\delta$  a parameter for tuning the soft degree. Using the Lagrange multiplier method, and taking the derivative with respect to  $U$  and  $\mathcal{C}$  gives

$$\mathcal{C}_i = \frac{\sum_{j=1}^n u_{i,j}^\delta \phi(\mathbf{x}_j)}{\sum_{j=1}^n u_{i,j}^\delta} \quad (2)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{d_{ij}}{d_{kj}} \right)^{2/(\delta-1)}} \quad (3)$$

By iterating (2) and (3), the optimal  $U$  can be resolved until it is converged. In (3),  $d_{i,j}^2$  is calculated in the form of

$$\begin{aligned} d_{i,j}^2 &= \|\phi(\mathbf{x}_j) - \mathcal{C}_i\|_E^2 = (\phi^T(\mathbf{x}_j) - \mathcal{C}_i^T) (\phi(\mathbf{x}_j) - \mathcal{C}_i) \\ &= 1 - 2 \frac{\sum_{k=1}^n u_{i,k}^\delta \phi^T(\mathbf{x}_j) \phi(\mathbf{x}_k)}{\sum_{j=1}^n u_{i,j}^\delta} + \frac{\sum_{j=1}^n \sum_{k=1}^n u_{i,k}^{2\delta} \phi^T(\mathbf{x}_j) \phi(\mathbf{x}_k)}{\left( \sum_{j=1}^n u_{i,j}^\delta \right)^2} \end{aligned} \quad (4)$$

By introducing the kernel trick:  $K(\mathbf{x}_j, \mathbf{x}_k) = \phi^T(\mathbf{x}_j) \phi(\mathbf{x}_k)$ . All the equations are arithmetically solvable. Generally,  $K(\cdot)$  is called a kernel function [21], and the Gaussian version of it, namely  $K_g(x, y) = \exp\left(-\frac{\|x-y\|_E^2}{\sigma^2}\right)$  is widely used [21], where  $\sigma$  is a tunable parameter. After obtaining  $U$ , the data can be partitioned into  $c$  clusters based on entries of it, that is, the  $i^{\text{th}}$  data belongs to the  $j^{\text{th}}$  cluster with the degree of  $u_{i,j}$ .

The clustering number  $c$  shall be accurately determined beforehand, an optimal  $c$  will make the following fault diagnosis more reliable. In this paper a new between-within proportion (BWP) index will be introduced [20]. First defining two kinds of distances, the first one is

$$b(j, i) = \min_{1 \leq k \leq c, k \neq j} \left( \frac{1}{N_k} \sum_{p=1}^{N_k} \|\phi(\mathbf{x}_{(k)p}) - \phi(\mathbf{x}_{(j)i})\|_E^2 \right) \quad (5)$$

which measures the minimum distance between  $j^{\text{th}}$  element in the  $i^{\text{th}}$  cluster ( $\phi(\mathbf{x}_{(j)i})$ ) and the data in other clusters. The other one is

$$w(j, i) = \frac{1}{N_j - 1} \sum_{q=1, q \neq i}^{N_j} \|\phi(\mathbf{x}_{(j)q}) - \phi(\mathbf{x}_{(j)i})\|_E^2 \quad (6)$$

which measures the average distance between  $\phi(\mathbf{x}_{(j)i})$  and other data in the  $j^{\text{th}}$  cluster. Let  $BWP(j, i) = \frac{b(j,i) - w(j,i)}{b(j,i) + w(j,i)}$ . It can be seen when  $w(j, i) \ll b(j, i)$ , namely the clustering result is correct, the BWP index approximates +1. On the contrary, the BWP index approximates -1. Thus, the BWP

index can reflect the clustering validity of the examined method, that is, the better clustering methods have the larger the BWP (approximate +1). Inductively, let  $\overline{BWP}(c) = \frac{1}{n} \sum_{j=1}^c \sum_{i=1}^{N_j} BWP(j, i)$  be the average value of BWP, the optimal  $c_{opt}$  should be the one with the maximum  $\overline{BWP}(c)$ . Noted that  $2 < c < c_{max}$ ,  $c_{max}$  can be defined depending on different processes. Given  $c_{opt}$  and  $U$ ,  $\mathbf{X}$  can be successfully classified into  $\mathbf{X}^{(i)}$  for  $i = 1, \dots, c_{opt}$ .

### B. Improved kernel PLS-based QFDD

For nonlinear QFDD purpose, KPLS-based approaches have many advantages. Compared with the nonlinear methods in [30], where nonlinear optimizations such as gradient descent, and nonlinear curve fitting-based approaches are used, the modelling of KPLS is free of this design and more straightforward to address the nonlinearity. Given the process and quality data within a single operating mode, which are expressed by  $\Phi \in \mathbb{R}^{N \times \hat{m}}$  and  $\mathbf{Y} \in \mathbb{R}^{N \times l}$ , respectively. Let a single process measurement be  $\phi \in \mathbb{R}^{\hat{m}}$ , which is an abbreviation of  $\phi(\mathbf{z})$ , by the convenience of kernel trick,  $\mathbf{K} = \Phi \Phi^T \in \mathbb{R}^{N \times N}$  and  $\mathbf{k} = \Phi \phi \in \mathbb{R}^N$  are satisfied. Actually, they are the normalized value based on the raw mapped data in a way [5]:

$$\begin{aligned} \mathbf{K} &= (\mathbf{I}_N - (1/N) \mathbf{1}_N \mathbf{1}_N^T) \mathbf{K}_{raw} (\mathbf{I}_N - (1/N) \mathbf{1}_N \mathbf{1}_N^T) \\ \mathbf{k} &= (\mathbf{I}_N - (1/N) \mathbf{1}_N \mathbf{1}_N^T) (\mathbf{k} - (1/N) \mathbf{K}_{raw} \mathbf{1}_N) \end{aligned} \quad (7)$$

where  $\mathbf{K}_{raw}$  is directly calculated from the kernel function. Like the PLS model, the nonlinear iterative KPLS model is shown in Table II.

TABLE II  
KPLS ALGORITHM

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

$$\begin{cases} \Phi_{i+1} = \left( \mathbf{I}_N - \mathbf{t}_i (\mathbf{t}_i^T \mathbf{t}_i)^{-1} \mathbf{t}_i^T \right) \Phi_i \\ \mathbf{Y}_{i+1} = \left( \mathbf{I}_N - \mathbf{t}_i (\mathbf{t}_i^T \mathbf{t}_i)^{-1} \mathbf{t}_i^T \right) \mathbf{Y}_i \\ \mathbf{K}_{i+1} = \left( \mathbf{I}_N - \mathbf{t}_i (\mathbf{t}_i^T \mathbf{t}_i)^{-1} \mathbf{t}_i^T \right) \mathbf{K}_i \left( \mathbf{I}_N - \mathbf{t}_i (\mathbf{t}_i^T \mathbf{t}_i)^{-1} \mathbf{t}_i^T \right) \end{cases}$$
  - (5) Set  $i = i + 1$ , loop to step (1), until  $i > A$ .
  - (6) Let  $\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_A] \in \mathbb{R}^{N \times A}$ ,  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_A] \in \mathbb{R}^{N \times A}$ ,  
 $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_A] \in \mathbb{R}^{l \times A}$
- 

After the KPLS model,  $\Phi$  and  $\mathbf{Y}$  can be modeled as

$$\begin{aligned} \Phi &= \hat{\Phi} + \tilde{\Phi} = \Pi_{\mathbf{T}} \Phi + \tilde{\Phi} \\ \mathbf{Y} &= \hat{\mathbf{Y}} + \tilde{\mathbf{Y}} = \Pi_{\mathbf{T}} \mathbf{Y} + \tilde{\mathbf{Y}} \end{aligned} \quad (8)$$

with  $\Pi_{\mathbf{T}} = \mathbf{T} (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T$ .  $\mathbf{T}$  can be directly derived from  $\Phi$  by  $\mathbf{T} = \Phi \Phi^T \mathbf{U} \left( (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{K} \mathbf{U} \right)^{-1} = \mathbf{K} \mathbf{U} \left( (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{K} \mathbf{U} \right)^{-1}$ . The prediction model of KLS

can be shown using the formation:

$$\hat{\mathbf{y}}_{new} = \mathbf{Q}\mathbf{t}_{new} = \mathbf{Q} \underbrace{\left( \mathbf{U}^T \mathbf{K} \mathbf{T} (\mathbf{T}^T \mathbf{T})^{-1} \right)^{-1} \mathbf{U}^T \mathbf{\Phi} \phi(\mathbf{z}_{new})}_{\psi_y^T} \quad (9)$$

Let the projector  $\Pi_y$  equal  $\psi_y(\psi_y^T \psi_y)^{-1} \psi_y^T$ , by projecting  $\phi$  onto  $\Pi_y$ :  $\phi_y = \psi_y(\psi_y^T \psi_y)^{-1} \psi_y^T \phi$ . Then the  $\hat{\mathbf{y}}$ -related part of  $\mathbf{\Phi}$  can be written using

$$\mathbf{\Phi}_{\hat{y}} = \mathbf{\Phi} \psi_y (\psi_y^T \psi_y)^{-1} \psi_y^T \quad (10)$$

Instead, projecting  $\phi$  onto the complementary space of  $\psi_y$  gives  $\phi_{\hat{y}}^\perp = (\mathbf{I}_{\hat{m}} - \psi_y(\psi_y^T \psi_y)^{-1} \psi_y^T) \phi$ . Correspondingly,

$$\mathbf{\Phi}_{\hat{y}}^\perp = \mathbf{\Phi} \left( \mathbf{I}_{\hat{m}} - \psi_y(\psi_y^T \psi_y)^{-1} \psi_y^T \right) \quad (11)$$

Since, this part may include the  $\hat{\mathbf{y}}$  related part [5], a further PCA decomposition on it is needed. To this end, the PCA model is developed below:

- (1) Do an SVD on  $\frac{\mathbf{K}_{\hat{y}}^\perp}{N-1} = \frac{\mathbf{\Phi}_{\hat{y}}^\perp \mathbf{\Phi}_{\hat{y}}^{\perp T}}{N-1}$ ;
- (2) Store  $A_{\hat{y}}^\perp$  eigenvectors as  $\omega$ , and scale them with the eigenvalues [21].
- (3) Let  $\mathbf{\Phi}_{\hat{y}}^\perp = \mathbf{K}_{\hat{y}}^\perp \omega \omega^T \mathbf{\Phi}_{\hat{y}}^\perp$ ,  $\mathbf{\Phi}_{\hat{y}} = \mathbf{\Phi}_{\hat{y}}^\perp - \mathbf{K}_{\hat{y}}^\perp \omega \omega^T \mathbf{\Phi}_{\hat{y}}^\perp$ .

In order to thoroughly detect the quality-related faults, both  $\mathbf{\Phi}_{\hat{y}}$  and  $\mathbf{\Phi}_{\hat{y}}^\perp$  should be both taken into account. For monitoring the first part, it can be observed that  $\mathbf{\Phi}_{\hat{y}} \psi_y = \mathbf{\Phi} \psi_y = \hat{\mathbf{Y}}$ , thus this part can be monitored using a  $T^2$  statistic [10]:

$$T_{\hat{y}}^2 = \phi_{\hat{y}}^T \psi_y \left( \frac{\psi_y^T \mathbf{\Phi}_{\hat{y}} \mathbf{\Phi}_{\hat{y}}^T \psi_y}{N-1} \right)^{-1} \psi_y^T \phi_{\hat{y}} \quad (12)$$

where  $\psi_y^T \mathbf{\Phi}_{\hat{y}} \mathbf{\Phi}_{\hat{y}}^T \psi_y = \hat{\mathbf{Y}}^T \hat{\mathbf{Y}}$ , and

$$\phi_{\hat{y}}^T \psi_y = \phi^T \psi_y = \mathbf{k} \mathbf{U} \left( (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{K} \mathbf{U} \right)^{-1} \quad (13)$$

The threshold is normally derived based on the  $J_{th, T^2} = \frac{l(N^2-l)}{N(N-l)} F_{1-\alpha}(l, N-l)$  [10].

For the second part, since it only consists of very few variations, the SPE statistic could be given:

$$SPE = \|\phi_{\hat{y}}^\perp\|^2 = \phi_{\hat{y}}^{\perp T} \phi_{\hat{y}}^\perp \quad (14)$$

with  $\phi_{\hat{y}}^\perp = \left( \mathbf{I}_{\hat{m}} - \mathbf{\Phi}_{\hat{y}}^\perp \omega \omega^T \mathbf{\Phi}_{\hat{y}}^\perp \right) \left( \mathbf{I}_{\hat{m}} - \psi_y(\psi_y^T \psi_y)^{-1} \psi_y^T \right) \phi$ . The threshold is calculated using  $J_{th, SPE} = g \chi_{1-\alpha}^2(h)$ , where  $g = S/2\mu$ ,  $h = 2\mu/S$ .  $S$  and  $\mu$  are trained with the off-line training data.

For the seek of simplification, a combined statistic shall be developed based on the following approach [5]:

$$\varphi_y = \frac{T_{\hat{y}}^2}{J_{th, T^2}} + \frac{SPE}{J_{th, SPE}} = \frac{\bar{K}_g(z_{new}, z_{new})}{J_{th, SPE}} + \mathbf{k}_{new}^T \Xi \mathbf{k}_{new} \quad (15)$$

where  $\bar{K}_g(\mathbf{z}_{new}, \mathbf{z}_{new}) = 1 - \frac{2}{N} \sum_{i=1}^N K_g(\mathbf{z}_{new}, \mathbf{z}_i) + \sum_{i=1}^N \sum_{j=1}^N K_g(\mathbf{z}_i, \mathbf{z}_j)$  and  $\Xi \in \mathbb{R}^{N \times N}$ , details of it can be derived by integrating (9-13). The new index can be bounded by kernel density

estimation (KDE)-based method with  $J_{th, \varphi}$ , and then the final detection decision logic is

$$\begin{cases} \varphi_y \leq J_{th, \varphi} \Rightarrow \text{Quality is fault-free} \\ \varphi_y > J_{th, \varphi} \Rightarrow \text{Quality is faulty} \end{cases} \quad (16)$$

Regarding the method, it is worth noting: (1) it can deliver higher fault detection performances, as it avoids the quality-orthogonal part in  $\hat{\mathbf{\Phi}}$ , which is the completely quality-related part in KPLS; (2) it further includes the possible quality-related parts in  $\hat{\mathbf{\Phi}}$ , while this part was left by KPLS modelling; (3) it involves a simplified computation process compared with the method in [5]. The method involves an extra PCA model on an  $N \times N$  ( $O(N^3)$ ) matrix than KPLS model, however, in [5], the method includes one extra PCA on an  $l \times l$  matrix and two additional PCA models on  $N \times N$  matrices. It is evident that the new method can save many calculation efforts. Besides, the efficient implementation performance will be shown in Section III.

After a quality-related fault has been detected, it is urgent to find out the root-cause for further corrective activities. The methods developed for KPLS-based diagnosis can be found in [5], [24]. The contribution rate-based approach in [5] has been recently concerned, as it has more clear physical interpretation. the contribution rate of variable  $i$  to the increment of  $\varphi_y$  when a fault occurs can be shown as

$$Cont_{z_{new}}^i = \left| \frac{\partial \varphi_y(\mathbf{z}_{new} \odot \mathbf{v})}{\partial v_i} \Big|_{\mathbf{v}=\mathbf{1}_m} \right| = \left| \frac{1}{J_{th, SPE}} \frac{\partial \bar{K}_g(\mathbf{z}_{new} \odot \mathbf{1}_m, \mathbf{z}_{new} \odot \mathbf{1}_m)}{\partial v_i} + \frac{\partial (\mathbf{k}_{new}^T \Xi \mathbf{k}_{new})}{\partial v_i} \right| \quad (17)$$

where  $\mathbf{z}_{new} \odot \mathbf{v} = [v_1 \mathbf{z}_{new,1}, \dots, v_m \mathbf{z}_{new,m}]^T$ ,  $\varphi_y(\mathbf{z}_{new} \odot \mathbf{v}) \Big|_{\mathbf{v}=\mathbf{1}_m} = \varphi_y(\mathbf{z}_{new})$ .  $v_i$  is a scaling factor that can signify the change of the  $i^{\text{th}}$  variable in the form of  $v_i \mathbf{z}_i$ . The first part of (17) could be calculated as

$$\begin{aligned} \frac{\partial \bar{K}_g(\mathbf{z}_{new} \odot \mathbf{1}_m, \mathbf{z}_{new} \odot \mathbf{1}_m)}{\partial v_i} &= -\frac{2}{N} \sum_{j=1}^N \frac{\partial K_g(\mathbf{z}_{new} \odot \mathbf{v}, \mathbf{z}_j)}{\partial v_i} \Big|_{\mathbf{v}=\mathbf{1}_m} \\ &= -\frac{2}{N} \sum_{j=1}^N \mathbf{z}_{new,i} (\mathbf{z}_{new,i} - \mathbf{z}_{j,i}) K_g(\mathbf{z}_{new}, \mathbf{z}_j) \end{aligned} \quad (18)$$

Then the second part is given using

$$\frac{\partial (\mathbf{k}_{new}^T \Xi \mathbf{k}_{new})}{\partial v_i} = \left( \frac{\partial \mathbf{k}_{new}}{\partial v_i} \right)^T (\Xi + \Xi^T) \mathbf{k}_{new} \quad (19)$$

where the  $i^{\text{th}}$  element of  $\frac{\partial \mathbf{k}_{new}}{\partial v_i} \Big|_{\mathbf{v}=\mathbf{1}_m}$  is shown in the following

$$\begin{aligned} &\frac{\partial \mathbf{k}_{new}(j)}{\partial v_i} \Big|_{\mathbf{v}=\mathbf{1}_m} \\ &= -\frac{2}{\sigma^2} \left\{ \begin{aligned} &\mathbf{z}_{new,i} (\mathbf{z}_{new,i} - \mathbf{z}_{j,i}) K_g(\mathbf{z}_{new}, \mathbf{z}_j) \\ &- \sum_{k=1}^N \mathbf{z}_{new,i} (\mathbf{z}_{new,i} - \mathbf{z}_{k,i}) \bar{K}_g(\mathbf{z}_{new}, \mathbf{z}_k) \end{aligned} \right\} \quad (20) \end{aligned}$$

In the end, the variables that have larger contribution rates are thought as faulty potentials.

### C. A hybrid KPLS regression and Bayes inference for online classification

The proposed QFDD framework will be used for the online measurement  $\mathbf{z}_{new}$ . The first step should identify its hidden mode. This section proposes a Bayes inference-based one. By employing the KPLS-based prediction model:

$$\hat{\mathbf{y}}_{new}^{(i)} = \left( \psi_y^{(i)} \right)^T \phi_{new}^{(i)} \quad (21)$$

The priori probability of  $\mathbf{z}_{new}$  belonging to the  $i^{\text{th}}$  class can be defined as

$$\begin{aligned} & \text{prob}(\mathbf{z}_{new} | i) = \\ & \text{prob} \left( \left\| \hat{\mathbf{y}}_{new}^{(i)} - \mathbf{y}_j^{(i)} \right\|_E \leq \left\| \hat{\mathbf{y}}_j^{(i)} - \mathbf{y}_j^{(i)} \right\|_E \mid \forall j=1, \dots, N_i \right) \end{aligned} \quad (22)$$

where  $\mathbf{y}_j^{(i)}$  is the  $j^{\text{th}}$  actual measurement in  $i^{\text{th}}$  class, alternatively  $\hat{\mathbf{y}}_j^{(i)}$  stands for the predicted value.  $N_i$  is the training data number of  $i^{\text{th}}$  mode with  $\sum_{i=1}^c N_i = n$ . By bayes' theorem, the posterior probability that  $\mathbf{z}_{new}$  is within the  $i^{\text{th}}$  mode is

$$\text{prob}(i | \mathbf{z}_{new}) = \frac{\text{prob}(\mathbf{z}_{new} | i) \text{prob}(i)}{\sum_{i=1}^c \text{prob}(\mathbf{z}_{new} | i) \text{prob}(i)} \quad (23)$$

where the priori probability of the  $i^{\text{th}}$  cluster could be given by its proportion in the training data, or by the scheduled output volumes of this class. In the end,  $\mathbf{z}_{new}$  can be determined by

$$i^{opt} = \arg \max (\text{prob}(i | \mathbf{z}_{new})) \quad (24)$$

### D. Summary of the proposed methods

Offline use of the framework consists of:

- (I) Unfold the data  $\bar{\mathbf{X}}$  to  $\mathbf{X}$ , and identify the  $c$  modes using KFCM-based method and BWP index,
- (II) Build the improved KPLS model using  $\mathbf{X}$ , and train the regression model (9), detection model (15), and diagnosis model (17).

Online use consists of:

- (I) Collect  $\mathbf{z}_{new}$  online,
- (II) Classify it into its mode using (24),
- (III) Fault detection using (15), judge logic (16), if a fault occurs, turn to step (IV), else turn to step (I),
- (IV) Root-cause diagnosis using (17), and loop to step (I).

## III. VERIFICATION STUDY: APPLICATION TO HSMR PROCESS

This part first presents an overview of HSMR process, and then shows the application and comparison results of the proposed framework.

### A. Introduction to HSMR process

The HSMR process is a complex system encountered in iron and steel industries. It can decrease the thickness of hot strip steel to the desired thickness, and keep the high-quality requirements for the flatness and width of the steel. In general, the HSMR process consists of six subsections:

reheating furnace, rough mill, transfer table and crop shear, finishing mill, run-out table cooling, and coiler. A generic streamline of HSMR can be found in [5]. The incoming strip steel is first reheated in the reheating furnace to reach the required temperature, and, then, in the rough mill section, it is roughly shaped to the desired thickness and width. After transported by the transfer table, the strip steel will arrive at the finishing mill section, where it will be accurately milled towards the preset width and thickness and fulfill the expected flatness. Then, the run-out table cooling section allows the strip to cool to the desired temperature, which allows the steel being of good mechanical property. The detailed descriptions of this process can be found in Peng *et al.* [5] and Ding *et al.* [1]. It can be observed that for the HSMR process, the four key quality variables are thickness, width, flatness and temperature, of which the first three are primarily determined by the finishing mill rolling process (FMRP). Therefore, the focus of this example will be on analysing and understanding the fault diagnosis issue in the finishing mill, as well, the thickness and flatness are selected as the basis of the work.

There are seven groups of stands in the FMRP. As shown in Fig. 1, each group of stand has four rolls: two rolls located in the middle work directly on the strip steel, while the other two rolls support the working ones. Normally, before a strip arrives at the stand, the rolling force directly imposed on the upper supporting roll is computed based on the desired thickness reduction rate and the weight of the upper supporting roll. As well, the bending force that mainly affects the flatness and can also affect the thickness is set beforehand using some empirical equations [5]. It is noted that the deformation of the thickness is affected not only by the rolling and bending force but also by the temperature, rolling speed, and also some other physical properties that depend on the specific strip steel. Physically speaking, besides the bending force, factors affecting flatness will be more complicated than thickness. Thus, it is hard to build precise, first principles models between the forces and output thickness and flatness for a single stand. In the overall FMRP system, the stands do not work individually, but are coupled with each other by different control methods. For example, in the 7<sup>th</sup> stand, the thickness is compared with the desired value and the difference can be fed back to adjust the rolling force in that or previous stands. It is noted that the thickness cannot be measured between two stands, instead the gap measurements between two working rolls can be measured by the height of the upper working roll. Due to the rebounding phenomenon, the thickness is approximately equal to the gap subtract the impact of the roll's stiffness. However, since the stiffness is hard to precisely calculate, it is impossible to adjust the downstream stands based on the upstream thickness. The thickness can only be measured using an X-ray device located at a distance from the stands, thus, it will cause time delay in the feedback control system. To be worse, the flatness cannot be online measured, instead, it has to be offline analyzed based on the signals from the CCD camera. The measurable variables and quality variables concerned in this example are shown in Table III. Note that the flatness has the unit of I, which is only a symbolic without any physical meaning. This study uses  $\mathbf{z} = [\mathbf{z}_{for,1} \sim \mathbf{z}_{for,7}, \mathbf{z}_{ben,2} \sim \mathbf{z}_{ben,7}, \mathbf{z}_{gap,1} \sim \mathbf{z}_{gap,7}] \in \mathbb{R}^{20}$

and  $\mathbf{y} = \mathbf{y}_{fla}$  or  $\mathbf{y}_{thi}$ .

*Remark 1:* Corresponding to the chemical process, the multimode and multibatch properties in hot rolling processes are slightly different. It is clear in chemical process that the process may work in different modes which are always referred to the production rates to fulfil the market demands e.g. Tennessee Eastman process [10]. However, the rolling process, multimode typically refers to different specifications of steel. It is a profitable reason to produce different steels in the same streamline. In chemistry, polymerization sector in special, batch process is very appealing to save economic costs and decrease the control complexity. While in hot rolling, a batch represents a coil. The strip steel is produced coil by coil for convenient shipment.

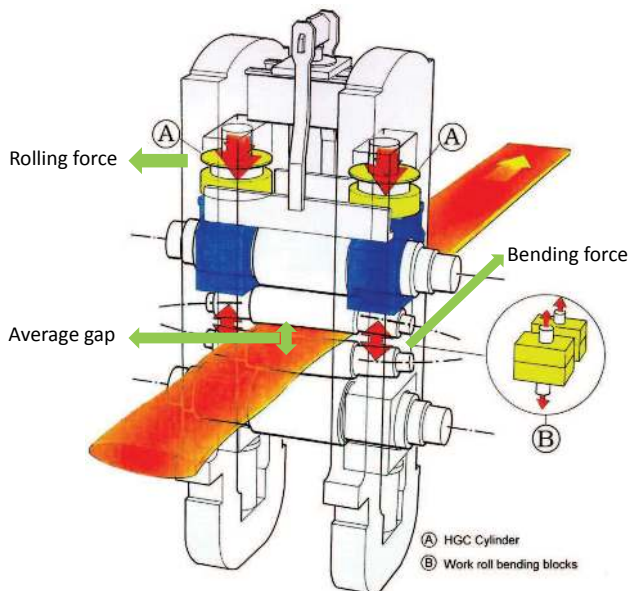


Fig. 1. The structure of a finishing mill stand

### B. Off-line training phase of the framework

FMRP is a typical batch process, where the strip steels are produced coil by coil with the length varying from 1000 m to 1500 m. As well mentioned in Table I, different grades of products underly that there are different operating modes for the system. Based on the two features, Fig. 2 shows a schematic of the FMRP data. It can be observed that the data set is three-dimensional with time axis, batch axis and variable axis. The different batches also contain different operating modes which are distinguished by distinct colors. To conveniently analyze the data, it should be unfolded to be two-dimensional in a variable-wise way. The method is shown in Fig. 2. After that, the resulting data matrix  $\mathbf{X}$  can be modeled. First of all, the KFCM-based mode partition method is examined. Consider the flatness as the quality variable, four modes of data are collected from the plant historian with each mode having eight batches. Each run of batch contains 500 samples, thus,  $\mathbf{X} \in \mathbb{R}^{21 \times 16000}$  is developed. Fig. 3 shows the

result of BMP index obtained from KFCM method. It can be seen that the proposed method gives a consistent clustering number without *a priori* mode information. By contrast, the method using KMC and BMP index cannot show the correct optimal clustering number. It is worth noting that the proposed method assumes the case that parameters:  $\delta$ ,  $\sigma^2$ ,  $c_{max}$  have been appropriately selected beforehand. Since the major focus of this paper is not here, thus, this process is omitted and one can refer to [5] for details. Finally,  $\mathbf{X}^{(j)}$ , for  $j = 1, \dots, 4$  that corresponds to different modes can be formed by comparing the entries in  $U \in \mathbb{R}^{4 \times 16000}$ .

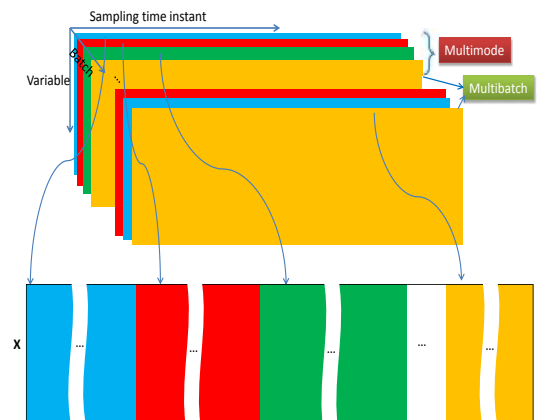


Fig. 2. The property of the HSMR process data

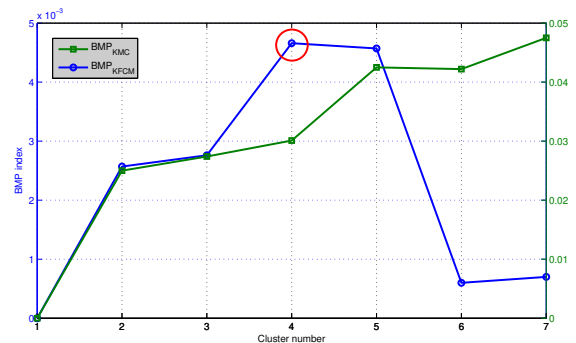


Fig. 3. Classification signal for the fault scenario 1 data

Since both quality variables cannot be immediately measured, the timely fault detection and diagnosis for them have to use the process measurement  $\mathbf{z}$ . In this case, the improved KPLS model should be developed. Regarding each  $\mathbf{X}^{(i)}$ , the detection model (15) and diagnosis model (17) can be established.

For the thickness-related example, the same procedures are repeated. Two modes with thickness equaling 2.7 mm and 3.95 mm are used. 10 batch runs of them are attained from historian such that  $\mathbf{X} \in \mathbb{R}^{21 \times 10000}$ . After the clustering phase, two models are obtained. Then, the fault detection and diagnosis models (15) and (17) are available for online use.

### C. Online application to FMRP

Two different faulty scenarios will be considered:

TABLE III  
OUTLINE OF VARIABLES MEASURED IN FMRP

Variable	Description	Unit	Notes
<b>Process variable</b>			
$\mathbf{z}_{gap,1} \sim \mathbf{z}_{gap,7}$	Average gap measurement in the $i^{\text{th}}$ stand, $i = 1, \dots, 7$	mm	Sampling interval: 0.01s (changeable)
$\mathbf{z}_{for,1} \sim \mathbf{z}_{for,7}$	The total force imposed on the $i^{\text{th}}$ stand, $i = 1, \dots, 7$	MN	Sampling interval: 0.01s (changeable)
$\mathbf{z}_{ben,2} \sim \mathbf{z}_{ben,7}$	The bending force in the working roll of the $i^{\text{th}}$ stand, $i = 2, \dots, 7$	MN	Sampling interval: 0.01s (changeable)
<b>Quality variable</b>			
$y_{fla}$	The flatness of the strip in the exit of FMRP	I	Sampling interval: 0.01s (changeable) with delay (fixed)
$y_{thi}$	The thickness of strip steels in exit of FMRP	mm	Online unavailable

- Fault Scenario 1: For  $1 \leq i \leq 1400$ , the process operates in mode 2 without fault; for  $1401 < i \leq 2100$ , the process runs within mode 1 with fault occurred from the 1601<sup>st</sup> to the 1800<sup>th</sup> sample; for  $2101 < i \leq 3500$ , the process runs within mode 3 without fault.
- Fault Scenario 2: For  $1 \leq i \leq 3000$ , the process operates under mode 1 with fault from the 20<sup>th</sup> s.

Scenario 1 is a flatness-related fault, it occurred when the gap sensor in the 4<sup>th</sup> stand malfunctioned. This fault will directly affect the gap and bending force measurements in the downstream stands, then, until the flatness value. Using the developed model, first of all, Fig. 4 shows the online classification results. It can be observed from which that the proposed method has successfully classified all the data to their constituent modes.

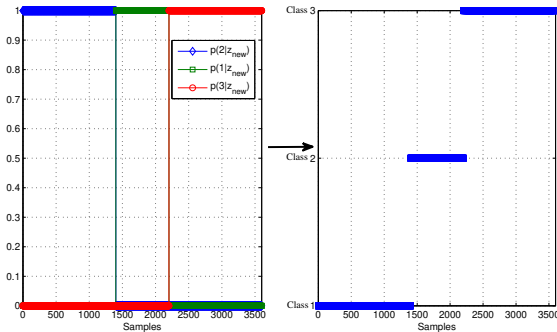


Fig. 4. Classification signal for the fault scenario 1 data

Fig. 5 shows the detection results of Scenario 1. It can be seen that  $\varphi_y$  detected this fault at around the 1600<sup>th</sup> sample. During the fault, the method kept a high detection rate until the fault vanished. The results can also be verified by the real flatness measurements, which is shown in the subfigure of Fig. 5. This can prove that the method can correctly reflect the status of the quality variable even there is no accurate quality measurement. As well-known that PLS, KPLS can also resolve this problem, these three methods were compared in terms of the FDR using Scenario 1. Table IV shows the results. It can be seen that improved KPLS behaves better than PLS and KPLS. Fig. 6 shows the contribution rate-based diagnosis results from the 1500<sup>th</sup> to the 1700<sup>th</sup> samples. It can be observed from around the 1600<sup>th</sup> sample that the contribution rate values of some variables started to increase. The 18<sup>th</sup> variable was particular, as it had the largest contribution rate. The result is consistent with the description of this fault. It is noted that the other variables also had large contribution rates, e.g. the 7<sup>th</sup>

(total force in 7<sup>th</sup> stand), the 17<sup>th</sup> (gap in 4<sup>th</sup> stand). This can be understood by the fault smearing effect.

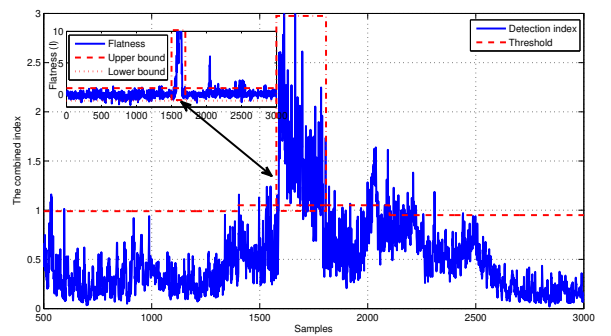


Fig. 5. Detection signal for flatness-related fault with improved KPLS-based method

TABLE IV  
COMPARISONS OF THE PROPOSED METHOD, PLS AND KPLS

Scenario	Performance type	PLS	KPLS	Improved KPLS
1	FDR <sup>1</sup>	0.8021	0.7564	0.9085
2	FDR	0.8852	0.9081	0.9952

<sup>1</sup> Refer to [8] for the calculation formula of FDR.

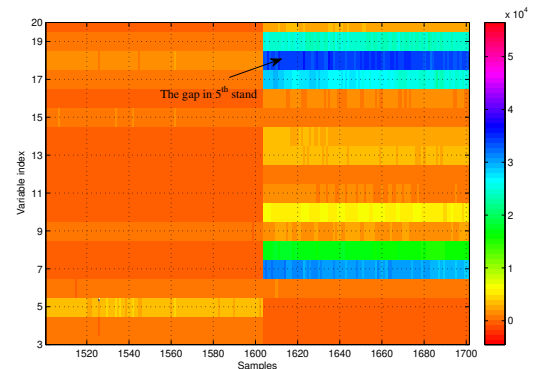


Fig. 6. Contribution rate-based diagnosis signal for flatness-related fault

The fault in Scenario 2 was defined as the malfunction of the gap control loop in the 4<sup>th</sup> stand. It directly affected the gap measurement in this stand. Typically, measurements in the downstream stand will echo this change. Fig. 7 shows the detection results for Scenario 2. It can be seen that  $\varphi_y$  detected this fault at around the 2000<sup>th</sup> sample, which behaves efficiently, as the fault occurred at this time instant. In addition,

the results are consistent with the real thickness measurements as displayed in the subfigure of Fig. 7. It is also worth noting that the thickness data become changed at around the 2350<sup>th</sup> sample, which also points out that the proposed method can identify faults largely before they are even noticed. Table IV provides the detection rates delivered by PLS, KPLS and improved KPLS, where it verifies the excellence of the proposed method. Fig. 8 demonstrates the diagnosis results of this fault. It was shown that variable 5 (total force in the 5<sup>th</sup> stand) and 17 (gap in the 4<sup>th</sup> stand) had the largest contribution rates. The results are correct consistently with the definition of this fault.

Thus, it can be seen that the proposed nonlinear framework can effectively detect and diagnose faults before they are even noticed in the quality variables. This allows for efficient and prompting resolution of any potential problems. Finally, regarding the realtime implementation of the framework, it should be noted:

*Remark 2:* Most plant engineers concern very much about the calculation efficiency of the proposed method due to the limited computing resources. The present approach is computation-intensive, and it is largely attributed to the contribution rate-based fault diagnosis step. As shown in Section 2-B, for each new coming process sample, the diagnosis step involves  $m$  repeated calculations of the contribution rate for each sample. The realtime operation of the model is time-consuming when using the laboratory-sized computer. According the test using the computer configured with Intel Core i7-3770 CPU, 8 GB memory, the overall operating time is larger than the real operating time (3 min). However, for the plant-sized computer, the processor is significantly high-efficiency, which make it possible for realtime implementation of the method.

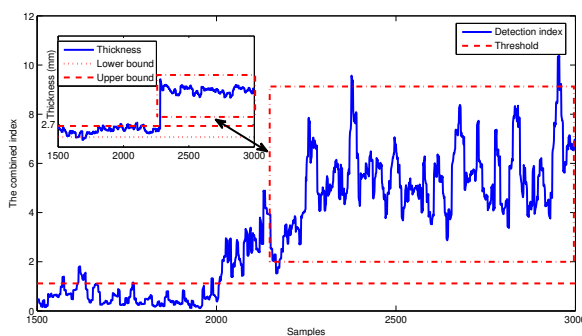


Fig. 7. Detection signal for thickness-related fault with improved KPLS-based method

#### IV. CONCLUSIONS

This paper proposed the framework for QFDD with application to batch multimode processes. The framework consists three relevant methods: multimode clustering, a nonlinear fault diagnosis method, and online classification of the new measurement. The training data from different operating modes were clustered using the nonlinear KFCM-based method, and a BWP index was developed for determining the optimal mode

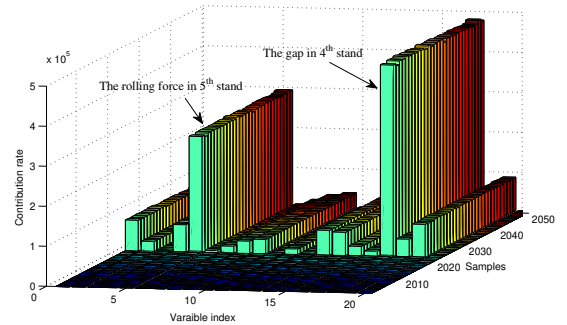


Fig. 8. Contribution rate-based diagnosis signal for thickness-related fault

number. The KPLS method was improved by considering a higher FDR, and based on it, the contribution rate method was used for fault diagnosis. The online data was classified using the abilities of KPLS regression and Bayes inference.

The proposed methods were applied to a batch HSMR process to diagnose faults that affect the product's thickness and flatness. It was seen that the framework can show accurate clustering results, higher detection and precise diagnosis performance.

Future work considers topics with dynamics and non-Gaussian dataset in batch multimode processes to achieve optimal operating performance.

#### REFERENCES

- [1] S. X. Ding, S. Yin, K. X. Peng, H. Y. Hao, and B. Shen, "A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip mill," *IEEE Trans. Ind. Inf.*, vol. 9, no. 4, pp. 2239–2247, 2013.
- [2] J. Zhang, P. D. Roberts, "A hierarchical structure for on-Line process fault diagnosis based on deep qualitative modelling," Springer-Verlag London 2000.
- [3] D.G. Bobrow, "Qualitative reasoning about physical systems," Massachusetts Institute of Technology Cambridge, MA, US, 1985.
- [4] J. Yu, S. J. Qin, "Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models," *AIChE J.*, vol.54, no.7, pp. 1811–1829, 2008.
- [5] K. X. Peng, K. Zhang, G. Li, and D. H. Zhou, "Contribution rate plot for nonlinear quality-related fault diagnosis with application to the hot strip mill process," *Contr. Eng. Pract.*, vol. 21, no. 4, pp. 360–369, 2013.
- [6] S. Yin, H. Luo, and S. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 2402–2411, 2014.
- [7] Z. Gao, C. Cecati, S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques part I: fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, 2015.
- [8] S. Yin, S. X. Ding, A. Haghani, H. Y. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *J. Process Contr.*, vol. 22, no. 9, pp. 1567–1581, 2012.
- [9] G. Li, S. J. Qin, and D. H. Zhou, "A new method of dynamic latent variable modelling for process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6438–6445, 2014.
- [10] S. X. Ding, *Data-driven design of fault diagnosis and fault-tolerant control systems*, London: Springer-Verlag, 2014.
- [11] E. B. Martin, A. J. Morris, "Process Monitoring and Fault Detection Using Multivariate SPC," Springer-Verlag London 2000.
- [12] Y. Yang, Q. Yu, and S. Qin, "Fault Identification for Industrial Process Based on KPCA-SSVM," In *Lecture Notes in Electrical Engineering 125*, Editors: Z. Qian, L. Cao, W. Su, T. W, and H. Yang, Springer-Verlag Berlin Heidelberg 2012.



- [13] S. J. Zhao, J. Zhang, Y. M. Xu, "Monitoring of processes with multiple operating modes through multiple principal component analysis models," *Ind. Eng. Chem. Res.*, vol. 43, no.22, pp. 7025-7035, 2004.
- [14] S. J. Zhao, J. Zhang, Y. M. Xu, "Performance monitoring of processes with multiple operating modes through multiple PLS models," *J. Process Contr.*, vol.16, no.7, pp. 763-772, 2006.
- [15] D. H. Hwang, C. Han, "Real-time monitoring for a process with multiple operating modes," *Control Eng. Pract.*, vol.7, pp. 891-902, 1999.
- [16] X-T Doan, R. Srinivasan, "Online monitoring of multi-phase batch processes using phase-based multivariate statistical process control," *Comput. Chem Eng.*, vol.32, no.1-2, pp. 230-243, 2008.
- [17] J. Yu, S. J. Qin, "Multiway Gaussian Mixture Model Based Multiphase Batch Process Monitoring," *Ind. Eng. Chem. Res.*, vol. 48, no.18, pp. 8585-8594, 2009.
- [18] A. Haghani, T. Jeinsch, S. X. Ding, "Quality-related fault detection in industrial multimode dynamic processes," *IEEE Trans. Ind. Electron.* vol. 61, no. 11, pp. 6446-6453, 2014.
- [19] B. Biswal, P. K. Dash, B. K. Panigrahi, "Power Quality Disturbance Classification Using Fuzzy C-Means Algorithm and Adaptive Particle Swarm Optimization," *IEEE Trans. Ind. Electron.*, vol. 56, no. 1, pp. 212-220, 2009.
- [20] S. B. Zhou, Z. Y. Xu, X. Q. Tang, "Method for determining optimal number of clusters in K-means clustering algorithm," *J. Comput. Appl.*, vol.30, no.8, pp. 1995-1998, 2010.
- [21] S. W. Choi, J. H. Park, I. B. Lee, "Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis," *Comput. Chem. Eng.*, vol.28, no.8, pp. 1377-1387, 2004.
- [22] R. Rosipal, L. J. Trejo, "Kernel partial least squares regression in reproducing kernel Hilbert space," *J. Mach. Learn. Res.*, vol.2, pp. 97-123, 2001
- [23] J. M. Lee, C. K. Yoo, S. W. Choi, P. A. Vanrolleghem, I. B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chem. Eng. Sci.*, vol.59, pp. 223-234, 2004.
- [24] J. H. Cho, J. M. Lee, S. W. Choi, D. W. Lee, I. B. Lee, "Fault identification for process monitoring using kernel principal component analysis," *Chem. Eng. Sci.*, vol.60, pp. 279-288, 2005.
- [25] M. X. Jia, F. Chu, F. L. Wang, W. Wang, "On-line batch process monitoring using batch dynamic kernel principal component analysis," *Chemometr. Intell. Lab.*, vol.101, pp. 110-122, 2010.
- [26] A. Raich, A. Cinar, "Statistical process monitoring and disturbance diagnosis in multivariable continuous processes," *AIChE J.*, vol.42, no.4, pp. 995-1009, 1996.
- [27] M. Kano, S. Hasebe, I. Hashimoto, H. Ohno, "Statistical process monitoring based on dissimilarity of process data," *AIChE J.*, vol.48, no.6, pp. 1231-1240, 2002.
- [28] J. A. Westerhuis, S. P. Gurden, A. K. Smilde, "Generalized contribution plots in multivariate statistical process monitoring," *Chemometr. Intell. Lab.*, vol.51, no.1, pp. 95-114, 2000.
- [29] C. Alcalá, S. J. Qin, "Reconstruction-based contribution for process monitoring," *Automatica*, vol.45, no.7, pp. 1593-1600, 2009.
- [30] R. Rosipal, "Nonlinear Partial Least Squares An Overview," *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques* Chap. 9, pp.169-189, 2011.