Proceedings of the 2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Auckland, New Zealand, July 9-12, 2018

Improved Weighted PLS for Quality-Relevant Fault Monitoring Based on Inner Matrix Similarity

Xiwei Bai[1,2], Xuelei Wang[1], Jie Tan[1], Wei Sun[3], Zhiyong Zhang[3], Zhonghao Zhang[1]

Abstract—**Monitoring the influence of fault towards the product quality is of great importance to modern manufacturing enterprise. Traditional projection to latent structures (PLS) method as well as its variants still face many problems. In this paper, a new improved weighted PLS (IWPLS) is proposed to utilize the local information of the process data, handle noises and build regression models with better generalization capability. The objective function of IWPLS is weighted through calculating the similarity between the target inner matrix (IM) and the other inner matrices (IMs). Two types of weight matrices are given for different process data set. The IWPLS-based monitoring scheme is developed with additional restrains and decomposition operation. A designed numerical experiments and Tennessee Eastman Process (TEP) are employed to evaluate the validity of the proposed method.**

I. Introduction

The rigorous standards of product quality in complex industrial process have prompted the advancement of quality-relevant fault monitoring technique[1]. Multivariate statistical process monitoring (MSPM) methods are deeply delved and widely used in chemical, steel, metallurgy, etc. process due to their low computation complexity and less dependence on process knowledge[2]–[5]. One basic method, partial least squares or projection to latent structures(PLS), with two basic statistics, Hoteling's T^2 and squared prediction error(SPE), are used to detect and differentiate the quality-relevant and irrelevant faults. PLS decomposes the process variable space into the residual subspace and the principle subspace, which reflects the underlying relationship between process and quality variables. Unfortunately, the performed decomposition is oblique and thus leads to inaccurate monitoring results[6]. Total projection to latent structures (TPLS) proposed by Zhou[1] solves this problem by further decomposing the principle and residual subspace into two orthogonal parts each. However, TPLS suffers in heavy computation and redundant subspaces[7]. To overcome these drawbacks, Yin[6] presented a simplified algorithm called improved projection to latent structures(IPLS), which performs orthogonal decomposition on process variable set directly with projection matrices obtained from the coefficient matrix of the PLS regression model.

Same problem exists in all above PLS-based methods, namely, while the global structures of the data set are wildly researched, the local structures are receiving little attention. Fortunately, the development of manifold learning offer new ideas for us to take local structures into account. Methods like Laplacian eigenmaps(LE)[8] and locally preserving projection (LPP)[9] perform dimensionality reduction while maintaining the local structures of the original data set. In simple terms, LPP is the linear version of LE. It utilizes the similarity information between every two samples to optimally preserve neighborhood structures by solving a generalized eigenvalue problem. As pointed out by He[9], LPP is capable to reduce the influence of noise and outliers effectively. It is deemed an alternative method to principle component analysis (PCA), therefore can be employed for process monitoring. Researchers proposed LGPCA[10], GLPP[11], QGLPLS[12], LPPLS[13] etc. in recent years, trying to find a way to integrate the merits of PCA/LPP or PLS/LPP. For the purpose of quality-relevant fault monitoring, QGLPLS reaches a compromise between maximizing the covariance of projections in PLS and preserving the local structures of process and quality data set in LPP; LPPLS merges LPP with PLS, which gives the model better properties of preserving local structure and dealing with nonlinear systems, by substituting LPP for the role of PCA.

In this paper, we present a novel PLS-based method called improved weighted PLS (IWPLS) inspired by both PLS and LPP. Different from QGLPLS and LPPLS, IWPLS uses the local similarity information between every two inner matrices instead of every two measurements to form the weight matrix for the objective function. This trick utilizes the local structures in both process and quality variable space. To improve the performance of IWPLS-based fault monitoring method, l_1 -norm sparseness constraints are added in to enhance the robustness of the IWPLS regression model. The coefficient matrix of the model is then used to implement orthogonal decomposition on process variable space. Numerical experiments and data from Tennessee Eastman chemical process simulations prove that the proposed method can prevent the negative influences of inaccurate

*Research supported by National Natural Science Foundation of China (U1701262) and New Intelligent Manufacturing Mode Application Project of Chinese Ministry of Industry and Information Technology(2016ZXFM06005).

X. W. Bai is with the Institute of Automation, Chinese Academy of Sciences, Beijing, 100190 and University of Chinese Academy of Sciences, Beijing, 100049 China(e-mail: xiwei_bai@163.com).

X. L. Wang is with the Institute of Automation, Chinese Academy of Sciences, Beijing, 100190 China (e-mail: xuelei.wang@ia.ac.cn).

*J. Tan is with the Institute of Automation, Chinese Academy of Sciences, Beijing, 100190 China (corresponding author to provide e-mail: jie.tan@ia.ac.cn).

W. Sun and Z. Y. Zhang are with Zhejiang Tianneng Energy Technology Co., Ltd., Changxing , 313100 China (e-mail: xiwei_bai@163.com).

Z. Zhang is with the Institute of Automation, Chinese Academy of Sciences, Beijing, 100190 China (e-mail: 1181984417@qq.com).

978-1-5386-1854-7/18/$31.00 ©2018 IEEE

measurements on process modelling, thus results in more accurate quality-relevant fault monitoring.

The reminder of this paper is organized as follow: Section 2, a brief introduction of PLS and LPP. Section 3, the key idea and the complete implementation of IWPLS. Section 4, a constructed numerical experiment for IWPLS evaluation. Section 6, a case study on Tennessee Eastman Process. Section 7, conclusion.

II. PLS AND LPP BASICS

For a typical industrial process with the input process variable set $X \in \Re^{N\times n}$ and the output quality variable set $Y \in \Re^{N\times m}$, where N refers to the number of samples and n, m refer to the number of input and output variables, PLS decomposes the scaled and mean-centered X and Y into two subspaces separately in terms of A latent variables:

$$\begin{aligned} X &= \hat{X} + \tilde{X} = TP^T + E \\ Y &= \hat{Y} + \tilde{Y} = TQ^T + F \end{aligned} \tag{1}$$

where $T = [t_1, t_2, ..., t_A] \in \Re^{N\times A}$ is the score matrix, $P = [p_1, p_2, ..., p_A] \in \Re^{n\times A}$ and E are the loading matrix and residual of X, $Q = [q_1, q_2, ..., q_A] \in \Re^{m\times A}$ and F are the loading matrix and residual of Y. In the original NIPALS algorithm[14], a pair of weight vectors w_i and v_i are calculated iteratively to maximize to following object function:

$$\begin{aligned} &J_{PLS} = \mathrm{t}_i^T u_i = w_i^T X_i^T Y_i v_i \\ &s.t. \|w_i\| = \|v_i\| = 1 \end{aligned} \tag{2}$$

where $t_i = X_i w_i$, X_i and Y_i are deflated after each iteration. The loading matrices P and Q are calculated by original least squares(OLS) with $p_i = \left(t_i^T t_i\right)^{-1} X_i^T t_i$ and $q_i = \left(t_i^T t_i\right)^{-1} Y_i^T t_i$. To bridge the gap between T and the original X, a new weight matrix $R = [r_1, r_2, ..., r_A] \in \Re^{n\times A}$ is introduced to establish the relation $T = XR$ with :

$$r_i = \prod_{j=1}^{i-1} \left(I_n - w_j p_j^T\right) w_i \tag{3}$$

Take X as an example, LPP finds a optimal projection matrix $T_p \in \Re^{n\times A}$ by minimizing the following object function:

$$J_{LPP} = \frac{1}{2}\sum_{i,j=1}^{N} W_{ij} \left\| x_i T_p - x_j T_p \right\|^2 \tag{4}$$

where $W \in \Re^{N\times N}$ is a weight matrix with $W_{ij} = e^{-\|x_i - x_j\|^2/t}$. W gives the similar $x_i \mid x_j$ pair a large weight and the dissimilar pair a small weight to preserve the original local structures of X. This optimization problem is reduced into the following form to avoid the degenerate solution:

$$\begin{aligned} &\min t_p^T X^T L X t_p \\ &s.t.\ t_p^T X^T D X t_p = 1 \end{aligned} \tag{5}$$

where $L = D - W$ and D is a diagonal matrix with $D_{ii} = \sum_{j=1}^{N} x_{ij}$. t_p is given by the eigenvectors of the A smallest eigenvalue of the equivalent generalized eigenvalue problem $X^T L X t_p = \lambda X^T D X t_p$.

III. IMPROVED WEIGHTED PLS

A new quality-relevant fault monitoring method with a name of improved weighted PLS(IWPLS) is proposed to make use of the local structures of both process and quality data. In this section, the key idea of IWPLS is illustrated firstly, then the complete implementation of the method for quality-relevant fault monitoring is discussed in details.

A. Key ideas

Essentially, IWPLS is a weighted version of traditional PLS. Similar to PLS, the weight vector pair in IWPLS, w_i and v_i, is optimally selected to maximize the weighted covariance of projections between X and Y. The object function is given as follow:

$$\begin{aligned} &J_{IWPLS} = \mathrm{t}_i^T S u_i = w_i^T X_i^T S Y_i v_i \\ &s.t. \|w_i\| = \|v_i\| = 1 \end{aligned} \tag{6}$$

$S \in \Re^{N\times N}$ is the weight matrix. Obviously, PLS is a simplified IWPLS, namely, S is an identity matrix in PLS. Inspired by the weight matrix W in LPP, S is designed to utilize the local structures of the original X and Y.

To make this statement more explanatory, we consider the vector form of the matrix $X^T SY$ (the subscription i is ignored here) in the object function. For $S = I_{N\times N}$, $X^T SY = X^T Y = \sum_{i=1}^{N} x_i^T y_i$. It is easy to observe that the $X^T Y$ can be represented as the sum of a series of $x_i^T y_i$. Therefore, $x_i^T y_i$ is named inner matrix(IM for short). For a more common S, we have $X^T SY = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i^T S_{ij} y_j$, where $X^T SY$ is the sum of weighted inner matrices(IMs).

It is meaningful to do research on IM because it carries local information of both process and quality measurements simultaneously. Evidently, IWPLS not only considers the IM formulated by the original corresponding $x_i \mid y_i$ pair, but also the IM by $x_i \mid y_j (i \neq j)$ from different measurements. In real industry process, data are usually contaminated by disturbance, noise and measurement error. The uncertainty of $x_i \mid y_i$ pair might impose huge negative effects on the developed model. However, although x_i and y_j seem irrelevant, with proper weighting and average operation, the variety of $x_i \mid y_j$ pair could reduce the influence of uncertain data. Accordingly, we present two variations of the weight matrix S :

$$S_{ij}^{(1)} = e^{\frac{-\left\|x_i y_j - x_i y_i\right\|_2}{t}}, i = 1, 2 \ldots N, j = 1, 2 \ldots N \quad (7)$$

or

$$S_{ij}^{(2)} = e^{\frac{-\left\|x_i y_j\right\|_2}{t}}, i = 1, 2 \ldots N, j = 1, 2 \ldots N \quad (8)$$

The $S^{(1)}$ in equation (7) gives large weight to $x_i \mid y_j$ pair which is similar to $x_i \mid y_i$ while the $S^{(2)}$ in equation (8) gives large weight to $x_i \mid y_j$ pair which is similar to $0 \mid 0$ .If the process works on the stable condition and the measurements are relatively accurate, $S^{(1)}$ should be selected to maintain the characteristics of the process and quality data; otherwise, if the signal-to-noise ratio (SNR) is low, $S^{(2)}$ is the optimal choice to improve the modelling precision.

The constrained optimization problem in equation (6) is transformed into unconstrained problem using Lagrange multiplier method:

$$\begin{aligned} L &= w_i^T X_i^T S Y_i v_i \\ &\quad -\lambda_1 \left(w_i^T w_i - 1\right) - \lambda_2 \left(v_i^T v_i - 1\right) \end{aligned} \quad (9)$$

where λ_1 and λ_2 are Lagrange multipliers. Let $\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial v_i} = 0$, we have:

$$\begin{aligned} X_i^T S Y_i v_i = 2\lambda_1 w_i \Rightarrow w_i^T X_i^T S Y_i v_i = 2\lambda_1 \\ Y_i^T S^T X_i w_i = 2\lambda_2 v_i \Rightarrow v_i^T Y_i^T S^T X_i w_i = 2\lambda_2 \end{aligned} \quad (10)$$

since $w_i^T X_i^T S Y_i v_i = v_i^T Y_i^T S^T X_i w_i$, it is easy to deduce that $\lambda_1 = \lambda_2$. Let $C_i = X_i^T S Y_i$, the optimal solutions of w_i and v_i are obtained by calculating the following two equations (11) recursively until convergences are realized:

$$\begin{aligned} C_i v_i &= w_i \\ C_i^T w_i &= v_i \end{aligned} \quad (11)$$

Next, same as PLS, p_i and q_i are computed by OLS. X_i and Y_i are deflated with $X_{i+1} = X_i - t_i p_i^T$ and $Y_{i+1} = Y_i - t_i q_i^T$.

B. IWPLS-based quality-relevant fault monitoring scheme

In order to make the process model more interpretable and robust, further achieve effective IWPLS-based quality-relevant fault monitoring. Two more procedures are added in to complete the IWPLS-based quality-relevant monitoring scheme.

Firstly, l_1 -norm sparseness constraints are imposed on w_i and v_i to realize sparse decomposition of X . The objection function (6) is rewritten as[15]:

$$\begin{aligned} J_{IWPLS} &= \mathrm{t}_i^T S u_i = w_i^T X_i^T S Y_i v_i \\ s.t. \left\|w_i\right\|_2^2 &\le 1, \quad \left\|v_i\right\|_2^2 \le 1, \\ \left\|w_i\right\|_1 &\le c_w, \quad \left\|v_i\right\|_1 \le c_v \end{aligned} \quad (12)$$

where c_w and c_v are hyper-parameters that control the tightness of the sparsity. Lower c_w and c_v impose tighter constraints on X and Y respectively. These two parameters are set in the range of $1 \le c_u \le \sqrt{n}$ and $1 \le c_v \le \sqrt{m}$, which enable all four constraints in (12). The solution of this constrained optimization problem is given by Witten[15].

Next, the coefficient matrix C of the constructed regression model is extracted to implement additional orthogonal decomposition. The structures of the subspaces in IWPLS are identical with PLS. We have $T = XR$ and $\hat{Y} = TQ^T$, thus $\hat{Y} = XRQ^T = XC$. Assume C is with full column rank(generally, $m \square n$), we then perform singular value decomposition (SVD) on C [16]:

$$C = UDV^T = \begin{bmatrix} \hat{U}_c & \tilde{U}_c \end{bmatrix} \begin{bmatrix} D_c & 0 \\ 0 & 0 \end{bmatrix} V^T \quad (13)$$

Since U is a orthogonal matrix, the following equation holds:

$$\begin{aligned} X &= X \begin{bmatrix} \hat{U}_c & \tilde{U}_c \end{bmatrix} \begin{bmatrix} \hat{U}_c^T \\ \tilde{U}_c^T \end{bmatrix} = X\hat{U}_c\hat{U}_c^T + X\tilde{U}_c\tilde{U}_c^T \\ &= \hat{T}_c\hat{U}_c^T + \tilde{T}_c\tilde{U}_c^T = \hat{X} + \tilde{X} \end{aligned} \quad (14)$$

where $\hat{T}_c$ and $\tilde{T}_c$ are the score matrices of $\hat{X}$ and $\tilde{X}$. $\hat{X}$ makes full contributions to Y with $Y = \hat{X} D_c V^T + \tilde{Y}$ while $\tilde{X}$ makes no contributions.

Given a new sample x_{new}, the score vectors are computed with $\hat{t}_{c\_new} = \hat{\mathrm{U}}_c x_{new}$ and $\tilde{t}_{c\_new} = \tilde{\mathrm{U}}_c x_{new}$. The statistics $\hat{T}_c^2$ and $\tilde{T}_c^2$ are employed for quality-relevant and irrelevant fault detection with their control limits $\hat{J}_c$ and $\tilde{J}_c$ [6]:

$$\hat{T}_c^2 = \hat{t}_{c\_new} \left(\frac{\hat{T}_c^T \hat{T}_c}{n-1} \right)^{-1} \hat{t}_{c\_new} \tag{15}$$

$$\tilde{T}_c^2 = \tilde{t}_{c\_new} \left(\frac{\tilde{T}_c^T \tilde{T}_c}{n-1} \right)^{-1} \tilde{t}_{c\_new} \tag{16}$$

$$\hat{J}_c = \frac{m\left(n^2-1\right)}{n\left(n-m\right)} F_{m,n-m} \tag{17}$$

$$\tilde{J}_c = \frac{\left(A-m\right)\left(n^2-1\right)}{n\left(n-A+m\right)} F_{A-m,n-A+m} \tag{18}$$

The implementation of the fault monitoring scheme is based on the following rules:

If $\hat{T}_c^2 > \hat{J}_c$, then quality-relevant fault occurs.

If $\tilde{T}_c^2 > \tilde{J}_c$ and $\hat{T}_c^2 < \hat{J}_c$, then quality-irrelevant fault occurs.

If $\tilde{T}_c^2 < \tilde{J}_c$ and $\hat{T}_c^2 < \hat{J}_c$, then no fault occurs.

IV. NUMERICAL EXPERIMENTS

The numerical example we adopted is based on Zhou[1] and Yin[6]:

$$\begin{cases} x_k = \mathrm{A} z_k + e_k \\ y_k = \mathrm{C} \mathrm{x}_k + v_k \end{cases} \tag{19}$$

where

$$\mathrm{A} = \begin{pmatrix} 1 & 3 & 4 & 4 & 0 \\ 3 & 0 & 1 & 4 & 1 \\ 1 & 1 & 3 & 0 & 0 \end{pmatrix}^T, \quad \mathrm{C} = \begin{pmatrix} 1 & 2 & 3 & 1 & 0 \end{pmatrix},$$

$z_{k,i} = \mathrm{U}\left(\left[0,1\right]\right), i = 1,2,3$ means z_k follows uniform distribution in the interval $\left[0,1\right]$, $v_k = \mathrm{N}\left(\left[0, 0.1^2\right]\right)$ and $e_{k,j} = \mathrm{N}\left(\left[0, 0.05^2\right]\right), j = 1,..,5$ means v_k and e_k follow normal distribution with zero means and the variances of 0.1^2, 0.05^2.

Fault is added in x_k in the form of:

$$x_k^f = x_k + \alpha f \tag{20}$$

where $\alpha \in \Re^{5\times 1}$ and $f \in \Re^{1\times 1}$ denote the direction and amplitude of the fault. For quality-relevant and irrelevant faults, f is set to $f_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \end{pmatrix}^T$ and $f_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix}^T$ respectively. Fault with f_1 has both quality-relevant and irrelevant influences, so it should be detected in both two statistics, which means the statistics should exceed the control limit. Fault with f_2 has quality-irrelevant influence and should be only detected in the corresponding quality-irrelevant statistic. The amplitude α varies between 0 and 20 to test the performance of existing and proposed models.

TABLE I. FARS (%) AND FDRS (%) FOR DIFFERENT FAULTS

Fault Type	Amplitude	IPLS		IWPLS	
		$T_{\hat{u}}^2$	$T_{\tilde{u}}^2$	$\hat{T}_c^2$	$\tilde{T}_c^2$
Normal	-	0	0.2	**0**	**0.2**
Quality-irrelevant	0.1	0	7.6	**0**	**7.8**
	0.2	0	71	**0**	**72.4**
	0.5	0	100	**0**	**100**
	1	0	100	**0**	**100**
	10	2.2	100	**0**	**100**
	20	11.2	100	**0**	**100**
Quality-relevant	1	2.2	85.4	**2.2**	**90.2**
	2	22.2	100	**22.4**	**100**
	3	49.6	100	**50.2**	**100**
	5	98	100	**98.4**	**100**
	10	100	100	**100**	**100**

According to Yin[6], IPLS outperforms PLS and TPLS in both detecting faults and differentiating whether the given fault has quality-relevant impact. Therefore, it is chosen for performance comparison. The number of latent variables A is set to 4 for both IPLS and IWPLS. The sparseness hyper-parameters are set to $c_w = 2.0$ and $c_v = 0$ in IWPLS to minimize the fitting error of the modelling data set. Meanwhile, $S^{(1)}$ is selected as the weight matrix for IMs in IWPLS.

The fault alarm rates (FARs) and fault detection rates (FDRs) of the two methods for two fault types with different amplitude are shown in table I.

Both two methods achieve good FARs(%). For quality-irrelevant fault, IWPLS show slightly better FDRs(%) when fault amplitude is low(α is between 0.1 to 1). However, when fault amplitude is high (α is 10 and 20), IPLS reports misleading quality-relevant faulty information, IWPLS nevertheless keeps excellent stability. The fault monitoring charts with $\alpha = 20$ are shown in figure 1. In the figure, the $T_{\hat{u}}^2$ in IPLS exceeds the control limit while the $\hat{T}_c^2$ in IWPLS stays below the control limit perfectly, which represents that the fault do not influence the product quality.

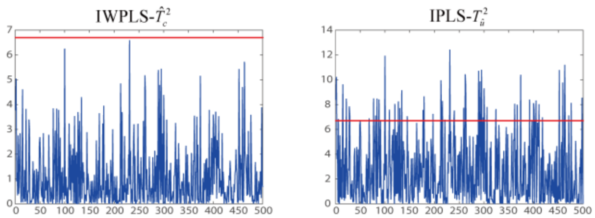

Figure 1. Quality-irrelevant Fault monitoring charts(α = 20)

For quality-relevant fault, IWPLS gives better FDRs(%) in both two statistics. The fault is completely detected after α increases beyond 10. The fault monitoring charts with α = 1 are shown in figure 2. Although both two methods detect the quality-irrelevant impact of the fault correctly, IWPLS gives relatively higher FDR(5% higher).

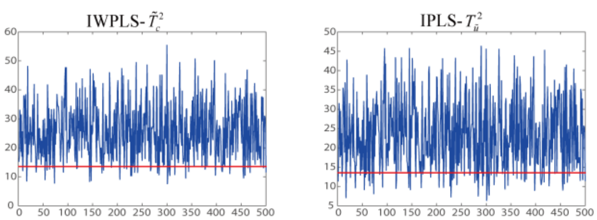

Figure 2. Qualtiy-relevant Fault monitoring charts(α =1)

V. A Case Study on Tennessee Eastman Process

The Tennessee Eastman Process(TEP) is a widely used chemical process simulation for evaluating the performance of process monitoring methods[17]. There are 5 units(stripper, separator, reactor, compressor and condenser) and 8 components(A, B, C, D, E, F and G) in this process. Among them, A, C, D and E are the reactants; G and H are the target products; B is the non-reactive inertial component and F is the by-product. Reactants A, B and C are fed to the stripper firstly. Reactants A, D, E, the vapor from the compressor and the remaining reactants from the stripper are fed to the reactor. The output stream goes through the condenser and enters the vapor-liquid separator. The vapor is sent back to the reactor by the compressor and the liquid flows into the stripper. The target product(G,H) mixture remains the base of the stripper.

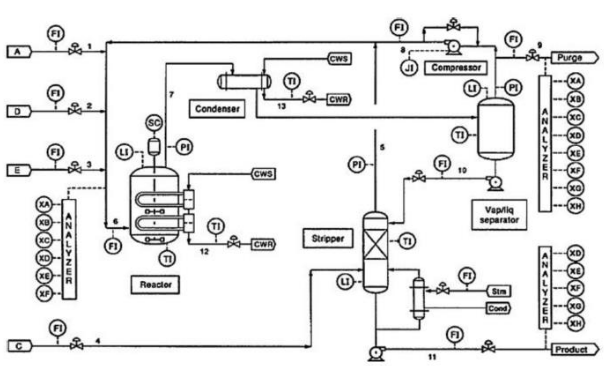

Figure 3. The flow chart of TEP

In the process, 11 manipulated variables(XMV(1-11)) and 41 measured variables(XMEAS(1-41)) are acquired from 21 designed faulty conditions(IDV(1-21)) and 1 normal condition(IDV(0)) at an interval of 3 minutes. Each condition contains 480 measurements for training and 960 measurements for testing. The description of designed faults are listed in table II[18].

TABLE II. Designed Faults in TEP

No.	Cause	Type
1	A/C feed ratio, B composition constant	Step
2	B composition, A/C ratio constant	Step
3	D feed temperature	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss	Step
7	C header pressure loss – reduced availability	Random
8	A, B, C feed composition	Random
9	D feed temperature	Random
10	C feed temperature	Random
11	Reactor cooling water inlet temperature	Random
12	Condenser cooling water inlet temperature	Random
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown
21	C feed valve	Sticking

In this paper, XMEAS(1-22) and XMV(1-11) are selected as process variables. XMEAS(35) and XMEAS(36), which correspond to the percentage composition of G and H, are selected as quality variables. Models are constructed in the basis of the testing data set of IDV(0) and are evaluated with the testing data set collected from faulty conditions.

In this case, the number of latent variables A is set to 6. The sparseness hyper-parameters are set to c_w=3.91 and c_v=1.40 and $S^{(2)}$ is selected as the weight matrix for IMs.

Firstly, in order to evaluate the modelling ability of IWPLS, the mean square error(MSE) is taken as the evaluation indicator. MSE is defined as $MSE = \left(Y - \hat{Y}\right)^2 / N$, where Y and $\hat{Y}$ are actual data and predicted data. N still represents the number of samples. For some typical faults, the comparison is listed in table III. XMEAS(35) and XMEAS(36) are abbreviated as X35 and X36 in this table.

TABLE III. MSE for different Faulty Conditions

Fault No.	PLS		IWPLS	
	X35	X36	X35	X36
IDV(2)	13.050	2.181	1.092	1.078
IDV(6)	6.443	16.657	4.626	2.110
IDV(7)	1.128	2.803	1.492	1.444
IDV(11)	0.981	1.097	0.924	1.055
IDV(17)	2.980	1.895	1.235	1.249

From table 3, under the same condition, we have reasons to believe that IWPLS performs better in predicting the value of both two quality variables. The contrasts are especially sharp for IDV(2) and IDV(6). In figure 4, the red, green and blue curves represent the original data, PLS predicted data and IWPLS predicted data. Model built by Traditional PLS cannot predict output data accurately because it only considers the corresponding input and output measurements under the normal condition. The modelling data is imprecise and lacking of variety, therefore the model is slightly overfitting. When input data change dramatically, the PLS model is unable to catch up with the changes of output data. To solve this problem, IWPLS takes IM into account. Intuitively, IWPLS increases the diversity of the original data and reduces the dependence of the model on them. Thus, IWPLS shows better performance.

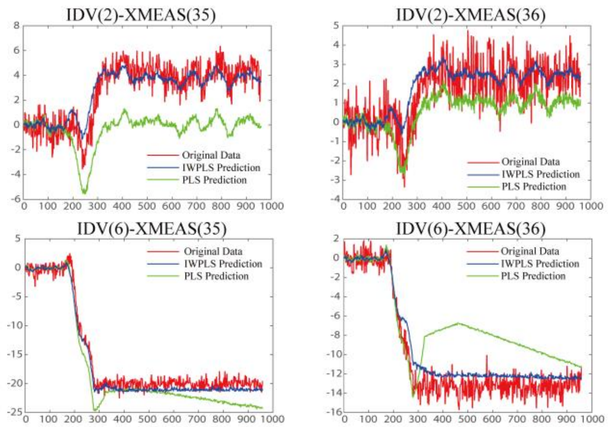

Figure 4. Prediction results of IDV(2) and IDV(6)

Next, the FDRs obtained from different statistics are shown in table IV and table V. In table II, it is clear that IDV(4,11,14) are quality-irrelevant faults. Both methods detect these faults successfully. However, in contrast with IPLS, the $\hat{T}_c^2$ statistic of IWPLS gives significantly lower FDRs for IDV(4,14) and well-matched FDR for IDV(11). Moreover, the $\tilde{T}_c^2$ statistic gives higher FDRs for IDV(11). The monitoring charts of IDV(14) with quality variable XMEAS(35) is shown in Figure 5. Fault occurs in 161th measurement, after that, IPLS reports quality-relevant fault successively while IWPLS makes correct judgment during almost the whole period.

TABLE IV. FDRS(%) FOR DIFFERENT FAULTS WITH XMEAS(35)

Fault Type	Fault No.	IPLS		IWPLS	
		$T_{\hat{u}}^2$	$T_{\tilde{u}}^2$	$\hat{T}_c^2$	$\tilde{T}_c^2$
Quality-irrelevant	IDV(4)	2.13	100	0.88	100
	IDV(11)	4.25	78.63	4.75	78.88
	IDV(14)	9.25	100	0.75	100
Quality-relevant	IDV(2)	20.50	98.40	86.88	98.38
	IDV(6)	97.63	100	97.13	100
	IDV(7)	28.00	100	27.13	100

IDV(2,6,7) are all quality-relevant faults but belong to different type. IDV(2,6) influence both two quality variables during the whole sampling period while IDV(7) only influence them for a specific period of time. IPLS can detect IDV(7) precisely but makes mistake in detecting IDV(2) due to its inaccurate modelling. Same problem exists in detecting IDV(6). Luckily, the predicted data still exceed the control limit, thus the monitoring results are correct. With regard to IWPLS, its predictive capability ensures the validity of monitoring. The monitoring charts in figure 6 prove this statement.

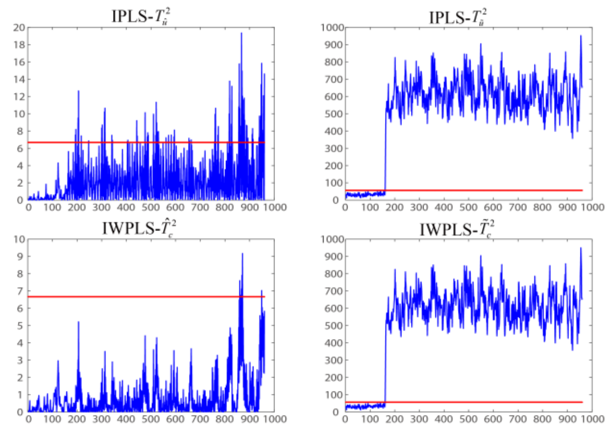

Figure 5. Monitoring charts of IDV(14) with XMEAS(35)

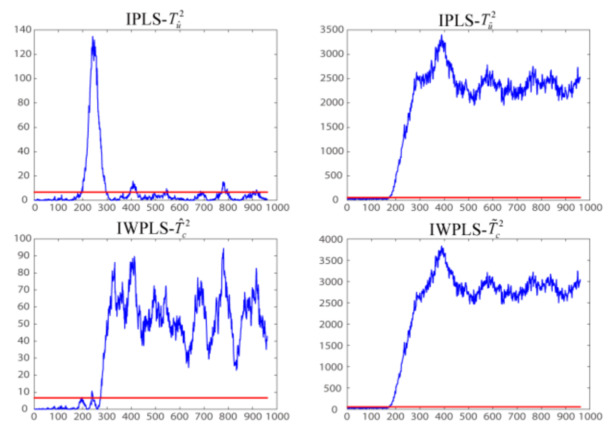

Figure 6. Monitoring charts of IDV(2) with XMEAS(35)

Table 5 shows the FDRs with XMEAS(36). For two quality variables, through analysis, 6 selected faults are classified into quality-relevant and irrelevant faults identically.

TABLE V. FDRS(%) FOR DIFFERENT FAULTS WITH XMEAS(36)

Fault Type	Fault No.	IPLS		IWPLS	
		$T_{\hat{u}}^2$	$T_{\tilde{u}}^2$	$\hat{T}_c^2$	$\tilde{T}_c^2$
Quality-irrelevant	IDV(4)	13.00	100	3.38	100
	IDV(11)	9.25	78.38	4.63	79.00
	IDV(14)	14.38	100	9.13	100
Quality-relevant	IDV(2)	89.63	98.25	89.88	98.38
	IDV(6)	97.38	100	96.75	100
	IDV(7)	88.63	100	33.50	100

For IDV(4,11,14), IWPLS reduces the FDR of quality-relevant statistic effectively. For IDV(7), the variations real XMEAS(36) values as well as its predicted values are shown in figure 7. Obviously, the quality-relevant impacts of IDV(7) is eliminated after about 400th measurements. In figure 8, we can see that the quality-relevant monitoring chart of IWPLS detects this fault accurately.

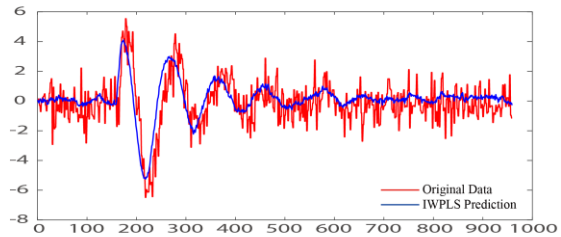

Figure 7. Variations of the original and predicted XMEAS(36)

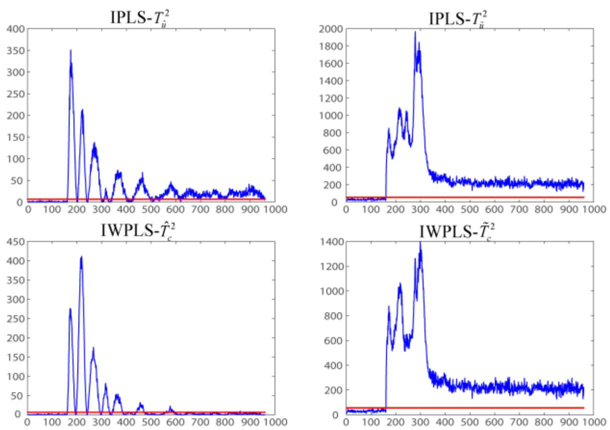

Figure 8. Monitoring charts of IDV(7) with XMEAS(36)

VI. Conclusion

A new IWPLS method for quality-relevant fault monitoring is developed. Inspired by LPP, IWPLS focuses on the ignored IM and build a weighted object function to obtain the weight vector according to the similarity among IMs. Compared to traditional PLS, IWPLS makes use of the local structures of both process and quality variable space. The constructed model is more robust and owns better generalization capability. With sparseness restraints and additional orthogonal decomposition, the proposed IWPLS-based monitoring scheme shows excellent prediction and monitoring accuracy in both numerical experiments and TEP.

Acknowledgment

We would like to extend our sincere thanks to the financial support from the National Natural Science Foundation of China (U1701262) and the New Intelligent Manufacturing Mode Application Project of Chinese Ministry of Industry and Information Technology(2016ZXFM06005).

References

[1] D. Zhou and G. Li, “Total Projection to Latent Structures for Process Monitoring,” AIChE J., vol. 2, no. No.1, pp. 168–178, 2010.

[2] S. Yin, G. Wang, and H. Gao, “Data-Driven Process Monitoring Based on Modified OrthogYin, S., Wang, G., & Gao, H. (2016). Data-Driven Process Monitoring Based on Modified Orthogonal Projections to Latent Structures. IEEE Transactions on Control Systems Technology, 24(4), 1480–1487. ht,” IEEE Trans. Control Syst. Technol., vol. 24, no. 4, pp. 1480–1487, 2016.

[3] K. Peng, K. Zhang, J. Dong, and B. You, “Quality-relevant fault detection and diagnosis for hot strip mill process with multi-specification and multi-batch measurements,” J. Franklin Inst., vol. 352, no. 3, pp. 987–1006, 2015.

[4] Y. Zhang, H. Zhou, S. J. Qin, and T. Chai, “Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares,” IEEE Trans. Ind. Informatics, vol. 6, no. 1, pp. 3–10, 2010.

[5] C. Zhao and F. Gao, “Fault-relevant Principal Component Analysis (FPCA) method for multivariate statistical modeling and process monitoring,” Chemom. Intell. Lab. Syst., vol. 133, pp. 1–16, 2014.

[6] S. Yin, X. Zhu, S. Member, and O. Kaynak, “Improved PLS Focused on Key-Performance- Indicator-Related Fault Diagnosis,” Trans. Ind. Electron., vol. 62, no. 3, pp. 1651–1658, 2015.

[7] K. Peng, K. Zhang, B. You, and J. Dong, “Quality-relevant fault monitoring based on efficient projection to latent structures with application to hot strip mill process,” IET Control Theory Appl., vol. 9, no. June 2014, pp. 1135–1145, 2014.

[8] M. Belkin and P. Niyogi, “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation,” Neural Comput., vol. 15, no. 6, pp. 1373–1396, 2003.

[9] X. He and P. Niyogi, “Locality preserving projections,” Neural Inf. Process. Syst., vol. 16, p. 153, 2004.

[10] J. Yu, “Local and global principal component analysis for process monitoring,” J. Process Control, vol. 22, no. 7, pp. 1358–1373, 2012.

[11] L. Luo, “Process monitoring with global-local preserving projections,” Ind. Eng. Chem. Res., vol. 53, no. 18, pp. 7696–7705, 2014.

[12] B. Zhong, J. Wang, J. Zhou, H. Wu, and Q. Jin, “Quality-Related Statistical Process Monitoring Method Based on Global and Local Partial Least-Squares Projection,” Ind. Eng. Chem. Res., vol. 55, no. 6, pp. 1609–1622, 2016.

[13] J. Wang, B. Zhong, and J. L. Zhou, “Quality-Relevant Fault Monitoring Based on Locality-Preserving Partial Least-Squares Statistical Models,” Ind. Eng. Chem. Res., vol. 56, no. 24, pp. 7009–7020, 2017.

[14] H. Wold, “Causal flows with latent variables,” Eur. Econ. Rev., vol. 5, no. 1, pp. 67–86, 1974.

[15] D. M. Witten, R. Tibshirani, and T. Hastie, “A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis,” Biostatistics, vol. 10, no. 3, pp. 515–534, 2009.

[16] S. X. Ding, S. Yin, K. Peng, H. Hao, and B. Shen, “A Novel Scheme for Key Performance Indicator Prediction and Diagnosis With Application to A Novel Scheme for Key Performance Indicator Prediction and Diagnosis With Application to an Industrial Hot Strip Mill,” IEEE Trans. Ind. INFORMATICS, vol. 9, no. November, pp. 2239–2247, 2013.

[17] J. J. Downs and E. F. Vogel, “A plant-wide industrial process control problem,” Comput. Chem. Eng., vol. 17, no. 3, pp. 245–255, 1993.

[18] S. Gajjar, M. Kulahci, and A. Palazoglu, “Real-time fault detection and diagnosis using sparse principal component analysis,” J. Process Control, 2017.