Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Comparison Study of Multivariate Statistics Based Key Performance Indicator Monitoring Approaches

Kai Zhang * Haiyang Hao * Zhiwen Chen * Steven X. Ding *
E.L. Ding **

* *Institute for Automatic Control and Complex Systems, University of Duisburg-Essen, Bismarckstrasse 81 BB, 47057, Duisburg, Germany (e-mail: kai.zhang@uni-due.de)*
** *Westphalian University of Applied Sciences, 45877, Gelsenkirchen, Germany*

Abstract: In this paper, multivariate statistical process monitoring approaches for key performance indicator (KPI) related static processes are reviewed under a unified framework. Based on their key nature in extracting KPI-related information from process variable space for performance monitoring, those approaches are analyzed and sorted into three categories: direct cross-correlation based decomposition method, modified least square regression based approaches, partial least square based approaches. In addition, their numerical properties and monitoring performance are compared in details. Finally the well-accepted TE benchmark process is utilized to demonstrate the theoretical comparison results and their monitoring performance from industrial viewpoint.

Keywords: Key performance indicator, process monitoring, multivariate statistical, least square, partial least square

1. INTRODUCTION

Key performance indicator (KPI) is a concept recently adopted in manufacturing industries to represent the key technical factors closely related to product quality, which are usually hard to be online quantified or sampled with significant time-delay. In practice, KPI are governed by the environment which consists of the actuators, sensors as well as the embedded control policies (Ding et al 2013), and the variables reflecting these information are usually available. Thus, for the purpose of on-line monitoring of the degradation in KPI, it is reasonable to model KPIs and process variables, then utilize the model to guide the real-time implementation. Model-based methods serve as the mainstream over the past decades, while for some processes like chemistry and semiconductor, it is extremely hard to extract their hidden process models. Supposing that, data-driven based one could be an efficient alternative, now has attracted soaring focus (Tracy et al. 1992, MacGregor et al. 1994).

Multivariate statistical process monitoring (MSPM) is a valuable member within data-driven community (Qin 2003; 2012). Taking KPI into account, MSPM methods are typically limited to significantly utilized partial least squares (PLS) based methodologies (MacGregor et al. 1994). PLS model is created for performing linear regression or coping with the collinear problem (Wold et al. 2001). It successfully overcomes the disadvantage of least square(LS) for computing the inversion of the covariance matrix. PLS model has been brought into the MSPM field at the beginning of 1990s. And since then, great developments have been achieved (MacGregor et al. 1994).

Very recently, based on PLS, many successful results have been reported for monitoring KPI relevant objects more accurately. Li et al. (2010) has revealed the geometric nature of PLS for process monitoring. Zhou et al. (2010) proposed a total PLS (T-PLS) model with a more detailed decomposition for process variable matrix. It improves PLS from two viewpoints, first of all, the scores include the redundance for explaining the predicted KPIs, secondly the residual part can not be neglected for detecting KPI related faults. Further more, concurrent PLS (C-PLS) was proposed by Qin et al. (2013) that addresses the same problem with T-PLS and claimed to be more efficient.

Although it is shown of great effectiveness, PLS, indeed bears heavy computation overhead (Yin et al. 2012), and also involves a parameter to be fixed in advance. To overcome this inconvenience, Yin et al. (2011) proposed a modification of PLS model, which is calculated economically and requires no parameter a priori. Its utility was proved to be superior than PLS in a benchmark study. Subsequently, Ding et al (2013) established a more efficient approach, which originates from classic LS regression, the result is convincing in the application to a hot strip mill process. In this review work, the two methods will naturally be sorted into LS based approaches and will be compared with PLS. It should be noticed that the original PLS and LS techniques both stem from linear regression field rather than process monitoring area. To strongly focus on the

[1] Kai Zhang and Zhiwen Chen appreciate the funding support from China Scholarship Council

theme of process monitoring, we would like to commence this paper from the monitoring purpose and introduce a basic scheme firstly, which is followed by LS and PLS based approaches. Further more, a unified projection structure involved in these approached is explored which can enlighten the deeper research in this area.

The rest of this paper is structured in the following. Section 2 presents the problem formulation of monitoring the KPI related processes, and a basic resolving scheme with SVD. In section 3, LS based methods are reviewed and discussed theoretically. Section 4 mainly concerns about PLS based approaches. Some comparisons and summaries are declared in Section 5. Section 6 includes a case study simulation to compare these approaches and verify some valuable consequences. The final conclusions will be drawn in Section 7.

2. PROBLEM FORMATION AND A SIMPLE SOLUTION

Acquire the raw process data $y_{obs,i} \in \mathcal{R}^m$ and key performance indicators $\theta_{obs,i} \in \mathcal{R}^l$ with $i \in [1, N]$. The data matrices could be formulated in forms of $Y_{obs} = [y_{obs,1}, y_{obs,2}, ..., y_{obs,N}]$ and $\Theta_{obs} = [\theta_{obs,1}, \theta_{obs,2}, ..., \theta_{obs,N}]$. Based on the normalized off-line data matrices $Y = [y_1, ..., y_N] \in \mathcal{R}^{m\times N}$ and $\Theta = [\theta_1, ..., \theta_N] \in \mathcal{R}^{l\times N}$, the problem is formulated of how to monitor the process evolution, KPI especially, by using the on-line normalized measurement $y_{on-line}$. Intuitively, a typical framework is to project $y_{on-line}$ into two subspaces as

$$y_{on-line} = \underbrace{\Pi_\theta y_{on-line}}_{\theta-releated} + \underbrace{\Pi_{\theta^\perp} y_{on-line}}_{\theta-unrelated} \quad (1)$$

where Π_θ and $\Pi_{\theta^\perp}$ represent the liner projectors. What follow are different manners to obtain Π_θ and $\Pi_{\theta^\perp}$, then build the fault detection indices statistically.

The cross-correlation between y and θ can be expressed as: $\frac{\Theta Y^T}{N-1}$. Performing a singular value decomposition on it yields:

$$SVD\left(\frac{\Theta Y^T}{N-1}\right) = V\Sigma U^T = V\left[\Sigma_{\theta y} 0\right]\begin{bmatrix} U_{\theta y}^T \\ U_{\theta y^\perp}^T \end{bmatrix} = V\Sigma_{\theta y} U_{\theta y}^T \quad (2)$$

Define two orthogonal projections $\hat{y} = U_{\theta y}U_{\theta y}^T y$ and $\tilde{y} = U_{\theta y^\perp}U_{\theta y^\perp}^T y$, namely $\Pi_\theta = U_{\theta y}U_{\theta y}^T$ and $\Pi_{\theta^\perp} = U_{\theta y^\perp}U_{\theta y^\perp}^T$. It is fairly straightforward to obtain the property that

$$\varepsilon\left(\tilde{y}\theta^T\right) = \frac{U_{\theta y^\perp}U_{\theta y^\perp}^T Y\Theta^T}{N-1} = 0 \quad (3)$$

Geometrically, $\hat{y}^T\tilde{y} = 0$ holds, which is achieved by projecting y into two orthogonal spaces $U_{\theta y}$ and $U_{\theta y^\perp}$. $\hat{y}$ covers the θ-correlated section in y, while $\tilde{y}$ is supplemented. It is reasonable to conclude that $rank\left(U_{\theta y}^T Y\right) = l$ and $rank\left(U_{\theta y^\perp}^T Y\right) = m - l$. Therefore, the two kinds of T^2 detection indices are constructed in the following,

$$\begin{aligned} T_{\theta y}^2 &= y^T U_{\theta y}\left(\frac{U_{\theta y}^T YY^T U_{\theta y}}{N-1}\right)^{-1} U_{\theta y}^T y \\ T_{\theta y^\perp}^2 &= y^T U_{\theta y^\perp}\left(\frac{U_{\theta y^\perp}^T YY^T U_{\theta y^\perp}}{N-1}\right)^{-1} U_{\theta y^\perp}^T y \end{aligned} \quad (4)$$

Remark that the involved two inverse computations may not be stable due to the fact that the minimum singular values of the two covariance matrices may be significantly small. To ease this problem, two robust statistics are presented by referring Ding et al. (2010):

$$T_{\theta y}^2 = y^T U_{\theta y}\Xi_1 U_{\theta y}^T y, T_{\theta y^\perp}^2 = y^T U_{\theta y^\perp}\Xi_2 U_{\theta y^\perp}^T y \quad (5)$$

where
$\Xi_1 = P_1 diag\left(\frac{\lambda_{1,\kappa}}{\lambda_{1,1}} \dots \frac{\lambda_{1,\kappa}}{\lambda_{1,\kappa}}\right)P_1^T, \frac{U_{\theta y}^T YY^T U_{\theta y}}{N-1} = P_1\Lambda_1 P_1^T$, $\Lambda_1 = diag\underbrace{(\lambda_{1,1} \dots \lambda_{1,\kappa})}_{\lambda_{1,1}\geq\cdots\geq\lambda_{1,\kappa}}$, $\frac{U_{\theta y^\perp}^T YY^T U_{\theta y^\perp}}{N-1} = P_2\Lambda_2 P_2^T$, $\Lambda_2 = diag\underbrace{(\lambda_{2,1} \dots \lambda_{2,m-\kappa})}_{\lambda_{2,1}\geq\cdots\geq\lambda_{2,m-\kappa}}$, $\Xi_2 = P_2 diag\left(\frac{\lambda_{2,m-\kappa}}{\lambda_{2,1}} \dots \frac{\lambda_{2,m-\kappa}}{\lambda_{2,m-\kappa}}\right)P_2^T$.

Under the popular assumption of Normal distribution, the thresholds are given according to Ding et al. (2010):

$$J_{th,T_{\theta y}^2} = \lambda_{1,\kappa}\chi_{1-\alpha}^2(\kappa), J_{th,T_{\theta y^\perp}^2} = \lambda_{2,m-\kappa}\chi_{1-\alpha}^2(m-\kappa) \quad (6)$$

Definitely, the above scheme is quite simple but originates from the monitoring purpose. Parallel to PCA, it could be viewed as a direct decomposition (DD) based solution of MSPM for KPI related issue, and possesses a strong statistical background.

3. LS BASED APPROACHES

LS plays a central role in linear regression theory, it predicts KPIs not only relying on the cross-correlation of y and θ but also the covariance of y. In a general case (Qin 1998, Yin et al. 2011), the regression model could be modified as:

$$\hat{\theta} = \Psi y, \Psi = \varepsilon\left(\theta y^T\right)\varepsilon\left(yy^T\right)^\dagger = \Theta Y^T\left(YY^T\right)^\dagger \quad (7)$$

where $\hat{\theta}$ represents the LS-prediction of θ. $\left(YY^T\right)^\dagger = P_y\Lambda_y^{-1}P_y$, Λ_y and P_y are the non-zero singular values and their corresponding vectors of YY^T. For on-line monitor θ using y, it is wise to decompose y governed by $\hat{\theta}$. Yin et al project y onto the subspace spanned by $col(\Psi^T)$, which is supposed to be responsible for KPI prediction, and onto the $col(\Psi^T)^\perp$, which contributes nothing to $\hat{\theta}$. The procedure could be realized subsequently. Perform a QR decomposition on Ψ^T:

$$\Psi^T = [Q_1, Q_2]\begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1 \quad (8)$$

Then $col\left(\Psi^T\right) = span\left(Q_1\right)$ and $col(\Psi^T)^\perp = span\left(Q_2\right)$. The projections can be achieved as $\hat{y} = Q_1{Q_1}^T y$ and $\tilde{y} = Q_2{Q_2}^T y$. They satisfy the requirements that

$$\begin{aligned} \Psi\hat{y} &= \Psi Q_1{Q_1}^T y = R_1^T{Q_1}^T y = \Psi y \\ \Psi\tilde{y} &= \Psi Q_2{Q_2}^T y = R_1^T{Q_1}^T Q_2{Q_2}^T y = 0 \end{aligned} \quad (9)$$

As $rank\left(Q_1{}^T y\right) = l$ and $rank\left(Q_2{}^T y\right) = m - l$ hold. The T^2 distribution can be shown for them:

$$\begin{aligned} T^2_{\hat{y}} &= y^T Q_1 \left(\frac{Q_1^T Y Y^T Q_1}{N-1}\right)^{-1} Q_1^T y \\ T^2_{\tilde{y}} &= y^T Q_2 \left(\frac{Q_2^T Y Y^T Q_2}{N-1}\right)^{-1} Q_2^T y \end{aligned} \tag{10}$$

The thresholds are calculated by scaled F distribution. Remark that $T^2_{\hat{y}} = y^T Q_1 R_1 \left(\frac{R_1^T Q_1^T Y Y^T Q_1 R_1}{N-1}\right)^{-1} R_1^T Q_1^T y = y^T \Psi^T \left(\frac{\Psi Y Y^T \Psi^T}{N-1}\right)^{-1} \Psi y = \hat{\theta}^T \left(\frac{\Theta\Theta^T}{N-1}\right) \hat{\theta}$, which is consistent with the purpose of KPI-related process monitoring. That is when this is no on-line KPI available, the predicted KPIs are employed instead. Very recently, Ding et al (2013) proposed another LS based method, of which, the principal components of y are taken account to estimate its covariance firstly. Then YY^T is replaced with $P_{y,pc}\Lambda_{y,pc}P_{y,pc}$, where $\Lambda_{y,pc} = diag\left(\lambda_1, ..., \lambda_{\bar{m}}\right)$ stands for the principal eignvalues, the $P_{y,res}\Lambda_{y,res}P^T_{y,res}$ is abandoned temporarily. Then the regression model could be modified as:

$$\hat{\theta} = \Theta Y^T P_{y,pc} \Lambda^{-1}_{y,pc} P^T_{y,pc} y = \bar{\Psi}\bar{y} \tag{11}$$

where $\bar{y} = \Lambda^{-1/2}_{y,pc} P^T_{y,pc} y$, $\bar{\Psi} = \Theta\left(\Lambda^{-1/2}_{y,pc} P^T_{y,pc} Y\right)^T$. (11) acts as a transformation of LS regression. Subsequently, the following procedures are similar with aforementioned one. Firstly, run a QR decomposition on $\bar{\Psi}^T = \left[\bar{Q}_1, \bar{Q}_2\right]\left[\bar{R}^T, 0\right]^T$. Then $\bar{Q}_1^T \bar{y}$ satisfies $\bar{Q}_1^T \bar{y} \sim N\left(0, I_{l\times l}\right)$ is applied for KPI related fault detection, while $\bar{Q}_2^T \bar{y} \sim N\left(0, I_{(\bar{m}-l)\times(\bar{m}-l)}\right)$ is adopted for KPI-unrelated faults. It also could be drawn that performing T^2 statistic on $\bar{Q}_1^T \bar{y}$ is analogous with the one on $\hat{\theta}$. Then,

$$T^2_{\hat{\theta}} = y^T P_{y,pc} \Lambda^{-1/2}_{y,pc} \bar{Q}_1 \bar{Q}_1^T \Lambda^{-1/2}_{y,pc} P^T_{y,pc} y \tag{12}$$

In addition, Ding et al (2013)'s approach covers the abandoned information of $P^T_{y,res}$ as the excitation for KPI-unrelated faults. Hao et al. (2013) combined them and proposed a robust index, which is

$$T^2_{\theta^\perp} = y^T \begin{pmatrix} \lambda^2_m P_{y,pc} \Lambda^{-1/2}_{y,pc} \bar{Q}_2 \bar{Q}_2^T \Lambda^{-1/2}_{y,pc} P^T_{y,pc} \\ + P_{y,res} \Xi_{res} P^T_{y,res} \end{pmatrix} y \tag{13}$$

where $\Xi_{res} = diag\left(\lambda^2_{\bar{m}}/\lambda^2_{\bar{m}+1}, ..., \lambda^2_m/\lambda^2_{m+1}, 1\right)$. The thresholds could be acquired by following (6).

To be summarized, LS based monitoring captures the capability of LS regression, induce an orthogonal decomposition on process variables, one is to account for predicted KPI, another contributes noting for KPI prediction. The monitoring policy considers the entire process variable space. Also, there is no parameter needed to be initialized by a priori.

4. PLS BASED APPROACHES

4.1 Process monitoring based on PLS

PLS regression is a multivariate analysis method that constructs a linear regression between two data sets expressed in form of data matrices, namely, the process variable measurements and KPIs. The basic idea behind PLS is to sequentially find the KPI-related information from the process variables, which will be represented with a group of uncorrelated score vectors. The score is a linear transformation of y, and chosen to maximize the correlation between Y and Θ. According to our understanding, the PLS algorithm is reorganized step-wise in Table 1, which is parallel to the original PLS model.

Table 1. PLS mdoel

1: Set $Y_1 = Y$ and recursively calculate for $i = 1, ..., \gamma$
$w_i^* = \arg\max_{\Vert w_i \Vert = 1} \left\Vert w_i^T Y_i \Theta^T \right\Vert_E$
$t_1 = w_i^T Y, p_i = Y_i t_i^T / t_i t_i{}^T$
$r_i = \begin{cases} w_1, i = 1 \\ \prod_{j=1}^{i-1} \left(I - w_j p_j^T\right) w_i, i > 1 \end{cases}, q_i = \Theta t_i^T / t_i t_i{}^T$
$Y_{i+1} = Y_i - p_i t_i$
where γ is a prior known stopping criterion.
2: Form the matrices T, P, Q,and R:
$T = [t_1^T, ..., t_\gamma^T]^T$, $P = [p_1, ..., p_\gamma]$, $Q = [q_1, ..., q_\gamma]$,
$R = [r_1, ..., r_\gamma]$

PLS decomposes y and θ into $y = \hat{y}_{PLS} + \tilde{y}_{PLS} = PR^T y + \left(I - PR^T\right) y$ and $\theta = \hat{\theta}_{PLS} + \tilde{\theta}_{PLS} = QR^T y + \tilde{\theta}_{PLS}$, respectively. $\hat{\theta}_{PLS}$ stands for the predictable part of θ with PLS. Some statistical properties could be found

$$\begin{aligned} \varepsilon\left(yy^T\right) &= \varepsilon\left(\hat{y}_{PLS}\hat{y}^T_{PLS}\right) + \varepsilon\left(\tilde{y}_{PLS}\tilde{y}^T_{PLS}\right) \\ \varepsilon\left(\theta\theta^T\right) &= \varepsilon\left(\hat{\theta}_{PLS}\hat{\theta}^T_{PLS}\right) + \varepsilon\left(\tilde{\theta}_{PLS}\tilde{\theta}^T_{PLS}\right) \\ \varepsilon\left(\hat{y}_{PLS}\tilde{\theta}^T_{PLS}\right) &= 0, \varepsilon\left(\tilde{y}_{PLS}\hat{\theta}^T_{PLS}\right) = 0 \end{aligned} \tag{14}$$

Note that $\varepsilon\left(\tilde{y}_{PLS}\tilde{\theta}^T_{PLS}\right) \neq 0$ or $\varepsilon\left(\tilde{y}_{PLS}\tilde{\theta}^T_{PLS}\right) \approx 0$. The monitoring strategy based on PLS utilizes $\hat{y}_{PLS}$ to interpret the KPI-related faults. As well known that $rank\left(R^T \hat{Y}_{PLS}\right) = rank\left(R^T Y\right) = rank\left(T\right) = \gamma$. It is reasonable to build the index on $R^T y$, and $R^T y \sim N\left(0, \frac{TT^T}{N-1}\right)$. So

$$T^2_{PLS} = y^T R \left(\frac{TT^T}{N-1}\right)^{-1} R^T y \tag{15}$$

Its control limit follows a scaled F statistic: $J_{th,T^2_{PLS}} = \frac{\gamma(N^2-1)}{N(N-\gamma)} F_\alpha\left(\gamma, N-\gamma\right)$. Q statistic is generally applied for $\tilde{y}_{PLS}$ in the case of $m \gg \gamma$. The index is developed in a form of

$$SPE = \tilde{y}^T_{PLS}\tilde{y}_{PLS} \tag{16}$$

Then its control limit is designed to be $J_{th,SPE_{PLS}} = g\chi^2_\alpha\left(h\right)$ according to Box (1954), where $g = S/2\mu_{SPE}$, $h = 2\mu^2_{SPE}/S$. S and μ_{SPE} are trained with the training data.

Li et al. (2010) have proved that

$$T^2_{PLS} \equiv \hat{y}^T_{PLS} \left(\frac{\hat{Y}_{PLS}\hat{Y}^T_{PLS}}{N-1}\right)^\dagger \hat{y}_{PLS} \tag{17}$$

holds, also it is obvious that T^2_{PLS} does not equal to the result of performing the T^2 statistic on $\hat{\Theta}_{PLS}$, which is different from LS based methods.

4.2 Process monitoring based on enhanced PLS

In this subsection, two enhanced models will be talked about, which are both established on the basis of PLS model.

T-PLS T-PLS proposed by Zhou et al. (2010) has shown greater effectiveness in the performance of handling KPI-related process monitoring. The motivation of T-PLS lies in the disadvantage of PLS when used for process monitoring. For one thing, $\hat{y}_{PLS}$ includes much more redundant information unrelated $\hat{\theta}_{PLS}$, for another, $\tilde{y}_{PLS}$ is weakly related with $\tilde{\theta}_{PLS}$, which may ultimately affect y. The detailed T-PLS algorithm can be referred in Zhou et al's work. The decomposition procedure for Y could be summarized in Table 2.

Table 2. T-PLS decomposition of Y

1: Do PLS to obtain $\hat{Y}_{PLS}$, $\tilde{Y}_{PLS}$, $\hat{\Theta}_{PLS}$.
2: Run PCA to extract the score vectors of $\hat{\Theta}_{PLS}$: $T_\theta = R_\theta^T Y$
3: Apply T_θ to reconstruct $\hat{Y}_{PLS,\theta}$ from $\hat{Y}_{PLS}$:$\hat{Y}_{PLS,\theta} = P_\theta T_\theta$
4: Perform PCA on $\hat{Y}_{PLS,o} = \hat{Y}_{PLS} - \hat{Y}_{PLS,\theta}$ yields: $T_o = R_o^T Y$, $\hat{Y}_{PLS,o} = P_o T_o$
5: Do a PCA decomposition on $\tilde{Y}_{PLS}$ yields: $\tilde{Y}_{PLS} = \hat{Y}_{PLS,p} + \tilde{Y}_{PLS,r}$, $T_r = R_r^T Y$, $\hat{Y}_{PLS,p} = P_r T_r$

T-PLS further decomposes y into four subspaces that are uncorrelated with each other: $y = \hat{y}_{PLS,\theta} + \hat{y}_{PLS,o} + \hat{y}_{PLS,p} + \tilde{y}_{PLS,r} = P_\theta R_\theta^T y + P_\theta R_o^T y + P_r R_r^T y + \tilde{y}_{PLS,r}$. $\hat{y}_{PLS,\theta}$ is supposed to be directly responsible for $\hat{\theta}_{PLS}$. $\tilde{y}_{PLS,p}$ and $\hat{y}_{PLS,o}$ are thought to be unrelated with y. As for $\tilde{y}_{PLS,r}$, it is nominated that possibly affect θ, and is included in the KPI-related detection. The monitoring indices based on T-PLS are comprehensive, which are established for these four parts, respectively. $\hat{y}_{PLS,\theta}$ could be represent using t_θ, thus it is able to build a $T^2_{\hat{y}_{PLS,\theta}}$ statistic on t_θ like T^2_{PLS}: $T^2_{\hat{y}_{PLS,\theta}} = t_\theta^T \left(T_\theta T_\theta^T/(N-1)\right)^{-1} t_\theta$. Remark that $T^2_{\hat{y}_{PLS,\theta}} = T^2_{\hat{\theta}_{PLS}} = \hat{\theta}^T_{PLS}\left(\hat{\Theta}_{PLS}\hat{\Theta}^T_{PLS}\big/(N-1)\right)^{-1}\hat{\theta}_{PLS}$, which is likewise in LS regression based monitoring. $\hat{y}_{PLS,o}$ could be monitored based on the T^2 statistic on t_o, namely $T^2_{\hat{y}_{PLS,o}} = t_o^T\left(T_o T_o^T/(N-1)\right)^{-1} t_o$. $T^2_{\tilde{y}_{PLS,p}}$ is similarly obtained with $T^2_{\tilde{y}_{PLS,p}} = t_r^T\left(T_r T_r^T/(N-1)\right)^{-1} t_r$. The finial part of $\tilde{Y}_{PLS,\theta}$ contains tiny variations that should be monitored by Q_{T-PLS} statistic, the same as SPE. Above mentioned score vectors t_θ, t_o and t_r are calculated from $R_\theta^T y, R_o^T y, R_r^T y$, respectively. Of the detection indices, the control limits of T^2 statistic are fixed with classic F distribution, Q statistic will be bounded using scaled χ^2 distribution.

C-PLS C-PLS is a recently proposed approach which argues to solve the KPI related process monitoring issue more accurately (Qin et al. 2013). It, on one hand, resembles LS based methods to obtain the directly KPI-related part, on the other hand, inherits the spirit of T-PLS to consider the possible KPI-related part. To be more convenient, it only divides y into three subspaces, compared with four subspace of T-PLS. The detailed C-PLS algorithm is shown in Table 3. C-PLS models y as $y = y_\theta + \hat{y}_{\theta^\perp} + \tilde{y}_{\theta^\perp} = (R_\theta^T)^\dagger R_\theta^T y + P^T_{\theta^\perp} R^T_{\theta^\perp} y + \tilde{y}_{\theta^\perp}$ and employs $y_{PLS,\theta}$ and $\tilde{y}_{PLS,\theta}$ for monitoring KPI-related faults, while $\hat{y}_{PLS,\theta^\perp}$ serves for KPI-unrelated faults.

Table 3. C-PLS decomposition of Y

1: Do a PLS decomposition to obtain the $\hat{\Theta}_{PLS}$.
2: Divide Y supervised by $\hat{\Theta}_{PLS}$: $Y = Y_\theta + Y_{\theta^\perp}$
2.1: Do a PCA on $\hat{\Theta}_{PLS}$ to extract the score T_θ: $T_\theta = R_\theta^T Y$
2.2: Project Y onto the orthogonal subspaces: $Y_\theta = (R_\theta^T)^\dagger R_\theta^T Y$ and $Y_{\theta^\perp} = \left(I - (R_\theta^T)^\dagger R_\theta^T\right) Y$
3: Do a PCA decomposition on $Y_{\theta^\perp}$:$Y_{\theta^\perp} = \hat{Y}_{\theta^\perp} + \tilde{Y}_{\theta^\perp}$: $T_{\theta^\perp} = R^T_{\theta^\perp} Y$, $\hat{Y}_{\theta^\perp} = P_{\theta^\perp} T_{\theta^\perp}$

$y_{PLS,\theta}$ could be adopted to design a index: $T^2_{y_\theta} = t_\theta^T\left(T_\theta T_\theta^T/(N-1)\right)^{-1} t_\theta$. Q statistic is more suitable for $\tilde{y}_{PLS,\theta}$, then a $Q_{C-PLS} = \tilde{y}_\theta^T \tilde{y}_\theta$ is generated. For $\hat{y}_{PLS,\theta^\perp}$, which is produced by a PCA decomposition, thus a suitable T^2 statistic: $T^2_{y_{\theta^\perp}} = t^T_{\theta^\perp}\left(T_{\theta^\perp} T^T_{\theta^\perp}/(N-1)\right)^{-1} t_{\theta^\perp}$ is given. All of their control limits can be chosen according to aforementioned corresponding approaches.

5. COMPARISON AND SUMMARY

First of all, the two projectors Π_θ and $\Pi_{\theta^\perp}$ induced by aforementioned approaches are summarized in Table 4. It is straightforward to prove that the condition $\Pi_\theta + \Pi_{\theta^\perp} = I$ is satisfied for all items. In addition, they are all idempotent, which follows the condition of being linear projectors. The expressions, meanwhile declare that DD, LS1 and LS2 deduce an orthogonal projection, while the remainders provide oblique ones.

By referring Table 5, DD based method is minimum in computation, which only includes one SVD on the $m \times l$ cross-covariance matrix. LS based methods involve an SVD on $m \times m$ covariance matrix and a QR decomposition on $m \times l$ matrix. PLS is not a time-efficient model, it needs γ times of SVDs on $m \times l$ matrix. The most intensive calculation could be found in T-PLS and C-PLS. T-PLS concludes γ times SVD on $m \times l$ matrix, two times on $m \times m$ and one SVD on $l \times l$. C-PLS is an SVD on $m \times m$ matrix less than T-PLS. Under the case of $\gamma \gg 2$, PLS based methods have to suffer even more severe computational disasters. Besides, γ is a user-predefined factor, which may determine the behaviors of PLS based methodologies. Cross-validation is popularly referred for picking up an appropriate γ. Although it is universally applied, the perplexing calculations still can not be avoided.

Table 5. Computational cost comparison

Approach	Computation complexity
DD	1 SVD on $m \times l$ matrix
LS1	1 SVD on $m \times m$ matrix +1 SVD on $m \times l$ matrix
LS2	the same with LS1
PLS	γ times of SVDs on $m \times l$ matrix
T-PLS	cost of PLS +2 SVD on $m \times m$ + 1 SVD on $l \times l$.
C-PLS	cost of PLS +1 SVD on $m \times m$ + 1 SVD on $l \times l$.

The direct SVD based method attempts to isolate the θ-uncorrelated part from the θ uncorrelated part, whilst LS based methods try to identify the the section of directly responsible for predicting θ from the section of completely indifferent for θ. As the two methods address

Table 4. Summary of projectors

Pr.	DD	LS1	LS2[1]	PLS	T-PLS	C-PLS
Π_θ	$U_{\theta y}U_{\theta y}^T$	$Q_1 Q_1^T$	$P_{y,pc}\Lambda_{y,pc}^{1/2}\bar{Q}_1\bar{Q}_1^T\Lambda_{y,pc}^{-1/2}P_{y,pc}^T$	PR^T	$I-P_oR_o^T-P_rR_r^T$	$I-P_{\theta\perp}R_{\theta\perp}^T$
$\Pi_{\theta\perp}$	$U_{\theta y^\perp}U_{\theta y^\perp}^T$	$Q_2 Q_2^T$	$P_{y,pc}\Lambda_{y,pc}^{1/2}\bar{Q}_2\bar{Q}_2^T\Lambda_{y,pc}^{-1/2}P_{y,pc}^T + P_{y,res}P_{y,res}^T$	$I-PR^T$	$P_oR_o^T+P_rR_r^T$	$P_{\theta\perp}R_{\theta\perp}^T$

[1] LS1:LS based method by Yin et al. LS2: LS based method by Ding et al.

Table 6. Summary of detection indices

Approach	KPI-related index	KPI-unrelated index
DD	$T^2_{\theta y}$	$T^2_{\theta y^\perp}$
LS1	$T^2_{\hat{y}}$	$T^2_{\tilde{y}}$
LS2	$T^2_{\hat{\theta}}$	$T^2_{\theta^\perp}$
PLS	T^2_{PLS}	SPE
T-PLS	$T^2_{\hat{y}_{PLS,\theta}}$ or Q_{T-PLS}	$T^2_{\hat{y}_{PLS,o}}$ or $T^2_{\hat{y}_{PLS,p}}$
C-PLS	$T^2_{y_\theta}$ or Q_{C-PLS}	$T^2_{y_{\theta^\perp}}$

the KPI related monitoring issue from distinct view points, it is ambiguous to judge which one overweighs another theoretically. Compared with LS, the scheme based on PLS is somewhat different. For monitoring the KPI related part, its detection index focuses on $\hat{Y}_{PLS}$ instead of $\hat{\theta}_{PLS}$, which will cause significant false alarms. Differently, T-PLS and C-PLS both place strong emphasis on $\hat{\theta}_{PLS}$, and extensively reduce the false alarms. Furthermore, they have also considered the fact that the residuals of PLS may also affect θ due to their weak connection with θ. Although it is still uncertain what the mechanism is, there appears no doubt that concluding this part as the potentials would further improve the detection result. Further corresponding to C-PLS and T-PLS, analog to (17), the statistics on scores are equivalent to that on their regarding section in y, but in a form of MP inverse. Finally, the involved detection indices are attached in Table 6.

6. CASE STUDY ON TEP BENCHMARK

TEP benchmark has been overwhelmingly applied to e-valuate different process monitoring strategies. The detailed technical description could be referred from(Chiang et al. 2000; 2001). In this work, a well-known TE simulator (Matlab file) will be utilized to generate the data. The recorded dataset include 12 manipulated variables and 41 process variables. In this work, 31 variables are picked out and form the process variable group, namely $y_{obs} = [xmeas\,(1-22)\,, xmv\,(1-4)\,, xmv\,(6-7)\,, xmv\,(10-11)]^T$, xmv represents the manipulated variable, $xmeas$ stands for the process variable. KPIs consist of xmeas 35 and 36, $\theta_{obs} = [xmeas35, xmeas36]$. TEP has predefined 20 fault patterns, all of which are injected from the 160th sample. For all tested approaches, the 960 normal operation samples are adopted for the training phase, then the 960 faulty samples serve for the implementing phase. γ for PLS is chosen to be 6 according to cross-validation. The significance level $\alpha = 0.05$ is set for an appropriate control limit.

The faults occurring herein are considerably complex, of which mechanisms are hardly to interpret. We firstly evaluate the performance of the approaches for dealing with all kinds of faults. The simulation results are summarized in Table 7. As can be seen, two types of LS based methods present the similar performance, so are like in the T-PLS and C-PLS. PLS is shown poorly comparatively, it is even worsen than the directly decomposition based method. LS based one is a bit better than direct decomposition, but is poorer than enhanced PLS based one. Consider the computational cost, LS based methods definitely could be an excellent candidate.

Table 7. The detection performance using both KPI-related and -unrelated indices

Fault	DD	LS1	LS2	PLS	T-PLS	C-PLS
1	1	1	1	1	1	1
2	0.9975	0.9975	0.9975	0.9975	0.9975	0.9975
3	0.1900	0.2450	0.1913	0.1925	0.3912	0.3113
4	1	1	1	1	1	1
5	0.0877	0.1088	0.0912	0.0938	0.1900	0.1437
6	1	1	1	1	1	1
7	1	1	1	1	1	1
8	0.9938	0.9938	0.9938	0.9925	0.9950	0.9962
9	0.2863	0.3250	0.2988	0.2612	0.4363	0.3575
10	0.9788	0.9788	0.9800	0.9300	0.9775	0.9712
11	0.9838	0.9825	0.9838	0.9788	0.9850	0.9838
12	0.7087	0.6937	0.6963	0.6075	0.7050	0.6513
13	0.9912	0.9938	0.9912	0.9900	0.9938	0.9930
14	0.9988	0.9988	0.9988	0.9975	0.9988	0.9988
15	0.1050	0.1275	0.1038	0.1163	0.2163	0.1725
16	0.0862	0.1075	0.0925	0.0998	0.1938	0.1475
17	0.9875	0.9875	0.9875	0.9838	0.9888	0.9862
18	0.9063	0.9075	0.9100	0.8588	0.9013	0.9087
19	0.9988	0.9988	0.9988	0.9888	1	1
20	0.9900	0.9900	0.9900	0.9875	0.9912	0.9988

As KPI has been increasingly concerned, the behavior of KPI related indices should be payed more attention. It should be first mentioned that all faults could be divided into two groups according to whether KPIs' properties (mean ε and variance σ^2) have been affected or not. The result is illustrated in Table 8, where KPI-unaffected faults 3, 5, 9, 14, 15 and 16 are colorfully distinguished. Comparing the figures in Table 8, the definitive KPI-related detection index in both T-PLS and C-PLS show the same result, which is consistent with the theoretical analysis. The direct decomposition based approach is the poorest for KPI-affected faults, but gives moderate false alarms for KPI-unaffected faults. LS based approaches improve the result, which are even better than PLS based methods in the case of false alarm rates (FAR). PLS presents fine fault detection rates (FDR) for KPI-related faults while creates significant FAR. As for T-PLS and C-PLS based methods, they both cover two indices, then PLS based detection has been enhanced by integrating the KPI weakly related part caused by PLS. So far, we draw the conclusion that, associated with KPI issue, if the FDR is much more focused, the methods in PLS family can be employed, whereas, if FAR is superiorly concerned, LS could be an effective alternative.

7. CONCLUSION

In this paper, the KPI-related process monitoring issues have been stressed of which a uniform projection modus

Table 8. KPI-related detection indices

Fault	KPI affected		DD	LS1	LS2	PLS	T-PLS		C-PLS	
	$\varepsilon(KPI)$	$\sigma^2(KPI)$	$T^2_{\theta y}$	$T^2_{\hat{y}}$	$T^2_{\hat{\theta}}$	T^2_{PLS}	$T^2_{\hat{y}_{PLS,\theta}}$	Q_{T-PLS}	$T^2_{y\theta}$	Q_{C-PLS}
1	√	√	0.9962	0.9487	0.9988	0.9988	0.9938	0.9988	0.9938	0.9988
2	√	√	0.9738	0.9675	0.9775	0.9950	0.9125	0.9938	0.9125	0.9938
4	√	×	0.7650	1	1	1	1	1	1	1
6	√	√	0.9930	1	0.9789	1	0.9930	1	0.9930	1
7	√	×	1	1	1	1	1	1	1	1
8	√	√	0.9625	0.9263	0.9600	0.9875	0.9350	0.9888	0.9350	0.9950
10	√	×	0.7275	0.8862	0.6900	0.8912	0.8325	0.9500	0.8325	0.9550
11	√	×	0.7963	0.8950	0.9437	0.9762	0.9013	0.9487	0.9013	0.9587
12	×	√	0.2637	0.2400	0.2025	0.4850	0.2512	0.4988	0.2512	0.2800
13	√	√	0.9738	0.9862	0.9587	0.9888	0.9862	0.9875	0.9862	0.9862
17	√	√	0.7662	0.9163	0.8600	0.9700	0.9125	0.9850	0.9125	0.9838
18	×	√	0.5112	0.6650	0.4713	0.7738	0.6550	0.8313	0.6550	0.8237
19	√	×	0.9387	0.9263	0.8525	0.9875	0.9275	0.9925	0.9275	0.9988
20	√	√	0.9775	0.9600	0.9425	0.9850	0.8488	0.9878	0.8488	0.9875
3	×	×	0.0500	0.0825	0.0500	0.1000	0.0900	0.1737	0.0900	0.1238
5	×	×	0.0488	0.0575	0.0512	0.0462	0.0512	0.0537	0.0512	0.0462
9	×	×	0.0712	0.0988	0.0638	0.1437	0.1050	0.1963	0.1050	0.0850
14	×	×	0.7250	0.9738	0.8400	0.9962	0.9875	0.9975	0.9875	0.9975
15	×	×	0.0600	0.0587	0.0537	0.0635	0.0537	0.0862	0.0537	0.0675
16	×	×	0.0500	0.0450	0.0512	0.0575	0.0462	0.0563	0.0462	0.0488

has been formed for the considered approaches in multivariate statistical process monitoring field. Three categories: direct SVD based one, modified LS regression based one and PLS based one are presented for the first time. The mechanisms of them have been explored to show how they are designed for monitoring KPI-related and -unrelated parts of process variables. Meanwhile, the comparison between them are shown in the aspects of computational intensity and detecting performance. TEP benchmark experiment has been implemented to demonstrate some valuable conclusions.

REFERENCES

S.X. Ding, S. Yin, K.X. Peng, H.Y. Hao, B. Shen. A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip mill. *IEEE Transactions on Industrial Informatics*, 9:2239–2247,2013.

S.X. Ding. *Model-based fault diagnosis techniques*. Springer, Berlin, 2nd edition, 2013.

N.D. Tracy, J.C. Young and R.L. Mason. Multivariate control charts for individual observations. *Journal of Quality Technology*, 24:88–95, 1992.

J.F. MacGregor, C. Jaeckle, C. Kiparissides, and M. Koutoudi. Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal*, 40:826–838 1994.

S.J. Qin. Statistical process monitoring: Basics and beyond. *Journal of Chemometrics*, 17:480–502, 2003.

S.J. Qin. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36: 220–234, 2012.

S. Wold, M. Sjstrma, L. Erikssonb. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58: 109–130, 2001.

G. Li, S.J. Qin, D.H. Zhou. Geometric properties of partial leasts quares for process monitoring. *Automatica*, 46:204–210, 2010.

G. Li, S.J. Qin, D.H. Zhou. Output relevant fault reconstruction and fault subspace extraction in total projection to latent structures models. *Industry Engineering Chemistry Research*, 49:9175–9183, 2010.

S.X. Ding, P. Zhang, E.L. Ding, S. Yin, A. Naik, P. Deng, and W. Gui. On the application of PCA technique to fault diagnosis. *Tsinghua Science and Technology*, 15:138–144, 2010.

H.Y. Hao, K. Zhang, S.X. Ding, Z.W. Chen, Y.G. Lei. A KPI-related multiplicative fault diagnosis scheme for industrial processes. *IEEE proceedings of the 10th international conference on control & automation*, Hangzhou (China), June 2013.

S.J. Qin. Recursive PLS algorithms for adaptive data modeling. *Computers and Chemical Engineering*, 22: 503–514, 1998.

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

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

S. Yin, S.X. Ding, P. Zhang, A. Hagahni, A. Naik. Study on modifications of PLS approach for process monitoring. *18th IFAC World Congress*, Milano (Italy), August 2011.

G. Box. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. effect of inequality of variance in the one-way classification. *The annals of mathematical statistics*, 25:290–302, 1954.

L.H. Chiang, E.L. Russell, R.D. Braatz. Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and intelligent laboratory systems*, 50:243–252, 2000.

L.H. Chiang, E.L. Russell, R.D. Braatz, *Fault detection and diagnosis in industrial systems*. Springer, London, 2nd edition, 2001.

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. *Journal of Process Control*, 22:1567–1581, 2012.