Data Driven Autonomous Fault Detection and Identification Methodologies for Honeywell Process Solutions

- Phase 1 Progress Report: DPCA and DPLS Based Anomaly Detection

Bahador Rashidi and Qing Zhao
Department of Electrical and Computer Engineering
Univ. of Alberta, Edmonton, Alberta, Canada

Date: Sept. 19, 2016

ABSTRACT

The material of this report is centered around fault detection and is focused on data-driven techniques. The strategy is to utilize available sensor measurements as well as process outputs, and apply a real-time statistical analysis to detect any existing abnormality including KPI-related and KPI-unrelated faults upon their immediate occurrences. KPI represents the key performance indicator. There are many statistical methods for fault detection and diagnosis, for example, see surveys in [1] and [2]. This report aims to demonstrate the applicability of the *improved dynamic schemes* of two commonly adopted methods (namely principal component analysis, in short, PCA; and partial least square, in short, PLS). These two schemes are called dynamic PCA (DPCA) and dynamic improved PLS (DIPLS). A brief review of the main results of DPCA and DIPLS is presented first. Moreover, a numerical example is given to demonstrate the fundamental concepts. Then the two scheme are applied to the well-known Tennessee Eastman benchmark system, and the detection results are presented to show the performance of DPCA and DIPLS.

Keywords: data-driven, statistical multivariate analysis, PCA, PLS, dynamic PCA (DPCA), dynamic improved PLS (DPLS)

I. DYNAMIC PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA-based methods have been used widely for abnormality detection in process control. In this section, dynamic PCA is presented and its difference with the ordinary PCA is highlighted. Moreover, a combined statistical index computed from the two well-known $T^2_{Hoteling}$ and SPE (i.e. squared prediction error) is derived and its upper control limits is shown. One of the applications of the PCA is to reduce the dimensionality of a set of data while keeping key features of it. Consider $X_{Original} \in R^{N\times m}$ as a matrix of process variables, arranged such that columns are different measurements and rows are observation values in different sample times. $Y \in R^{N\times l}$ is the matrix of outputs that sometimes are considered as key performance indicators (KPIs) in a process.

Owing to the fact that in majority of industrial processes, process variables have dynamic relations with each other, ordinary PCA may not be the appropriate monitoring tool as it is originally designed for a set of variables that have static and linear relations. Therefore, a method that considers the dynamic feature of the process under study should be more applicable. In this report, it is proposed to use dynamic PCA and dynamic IPLS for the phase 1 anomaly detection problem defined in this project, which is focused on fault detection based on using available sensor measurements. One would challenge the suggested idea by saying that both PCA and DPCA or PLS and DIPLS are defined for linear systems and most of the industrial processes are intrinsically nonlinear. The justification for this question is that any nonlinear system can be deemed linear by proper segmentation between different modes of the process.

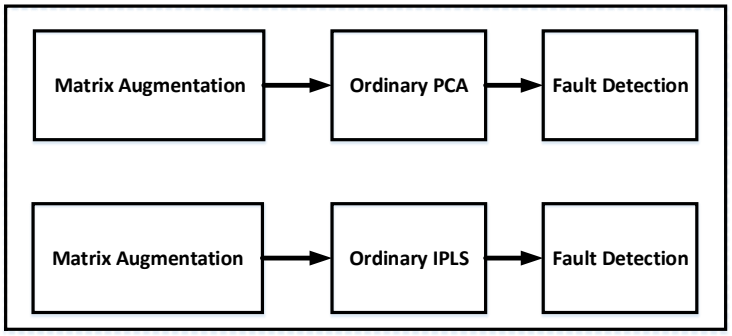


Fig. 1. Augmentation steps added to the ordinary PCA and IPLS to incorporate time correlation into process variables

In a dynamic process, variables are time correlated. Or in another word, evolutions of process variables are recursive and regressive. This means that the value of each process variable at current time is dependent on the past values of itself or other variables. For incorporating this feature, it is proposed in [3] to build an augmented matrix X from $X_{Orginal}$ as $X_i = [X_i^T \ X_{i-\tau}^T \ ... \ X_{i-(l-1)\tau}^T]$. As shown in Figure (I), the main difference of DPCA from PCA, and DIPLS from IPLS is the augmentation step in order to consider the dynamic feature of variables. In the above, τ is a time interval, which allows for scaling of the augmented matrix in time, and l is the number of sample lag that can be determined by an iterative method in [3]. From this point, the augmented matrix X is considered as the new set of process data that is able to reveal time correlation of process variables as well as their correlations with each other.

The underlying idea of PCA is to extract the dominant principal components (PCs) of a data-set. Singular value decomposition (SVD) in matrix algebra is a common tool to derive PCs, which is also deemed in this report. In this phase of the project, PCA is considered as a mechanism to check the presence of abnormality in sensor and process variable measurements. For this purpose, first of all, the matrix X needs to be mean-centered with the unit variance. Then by using SVD on $\Phi = \frac{1}{\sqrt{N-1}}X$ as in (1), the so-called loading and score matrices are determined.

$$\Phi = USV^T \tag{1}$$

S is a diagonal matrix including singular values of matrix Φ arranged in a descending order. V is an orthonormal matrix that is called loading matrix of ϕ. The number of latent variables or principal components of matrix X is p_{com} that is the smallest number satisfying the inequality,

$$\sum_{i=1}^{p \leq K} \frac{\delta_i \times 100}{\sum_{j=1}^{j=K} \delta_j} \geq CPV \tag{2}$$

where CPV is the cumulative percentage value and δ_i are singular values of Φ. Two corresponding loading vectors for matrix X can be defined as $\hat{P} = V_1 \in R^{p_{com} \times m}$ and $\tilde{P} = V_2 \in R^{m-p_{com} \times m}$. By using $\hat{P}$, the new data-set in principal and residual subspaces are $X_{PC} = X\hat{P}$ and $X_{RES} = X\tilde{P}$, respectively. Equation (3) shows two projection matrix $\hat{\Pi}$ and $\tilde{\Pi}$ that can decompose data matrix X into two parts aligned with PCs and residue subspace.

$$X = \hat{X} + \tilde{X} = X\hat{\Pi} + X\tilde{\Pi} = X\hat{P}\hat{P}^T + X\tilde{P}\tilde{P}^T \tag{3}$$

As can be concluded from (3), $\hat{X}$ is the part of original process data matrix X in principal planes. If a malfunction is present in a process, whether it is along with PCs or residual subspace, it might affect one or both of $\hat{X}$ and $\tilde{X}$. In order to detect presence of such anomalies, in literature, two indices, namely

T^2 Hotelling and square prediction error (SPE), are defined. By calculating upper control limits (UCL) for both indices as thresholds for the normal behaviour of the process or no-fault situation, faults can be detected when these thresholds (UCL) are violated. (4) and (5) show the formula for the calculation of T^2 and SPE and their UCL.

$$\begin{aligned} T^2_{Hotelling} &= ||\hat{X}||^2 = X\hat{P}S^{-2}\hat{P}^T X \\ UCL_{T^2} &= \frac{p_{com}(N-1)(N+1)}{N^2 - Np_{com}} F_{\alpha,p_{com},N-p_{com}} \end{aligned} \tag{4}$$

$$\begin{aligned} SPE &= ||\tilde{X}||^2 = ((I-\hat{\Pi})X)^T((I-\hat{\Pi})X) \\ UCL_{SPE} &= \theta_1 \left(\frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2} + 1 \right)^{1/h_0} \end{aligned} \tag{5}$$

$F_{\alpha,p_{com},N-p}$ is the upper α percentile of the F distribution with p_{com} and $N - p_{com}$ degree of freedom and $\theta_i = \sum_{j=p+1}^{K} \lambda_j^i$ for $i = 1, 2, 3$. After determining θ_i, h_0 can be calculated as $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$. c_α is a standard normal variable corresponding to the upper $1-\alpha$ percentile. Alcala in [4] demonstrated that T^2 and SPE residuals can not cover all possibilities of anomalies in process variables if adopted individually, and proposed a combined residual signal that can cover all possible abnormalities. By following the procedure in [4], a combined residual index is derived as follows,

$$\varphi = \frac{T^2_{Hoteling}}{UCL_{T^2}} + \frac{SPE}{UCL_{SPE}} \tag{6}$$

The control limit for φ is

$$\zeta^2 = g^\varphi \chi^2_\alpha(h^\varphi)$$

where $g^\varphi = (\frac{1}{UCL^2_{T^2}} + \frac{1\theta_2}{UCL^2_{SPE}})/(\frac{1}{UCL_{T^2}} + \frac{\theta_1}{UCL_{SPE}})$ and $h^\varphi = (\frac{1}{UCL_{T^2}} + \frac{\theta_1}{UCL_{SPE}})^2/(\frac{1}{UCL^2_{T^2}} + \frac{1\theta_2}{UCL^2_{SPE}})$ with $(1-\alpha)\times 100$ confidence level. The detailed derivation of this control limit can be found in [2].

Figure (2) shows the diagram of the fault detection using DPCA. When the combined residual φ exceeds the ζ^2 as its threshold, it means that there may be anomalies present in the process.Therefore, a binary trigger mechanism will switch the isolation module into the monitoring procedure.

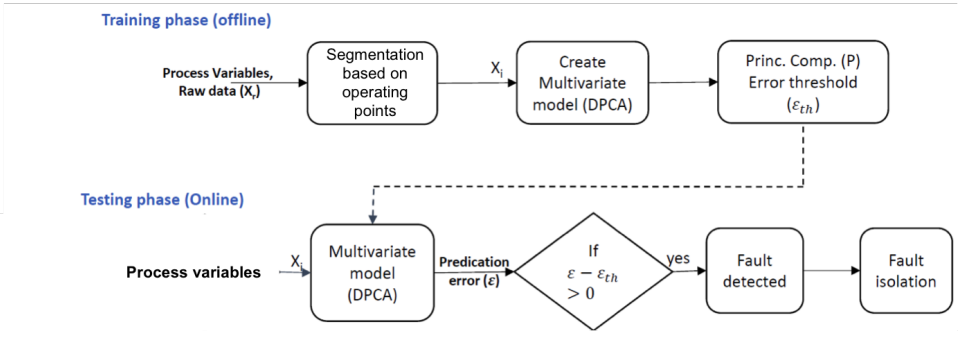


Fig. 2. Schematic diagram of fault detection based on process measurements

Remark 1: In this report, it is proposed to employ DPCA on sensor measurements for monitoring their performance and detecting anomalies in the sensor level. In other words, up to this point DPCA can be used on a dataset, given in matrix X, to detect sensor anomalies.

II. Dynamic Improved Partial Least Square (DIPLS)

Like PCA, PLS-based methods are widely used for monitoring purposes [5], [6]. This report suggests to use a dynamic version of improved PLS known as DIPLS [7]. As mentioned in the previous section, DPCA is a module operating on a sole set of data, including process variables and outputs. However, if two separate data matrices are available such as process variables and key performance indicators (KPIs), and users intend to monitor faults and their influence on KPIs, a PLS-based method is an appropriate choice. Authors of this report suggest to use DIPLS not only to detect abnormality in the process, but also determine whether the detected fault influences KPI variables or not. There are also other methods that can determine KPI-related and unrelated faults in a process such as [8]. However, DIPLS has advantages by considering dynamic relation of variables with each other, and using total orthonormal decomposition of process variables to estimate the KPI variable(s). The following is a brief summary for the DIPLS scheme.

Similar to dynamic PCA, augmented versions of process variables $X_{original} \in R^{N\times m}$ and KPIs $Y \in R^{N\times l}$ are constructed as follows,

$$X_b = \begin{bmatrix} x_{h+1}^T & x_h^T & ... & 1 \\ x_{h+2}^T & x_{h+1}^T & ... & x_2^T \\ ... & ... & ... & ... \\ x_N^T & x_{N-1}^T & ... & x_{N-h}^T \end{bmatrix} \in R^{(N-h)\times(mh+m)}, \quad Y_b = \begin{bmatrix} y_{h+1}^T \\ y_{h+2}^T \\ . \\ . \\ . \\ y_N^T \end{bmatrix} \in R^{(N-h)\times l} \tag{7}$$

where h is the order of the time delay for the process variables. As it is shown in figure (I), this augmentation incorporates the time-correlation to the process and is the main difference between ordinary PLS and DIPLS. By using X_b and Y_b, (8) provides the PLS model.

$$\begin{aligned} X_b &= \hat{X}_b + \tilde{X}_b = TP_b^T + \tilde{X}_b \\ Y_b &= \hat{Y}_b + \tilde{Y}_b = TQ_b^T + \tilde{Y}_b \end{aligned} \tag{8}$$

Let us assume that the matrix Ψ is the correlation matrix between process variables X_b and KPIs Y_b. Therefore, from the real-time implementation point of view, if it is not possible to measure KPIs in real-time, the correlation matrix Ψ can be used to estimate the KPI as follows,

$$\hat{Y}_b = X_b\Psi \tag{9}$$

Due to the fact that in ordinary PLS model (8), the decomposition is based on the orthogonal projection, the fault in process variable is not properly decomposed and projected onto Y_b-related and unrelated subspaces. To achieve this goal, a complete orthonormal decomposition on X_b and Y_b is necessary. Hence the following decomposition is desired to be reached.

$$\begin{aligned} X_b &= \hat{X}_b + \tilde{X}_b \\ Y_b &= \hat{Y}_b + \tilde{Y}_b \\ \hat{Y}_b &= X_b\Psi = \hat{X}_b\Psi \end{aligned} \tag{10}$$

It can be proved that the correlation matrix Ψ can be determined as (11).

$$\Psi = (X_b^T X_b)^{-1} X_b^T Y_b \tag{11}$$

In order to achieve (10), SVD is implemented on $\Psi\Psi^T$ as follows,

$$\Psi\Psi^T = \begin{bmatrix} \hat{\Gamma}_\Psi & \tilde{\Gamma}_\Psi \end{bmatrix} \begin{bmatrix} \Lambda_\Psi & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\Gamma}_\Psi^T \\ \tilde{\Gamma}_\Psi^T \end{bmatrix} \tag{12}$$

where, $\hat{\Gamma}_\Psi \in R^{m(h+1)\times l}$, $\tilde{\Gamma}_\Psi \in R^{m(h+1)\times(m(h+1)-l)}$ and $\Lambda_\Psi \in R^{l\times l}$. The corresponding projection matrices are as follows,

$$\begin{aligned} \Pi_\Psi &= \hat{\Gamma}_\Psi \hat{\Gamma}_\Psi^T \\ \Pi_\Psi^\perp &= \tilde{\Gamma}_\Psi \tilde{\Gamma}_\Psi^T \end{aligned} \tag{13}$$

Therefore, by using these two projection matrices, X_b can be written as,

$$\begin{aligned} \hat{X}_b &= X_b \Pi_\Psi = T_{\hat{X}_b} \hat{\Gamma}_\Psi \in \hat{S} \equiv span\{\Psi\} \\ \tilde{X}_b &= X_b \Pi_\Psi^\perp = T_{\tilde{X}_b} \tilde{\Gamma}_\Psi \in \tilde{S} \equiv span\{\Psi\}^\perp \end{aligned} \tag{14}$$

As can be seen from equation (14), the new score matrix correlated with the KPIs is $\hat{X}_b = X_b \hat{\Gamma}_\Psi$ that is aligned with KPI variables and can be monitored for presence of any KPI-related malfunctions (i.e. those affecting KPIs). It should be mentioned that $\tilde{X}_b = X_b \tilde{\Gamma}_\Psi$ is the part of process variables that is uncorrelated with KPIs. Hence, by defining a monitoring index such as T^2 and SPE on each of $\hat{X}_b$ and $\tilde{X}_b$, faults can be detected.

$$\begin{aligned} \hat{T}^2_{X_b} &= X^T_{new_b} \hat{\Gamma}_\Psi \left(\frac{\hat{\Gamma}_\Psi^T X^T_{new_b} X_{new_b} \hat{\Gamma}_\Psi}{N-h-1} \right) \hat{\Gamma}_\Psi^T X_{new_b} \\ \tilde{T}^2_{X_b} &= X^T_{new_b} \tilde{\Gamma}_\Psi \left(\frac{\tilde{\Gamma}_\Psi^T X^T_{new_b} X_{new_b} \tilde{\Gamma}_\Psi}{N-h-1} \right) \tilde{\Gamma}_\Psi^T X_{new_b} \end{aligned} \tag{15}$$

where, X_{new_b} contains the real-time process variables that can be used for calculating the real-time index signal for monitoring purposes. It should be mentioned that the matrix Ψ and correspondingly matrices $\hat{\Gamma}_\Psi$ and $\tilde{\Gamma}_\Psi$ are derived off-line in the training process from clean (no fault) data. Equation (16) shows the quality related and unrelated upper control limit (UCL) for the determined indices.

$$\begin{aligned} UCL_{\hat{T}_{X_b}} &= \frac{l(N_h^2-1)}{N_h(N_h-l)} F_\alpha(l, N_h-l) \\ UCL_{\tilde{T}_{X_b}} &= \frac{(m_h-l)(N_h^2-1)}{N_h(N_h-m_h+l)} F_\alpha(m_h-l, N_h-m_h+l) \end{aligned} \tag{16}$$

where, $N_h = N-h$ and $m_h = m(h+1)$.

Remark 2: The explained DIPLS method is suitable for KPI-related fault detection. If $\hat{T}^2_{X_b} \geq UCL_{\hat{T}_{X_b}}$, it can be concluded that a fault has occurred, which affects key performance indicators. If $\tilde{T}^2_{X_b} \geq UCL_{\tilde{T}_{X_b}}$ and $\hat{T}^2_{X_b} < UCL_{\hat{T}_{X_b}}$, it indicates the presence of a (hidden) 'fault' that is KPI-unrelated.

A. Numerical example

In this section a numerical example is considered to show the procedure and performance of DIPLS for dynamic systems. The governing equation of the numerical example is given as (17).

$$\begin{aligned} t_k &= a_1 t_{k-1} - a_2 t_{k-2} + t_k^\star \\ x_k &= P t_k + e_k \\ y_k &= C_1 x_k + C_2 x_{k-1} + v_k \end{aligned} \tag{17}$$

where,

$$a_1 = \begin{bmatrix} 0.4389 & 0.1210 & -0.0862 \\ -0.2966 & -0.0550 & 0.2274 \\ 0.4538 & -0.6573 & 0.4239 \end{bmatrix}, \; a_2 = \begin{bmatrix} -0.2998 & -0.1905 & -0.2669 \\ -0.0204 & -0.1585 & -0.2950 \\ 0.1461 & -0.0755 & 0.3475 \end{bmatrix}$$

$$P = \begin{bmatrix} 0.5586 & 0.2042 & 0.6370 \\ 0.2007 & 0.0492 & 0.4429 \\ 0.0874 & 0.6062 & 0.0664 \\ 0.9332 & 0.5463 & 0.3743 \\ 0.2594 & 0.0958 & 0.2491 \end{bmatrix}, \; C_1 = \begin{bmatrix} 0.9249 & 0.4350 \\ 0.6295 & 0.9811 \\ 0.8783 & 0.0900 \\ 0.6417 & 0.5275 \\ 0.7984 & 0.5456 \end{bmatrix}^T, \; C_2 = \begin{bmatrix} 1.7198 & -0.3715 \\ 0.5835 & 1.5011 \\ 1.4236 & 1.3226 \\ 0.4963 & -1.4145 \\ -2.5117 & 1.0696 \end{bmatrix}^T \tag{18}$$

$t_k^\star \sim \nu(0, 2^2 I_3)$, $e_k \sim (0, 0.01^2 I_5)$, $v_k \sim \nu(0, 0.1^2 I_2)$. The fault of the process is added as $x_k = x_k^\star + x_f$ such that the $x_k^\star$ is the no fault state and x_f is the additive process fault. In the simulation the first 1000 samples are used for training and the second 1000 samples used for testing the algorithm. the fault is is assumed to occur at the 500th sample of the testing set. Two fault scenarios are defined for the simulation,

1) Case 1: In this scenario, $X_f^\star = [2,\ 1,\ 3,\ -2,\ -5]$ that is a quality-related fault. As can be seen from figure (3(a)), KPI-related residual violated the upper control limit that refers to presence of such malfunction. Figure (3(b)) indicates the unrelated-KPI fault as well. Figure (3(c), 3(d)) is the estimation of the Y using X.

2) Case 2: In this scenario, $X_f^\star = [0.0054,\ 0.3145,\ -0.0432,\ 0.7516,\ -0.4440]$ is an KPI-unrelated fault introduced to the process. Figure (3(e)) shows that the KPI-related index does not cross the threshold, while the KPI-unrelated index in figure (3(f)) shows the presence of quality unrelated fault. From these two figures, it can be concluded that there is possibly a malfunction present in the process, which does not affect the KPI variables. In this case, one may choose to ignore it from the economical point of view.

DIPLS is able to successfully detect faults in a dynamic process as well as determining whether that fault is KPI-related/unrelated.

III. Results with Tenessee Estman (TE) Model Data

TE process model is a well-known simulation benchmark for validation of process control and data-monitoring techniques. This paper adopts the TE model originally given in [9] as a realistic simulation of its industrial prototype. Here, only essential information for the following case studies is provided and readers are referred to [9] for more details about the system's units and components. TE process has eight major components: gaseous reactants A, C, D, E and inert B as input of the main reactions as well as liquid outputs G and H. The total number of measurements is 41, i.e., XMEAS (1-41) and it has 12 manipulated variables, i.e., XMV (1-12). 22 out of 41 process measurements are continuous and the rest are sampled process measurements. The data-set used in this paper is from [10] that includes 21 different fault scenarios. All faults were introduced to the process at the 8th (161th sample) operation hour. 22 sets of training and testing data are considered such that one set is obtained from normal situation (no fault) that is used for building base-line model and 21 sets are run for 21 different fault scenarios. XMEAS (1-22) are considered as $(n = 22)$ outputs and XMV (1-11) are $(m = 11)$ inputs that is $(m + n = 33)$ total measurements.

A schematic of the TE model is presented in figure (4). Table (3) shows the process manipulated variables of TE process. Continuous process measurements and sampled process measurements are presented in Tables (4) and (5), respectively.

In the TE process, the final product component G (XMEAS (35)) is selected as the KPI variable y and another 33 variables, XMEAS (1-22) and XMV (1-11), are selected as process variable x. We simulate the TE model as if it runs for 72 hours with fault introduced at the 20th hour (the 2001st

sample). The sampling rate is 100 samples/hr. Faults are represented by IDV(0,1,2,,20) with different IDVs being different types of faults (Table (8)).

As mentioned before, DIPLS can be used for two purposes; fault detection and checking whether the present fault affects the KPI variables or not. However, the DPCA, can be used for fault detection in the sensor level. Figures (9(a)) to (9(f)) present some examples of using DPCA on TE benchmark for the fault IDV8. More results are available upon request.

IV. Discussion of simulations result

By using DPCA for monitoring purposes, if malfunction detection is the goal of users, T^2 and SPE statistics are the main outcome of the method and can detect presence of fault. The estimation of each process variable using DPCA does not contain any useful information for fault detection purposes and it can only be used for validation of the principal planes extraction from process variables. On the other side, DIPLS has one more advantage compared to DPCA, which is the decomposition of the detected malfunction into two types; KPI-related and unrelated. DIPLS is also able to estimate the KPI variables based on a correlation matrix that is calculated form the training part of the process. Figures (7(f), 7(i), 8(c), 5(f), 8(f), 5(i), 6(c), 6(f), 6(i), 7(c), 5(c)) are the estimated KPI for different fault scenarios. In some cases like (7(c), 6(i), 6(f)), the estimated KPI matches more with the real value. It can be concluded that when there is a KPI-unrelated fault in a process, the estimated KPI is relativity more consistent and accurate. However, from the point of view of fault detection, there is no need for users to expect the estimated value to be accurate for no the fault situation since the presence of fault can be detected by the provided residual indices as shown in figures (7(d),7(g), 8(a), 5(g), 8(d), 6(a), 6(d), 6(g), 7(a), 5(a)). It should also be mentioned that there is a need for estimating KPI variables when their measurements are not possible in real-time and monitoring the KPI variables is necessary.

References

[1] A. H. H. H. P. Z. Shen Yin, Steven X. Ding, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567 – 1581, 2012.

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

[3] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, no. 1, pp. 179 – 196, 1995.

[4] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring," *Automatica*, vol. 45, no. 7, pp. 1593 – 1600, 2009.

[5] R. Muradore and P. Fiorini, "A pls-based statistical approach for fault detection and isolation of robotic manipulators," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 8, pp. 3167–3175, 2012.

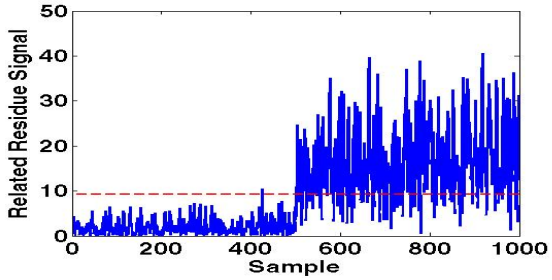
[6] S. Yin, S. X. Ding, P. Zhang, A. Hagahni, and A. Naik, "Study on modifications of {PLS} approach for process monitoring," {*IFAC*} *Proceedings Volumes*, vol. 44, no. 1, pp. 12 389 – 12 394, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667016456089

[7] J. Jiao, H. Yu, and G. Wang, "A quality-related fault detection approach based on dynamic least squares for process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 4, pp. 2625–2632, 2016.

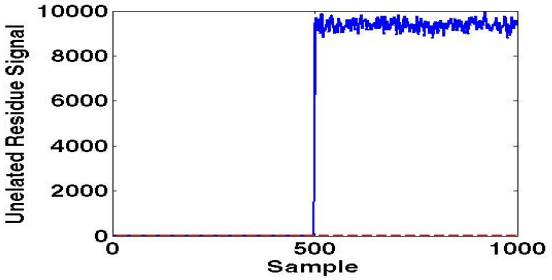
[8] D. Zhou, G. Li, and S. J. Qin, "Total projection to latent structures for process monitoring," *AIChE Journal*, vol. 56, no. 1, pp. 168–178, 2010.

[9] J. Downs and E. Vogel, "Industrial challenge problems in process control a plant-wide industrial process control problem," *Computers and Chemical Engineering*, vol. 17, no. 3, pp. 245 – 255, 1993.

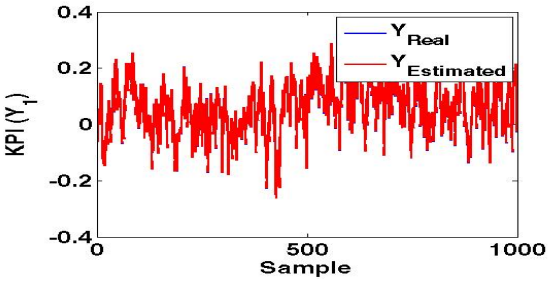
[10] R. D. B. L. H. Chiang and E. L. Russell, "Fault detection and diagnosis in industrial system," *New York, NY, USA, Springer*, 2001.



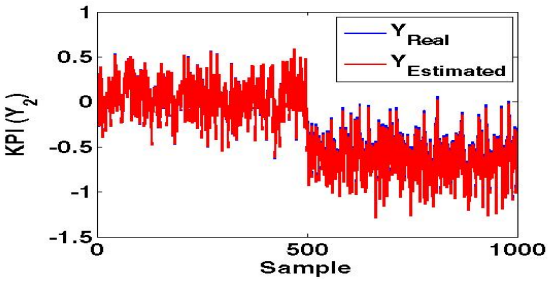
(a) Numerical example, Case 1, KPI-related index



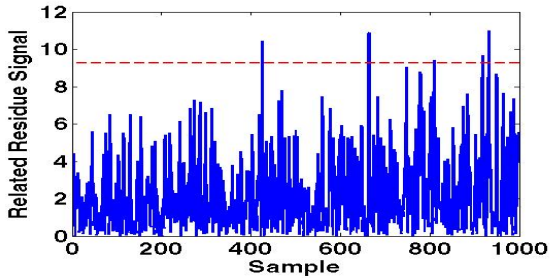
(b) Numerical example, Case 1, KPI-unrelated index



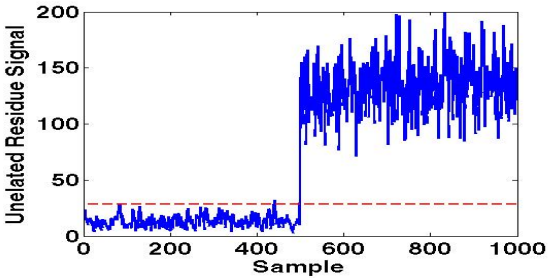
(c) Numerical example, Case 1, $\hat{Y}_1$



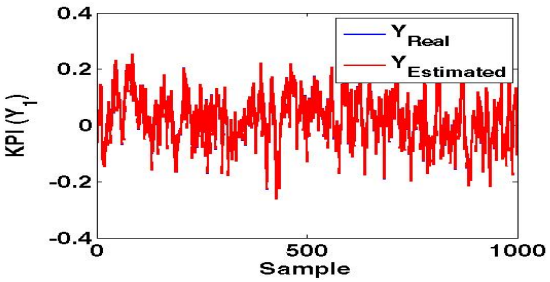
(d) Numerical example, Case 1, $\hat{Y}_2$



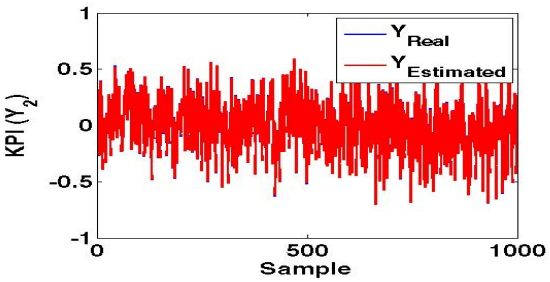
(e) Numerical example, Case 2, KPI-related index



(f) Numerical example, Case 2, KPI-unrelated index



(g) Numerical example, case 2, $\hat{Y}_1$



(h) Numerical example, Case 2, $\hat{Y}_2$

Fig. 3. Simulation results for numerical example introduced to KPI-related/unrelated fault

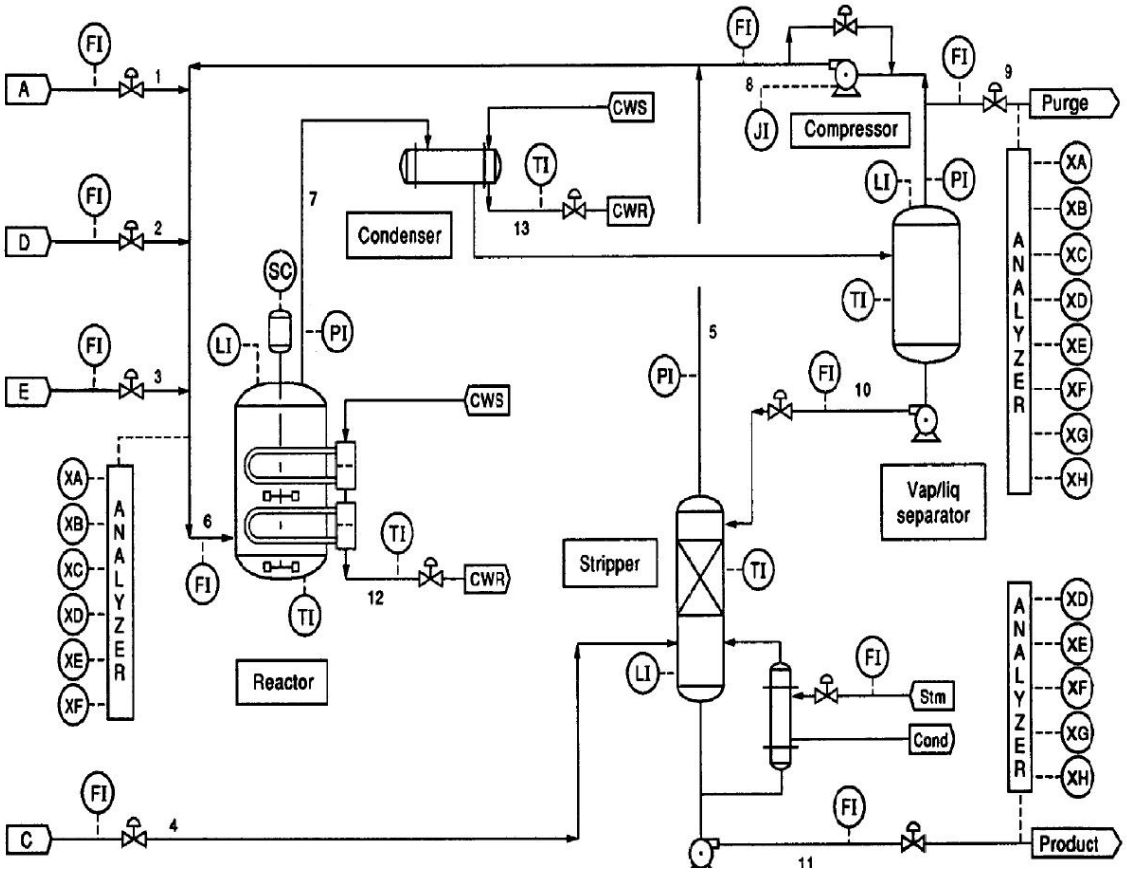


Fig. 4. Schematic diagram showing components and units of TE process

Table 3. Process manipulated variables

Variable name	Variable number	Base case value (%)	Low limit	High limit	Units
D feed flow (stream 2)	XMV (1)	63.053	0	5811	kg h^{-1}
E feed flow (stream 3)	XMV (2)	53.980	0	8354	kg h^{-1}
A feed flow (stream 1)	XMV (3)	24.644	0	1.017	kscmh
A and C feed flow (stream 4)	XMV (4)	61.302	0	15.25	kscmh
Compressor recycle valve	XMV (5)	22.210	0	100	%
Purge valve (stream 9)	XMV (6)	40.064	0	100	%
Separator pot liquid flow (stream 10)	XMV (7)	38.100	0	65.71	$m^3 h^{-1}$
Stripper liquid product flow (stream 11)	XMV (8)	46.534	0	49.10	$m^3 h^{-1}$
Stripper steam valve	XMV (9)	47.446	0	100	%
Reactor cooling water flow	XMV (10)	41.106	0	227.1	$m^3 h^{-1}$
Condenser cooling water flow	XMV (11)	18.114	0	272.6	$m^3 h^{-1}$
Agitator speed	XMV (12)	50.000	150	250	rpm

Table 4. Continuous process measurements

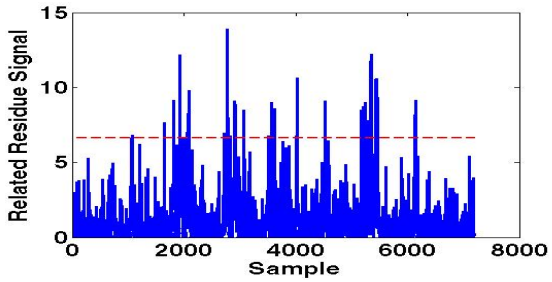
Variable name	Variable number	Base case value	Units
A feed (stream 1)	XMEAS (1)	0.25052	kscmh
D feed (stream 2)	XMEAS (2)	3664.0	kg h^{-1}
E feed (stream 3)	XMEAS (3)	4509.3	kg h^{-1}
A and C feed (stream 4)	XMEAS (4)	9.3477	kscmh
Recycle flow (stream 8)	XMEAS (5)	26.902	kscmh
Reactor feed rate (stream 6)	XMEAS (6)	42.339	kscmh
Reactor pressure	XMEAS (7)	2705.0	kPa gauge
Reactor level	XMEAS (8)	75.000	%
Reactor temperature	XMEAS (9)	120.40	°C
Purge rate (stream 9)	XMEAS (10)	0.33712	kscmh
Product separator temperature	XMEAS (11)	80.109	°C
Product separator level	XMEAS (12)	50.000	%
Product separator pressure	XMEAS (13)	2633.7	kPa gauge
Product separator underflow (stream 10)	XMEAS (14)	25.160	$m^3 h^{-1}$
Stripper level	XMEAS (15)	50.000	%
Stripper pressure	XMEAS (16)	3102.2	kPa gauge
Stripper underflow (stream 11)	XMEAS (17)	22.949	$m^3 h^{-1}$
Stripper temperature	XMEAS (18)	65.731	°C
Stripper steam flow	XMEAS (19)	230.31	kg h^{-1}
Compressor work	XMEAS (20)	341.43	kW
Reactor cooling water outlet temperature	XMEAS (21)	94.599	°C
Separator cooling water outlet temperature	XMEAS (22)	77.297	°C

Table 5. Sampled process measurements

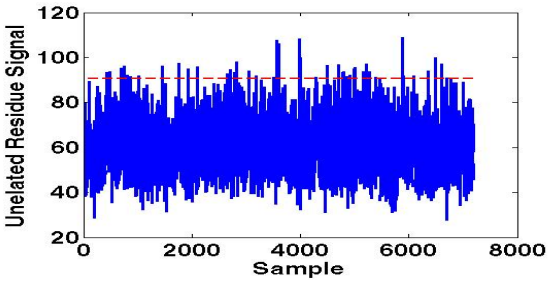
Reactor feed analysis (stream 6)				
Component	Variable number	Base case value	Units	Sampling frequency = 0.1 h Dead time = 0.1 h
A	XMEAS (23)	32.188	mol%	
B	XMEAS (24)	8.8933	mol%	
C	XMEAS (25)	26.383	mol%	
D	XMEAS (26)	6.8820	mol%	
E	XMEAS (27)	18.776	mol%	
F	XMEAS (28)	1.6567	mol%	
Purge gas analysis (stream 9)				
Component	Variable number	Base case value	Units	Sampling frequency = 0.1 h Dead time = 0.1 h
A	XMEAS (29)	32.958	mol%	
B	XMEAS (30)	13.823	mol%	
C	XMEAS (31)	23.978	mol%	
D	XMEAS (32)	1.2565	mol%	
E	XMEAS (33)	18.579	mol%	
F	XMEAS (34)	2.2633	mol%	
G	XMEAS (35)	4.8436	mol%	
H	XMEAS (36)	2.2986	mol%	
Product analysis (stream 11)				
Component	Variable number	Base case value	Units	Sampling frequency = 0.25 h Dead time = 0.25 h
D	XMEAS (37)	0.01787	mol%	
E	XMEAS (38)	0.83570	mol%	
F	XMEAS (39)	0.09858	mol%	
G	XMEAS (40)	53.724	mol%	
H	XMEAS (41)	43.828	mol%	

Table 8. Process disturbances

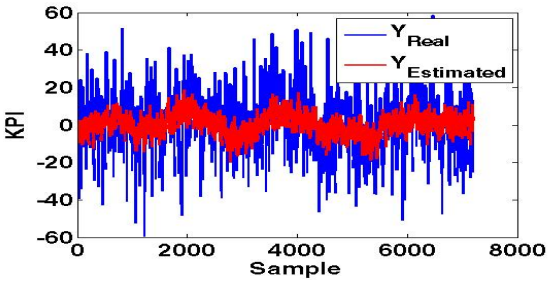
Variable number	Process variable	Type
IDV (1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV (2)	B composition, A/C ratio constant (stream 4)	Step
IDV (3)	D feed temperature (stream 2)	Step
IDV (4)	Reactor cooling water inlet temperature	Step
IDV (5)	Condenser cooling water inlet temperature	Step
IDV (6)	A feed loss (stream 1)	Step
IDV (7)	C header pressure loss—reduced availability (stream 4)	Step
IDV (8)	A, B, C feed composition (stream 4)	Random variation
IDV (9)	D feed temperature (stream 2)	Random variation
IDV (10)	C feed temperature (stream 4)	Random variation
IDV (11)	Reactor cooling water inlet temperature	Random variation
IDV (12)	Condenser cooling water inlet temperature	Random variation
IDV (13)	Reaction kinetics	Slow drift
IDV (14)	Reactor cooling water valve	Sticking
IDV (15)	Condensor cooling water valve	Sticking
IDV (16)	Unknown	Unknown
IDV (17)	Unknown	Unknown
IDV (18)	Unknown	Unknown
IDV (19)	Unknown	Unknown
IDV (20)	Unknown	Unknown



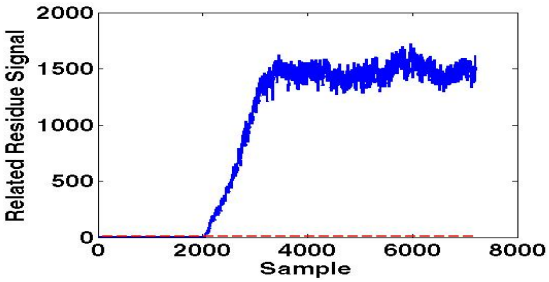
(a) KPI-related residue signal for no fault situation



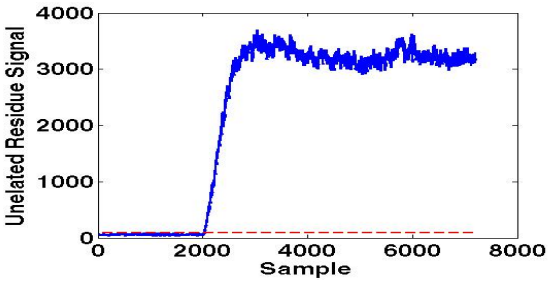
(b) KPI-unrelated residue signal for no fault situation



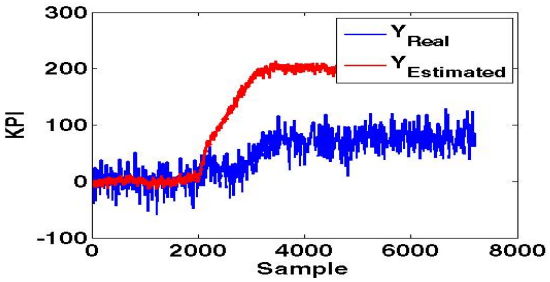
(c) KPI estimation using process signal



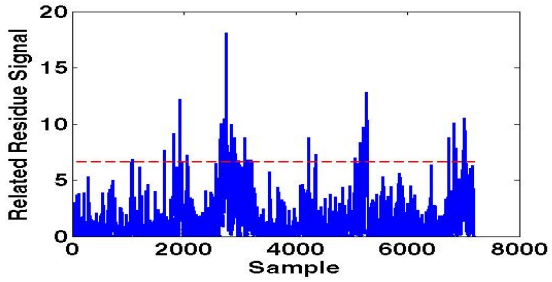
(d) KPI-related residue signal for fault IDV2



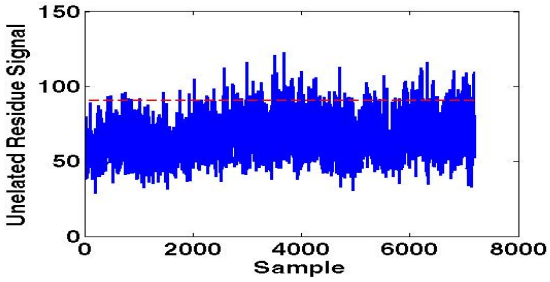
(e) KPI-unrelated residue signal for fault IDV2



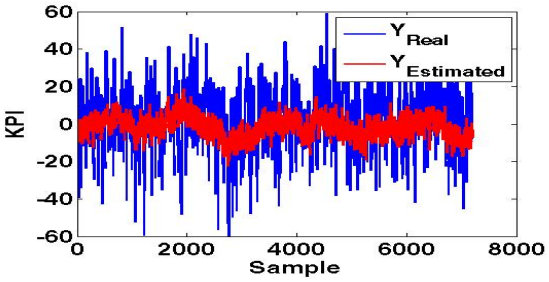
(f) KPI estimation using process signal with presence of IDV2



(g) KPI-related residue signal for fault IDV3

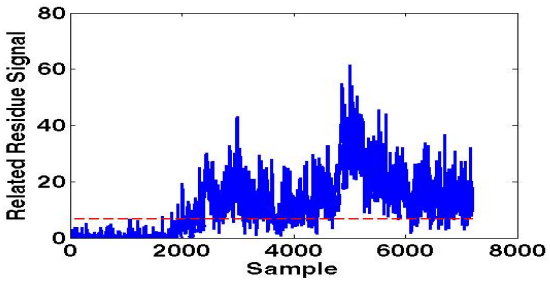


(h) KPI-unrelated residue signal for fault IDV3

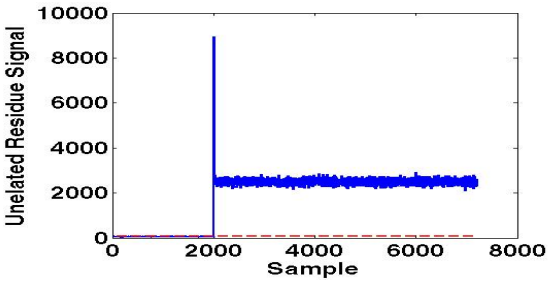


(i) KPI estimation using process signal with presence of IDV2

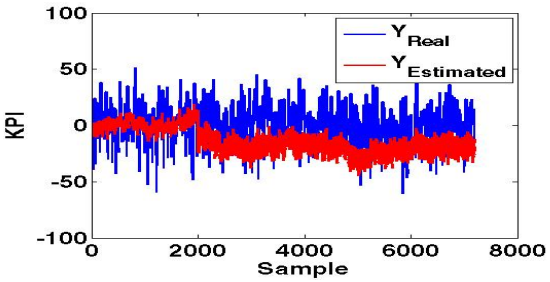
Fig. 5. Simulation result for TE benchmark using DIPLS for different fault scenarios presented in Table (8)



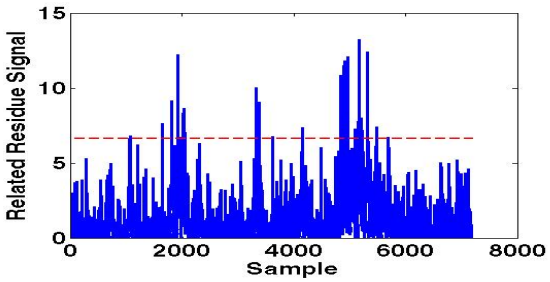
(a) KPI-related residue signal for fault IDV4



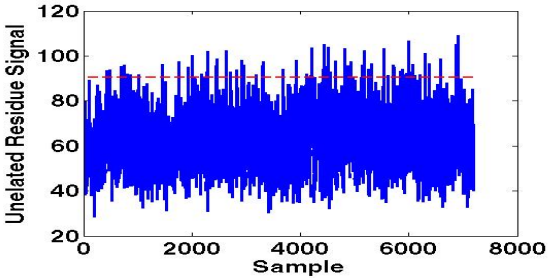
(b) KPI-unrelated residue signal for fault IDV4



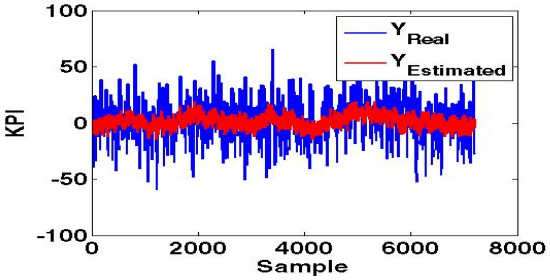
(c) KPI estimation using process signal with presence of IDV4



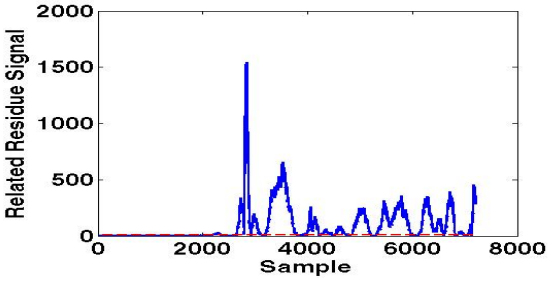
(d) KPI-related residue signal for fault IDV5



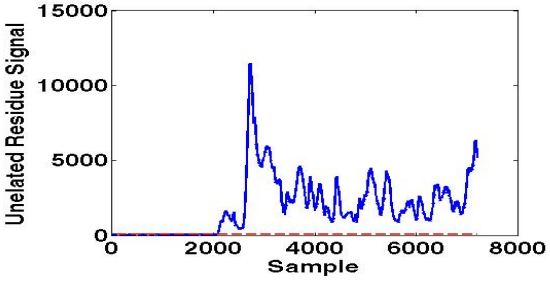
(e) KPI-unrelated residue signal for fault IDV5



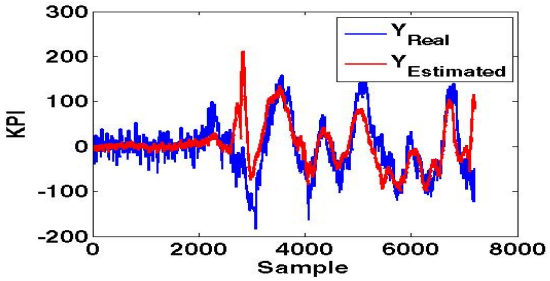
(f) KPI estimation using process signal with presence of IDV5



(g) KPI-related residue signal for fault IDV8

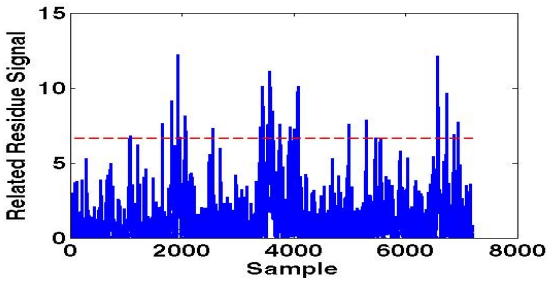


(h) KPI-unrelated residue signal for fault IDV8

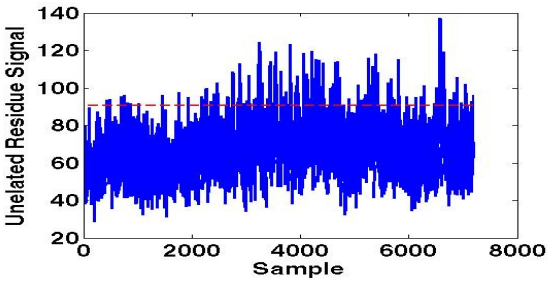


(i) KPI estimation using process signal with presence of IDV8

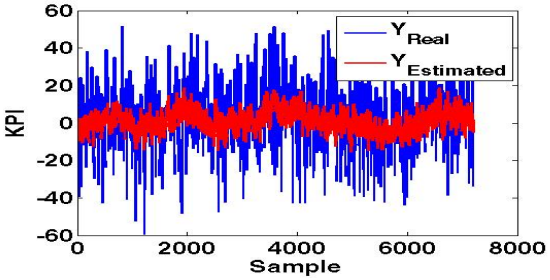
Fig. 6. Simulation result for TE benchmark using DIPLS for different fault scenarios presented in Table (8)



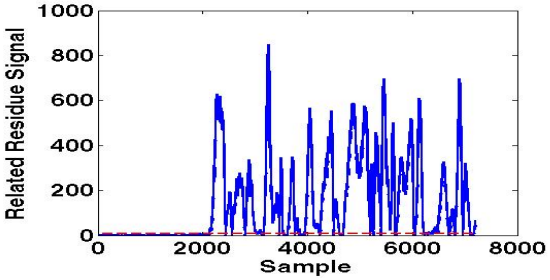
(a) KPI-related residue signal for fault IDV9



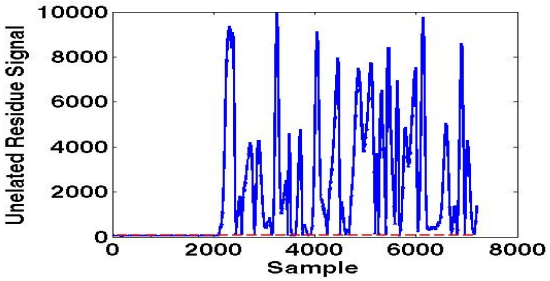
(b) KPI-unrelated residue signal for fault IDV9



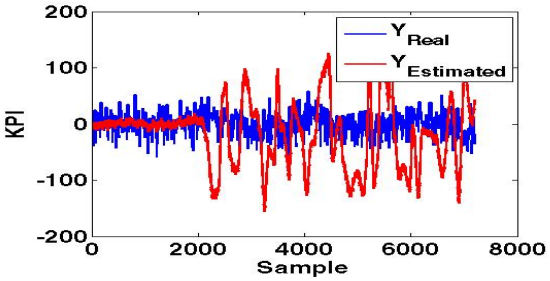
(c) KPI estimation using process signal with presence of IDV9



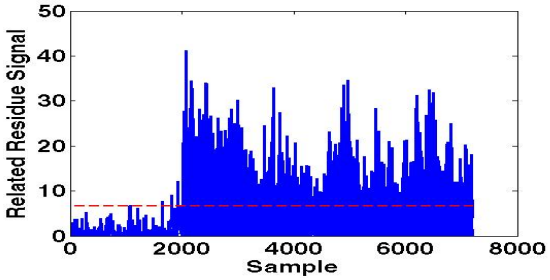
(d) KPI-related residue signal for fault IDV10



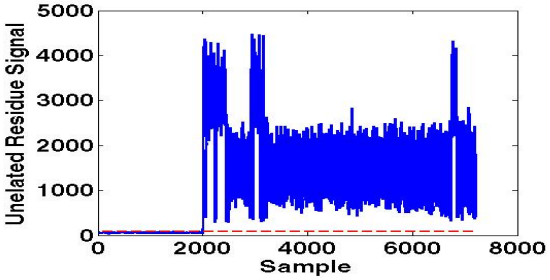
(e) KPI-unrelated residue signal for fault IDV10



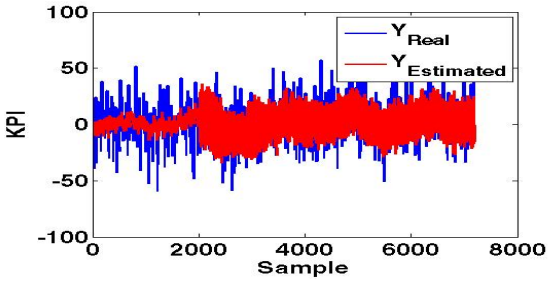
(f) KPI estimation using process signal with presence of IDV10



(g) KPI-related residue signal for fault IDV14

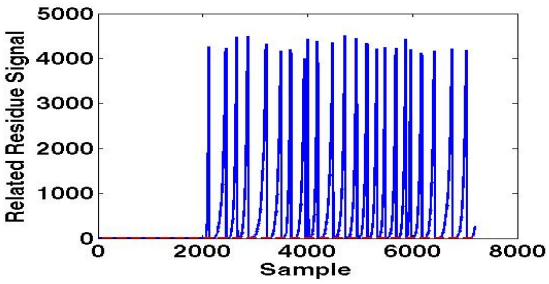


(h) KPI-unrelated residue signal for fault IDV14

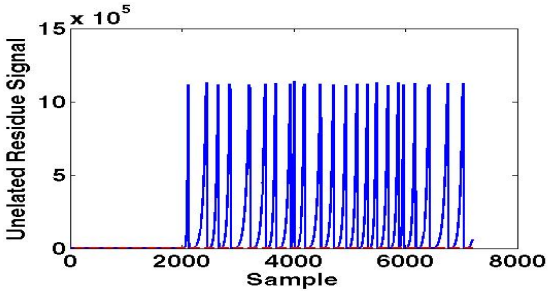


(i) KPI estimation using process signal with presence of IDV14

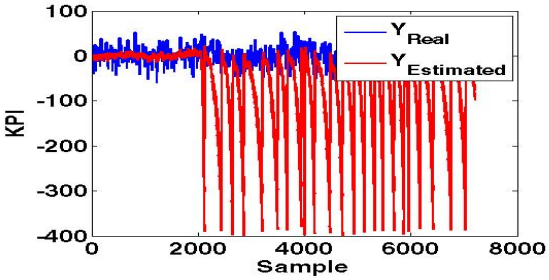
Fig. 7. Simulation result for TE benchmark using DIPLS for different fault scenarios presented in Table (8)



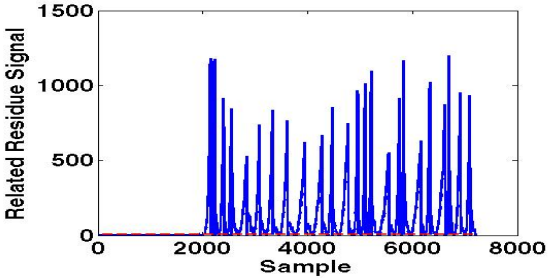
(a) KPI-related residue signal for fault IDV17



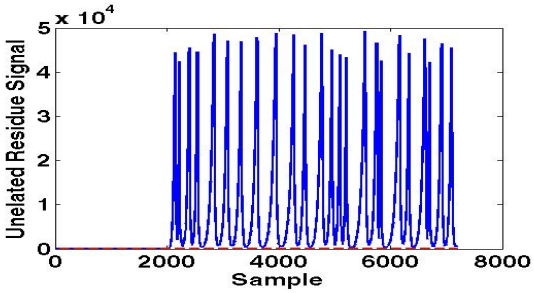
(b) KPI-unrelated residue signal for fault IDV17



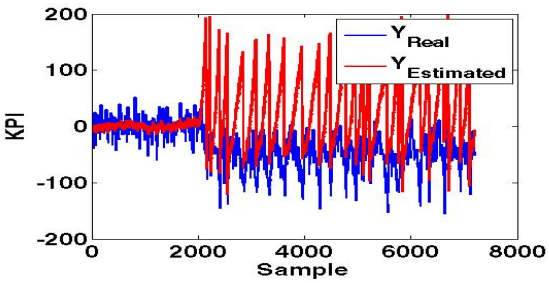
(c) KPI estimation using process signal with presence of IDV17



(d) KPI-related residue signal for fault IDV20

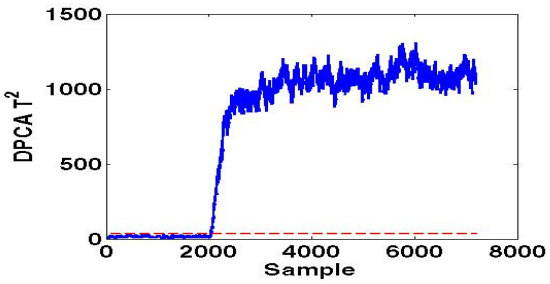


(e) KPI-unrelated residue signal for fault IDV20

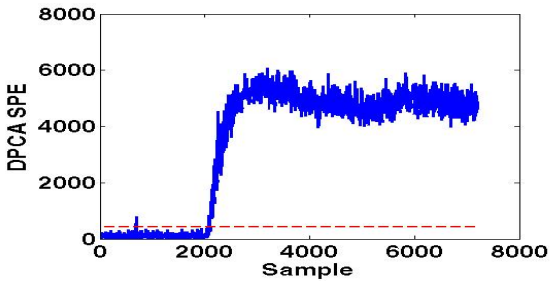


(f) KPI estimation using process signal with presence of IDV20

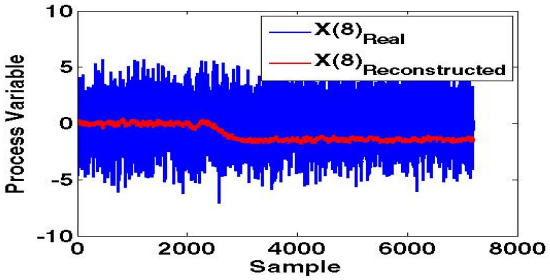
Fig. 8. Simulation result for TE benchmark using DIPLS for different fault scenarios presented in Table (8)



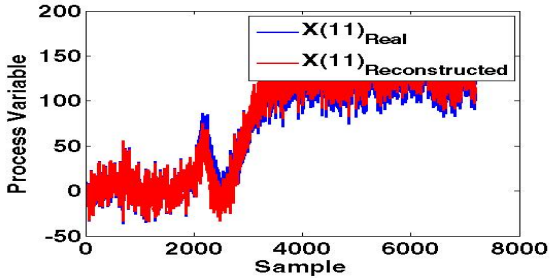
(a) T^2 statistic of DPCA employed on TE with fault IDV8



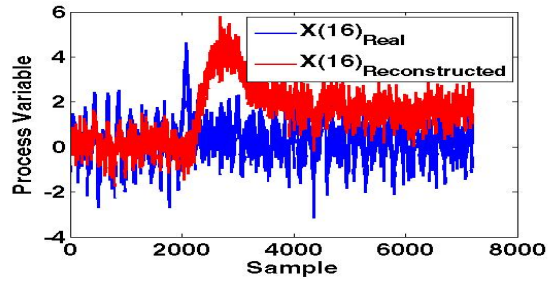
(b) SPE statistic of DPCA employed on TE with fault IDV8



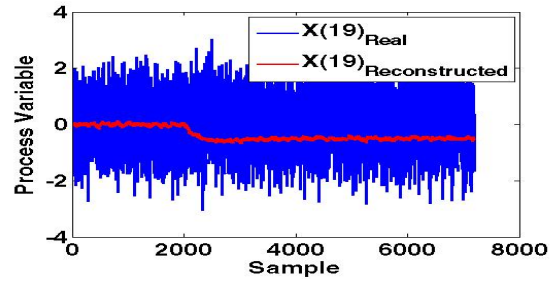
(c) Estimation of X_8 using principal loading of TE process variables



(d) Estimation of X_{11} using principal loading of TE process variables



(e) Estimation of X_{16} using principal loading of TE process variables



(f) Estimation of X_{19} using principal loading of TE process variables

Fig. 9. Simulation result for TE benchmark using DPCA for for fault IDV8