Enhanced Anomaly Detection via PLS Regression Models and Information Entropy Theory

Fouzi Harrou, *Member*, *IEEE*, Ying Sun
CEMSE Division, King Abdullah University of Science and Technology, Saudi Arabia
fouzi.harrou@kaust.edu.sa, ying.sun@kaust.edu.sa

***Abstract*—Accurate and effective fault detection and diagnosis of modern engineering systems is crucial for ensuring reliability, safety and maintaining the desired product quality. In this work, we propose an innovative method for detecting small faults in the highly correlated multivariate data. The developed method utilizes partial least square (PLS) method as a modelling framework, and the symmetrized Kullback-Leibler divergence (KLD) as a monitoring index, where it is used to quantify the dissimilarity between probability distributions of current PLS-based residual and reference one obtained using fault-free data. The performance of the PLS-based KLD fault detection algorithm is illustrated and compared to the conventional PLS-based fault detection methods. Using synthetic data, we have demonstrated the greater sensitivity and effectiveness of the developed method over the conventional methods, especially when data are highly correlated and small faults are of interest.**

I. Introduction

Process safety and product quality are two crucial issues for modern industrial processes. Fault detection (FD) and diagnosis play an important role from the view point of improving product quality and enhancing process safety. As a result of proper process monitoring, downtime is minimized, safety of process operations is improved, and manufacturing costs are reduced [1]. Of course, monitoring process can be defined as the set of actions carried out to detect, isolate faulty measurement sources and then remove them before they affect the process performance [1]. The purpose of fault detection is to identify any fault event indicating a distance from the process behavior compared to its nominal behavior. Whereas fault isolation is used to determine the location of the detected fault. This work focuses on fault detection.

Data-based process monitoring methods, also known as process history based methods or model-free methods[2], [3], are able to extract useful information from the historical data, computing the relationship between the variables without the need for an analytical model. Towards this end, data-based monitoring methods rely on the availability of historical data obtained from the monitored process under nominal operating conditions[4]. The fault-free data are first used to build an empirical model that describes the nominal process's behavior, which is then used to detect faults in future data. Then, the empirical model is used to estimate true values of new measurements, and faults are detected and diagnosed. Since no explicit models are required, the development of which is usually costly or time consuming, data-based methods has become very popular in industrial processes. However, the performance of data-based methods mainly depends on the availability of quantity and quality of the input data.

Various data-based FD techniques are referenced in the bibliography, and they can be broadly categorized into two main classes: univariate and multivariate techniques [4]. The univariate statistical monitoring methods such as EWMA (exponentially weighted moving average) chart and CUSUM (cumulative sum) chart are essentially used to monitor only one process variable [4], [5]. However, modern industrial processes often present a large number of highly correlated process variables. This is the area where univariate FD methods are unable to explain different aspects of the process and, therefore, it is not appropriate for modern day processes [4]. Moreover, to monitor several different process variables in the same time multivariate statistical monitoring were developed. A multivariate FD methods take into account the correlation between a process variables while univariate FD methods do not. Particularly, for multivariate process monitoring purpose, the latent variable regression (LVR)-based method has received much attention in last decades. The main idea of the LVR-based monitoring approach (e.g., partial least square (PLS) regression, principal component analysis (PCA)) is to extract the useful data information from the original data set, and construct some statistics for monitoring [6], [7], [8], [8].

PLS also known as projection to latent structure is among the most widely used multivariate statistical process monitoring (MSPM) method for monitoring multivariate processes [9]. PLS attempts to decompose the data in such a way that the correlation between predictor and predicted variables are maximized [10]. By extracting the useful data information from the original dataset, and then using monitoring indices such as T^2 and Q statistics, PLS has been used successfully for fault detection in multivariate process with highly correlated variables. PLS-based process monitoring method as well its variants has been largely exploited and used to different engineering applications [7], [11], [12].

Detecting small or incipient faults in highly correlated multivariate data is one of the most crucial and challenging tasks in the area of fault detection and diagnosis. Indeed, early detection of small or incipient faults can provides an early warning and helps to avoid catastrophic damage and subsequent financial loss. Unfortunately, the conventional

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.64

PLS-based monitoring indices such as T^2 and Q statistic are less efficient in detecting small faults as they use only the information enclosed about the process in the last observation to take decision. These shortcomings of the T^2 and Q statistics motivate the use of other alternatives in order to mitigate these disadvantages. To do so, this paper is focused in developing enhanced PLS fault detection method by using Kullback-Liebler divergence (KLD) [13]. Such a choice is mainly motivated by the greater ability of the KLD metric to evaluate the similarity between two distributions, which makes it very attractive as monitoring index. In particular, the KLD is used as a measure to quantify the similarity between the probability distributions of actual PLS-based residuals and reference one. Of course, the main contribution of this work is to exploit the advantages of the KLD index and those of PLS modeling for enhancing detection performances of conventional PLS, especially for detecting small faults in highly correlated multivariate data.

The reminder sections of this paper is organized as follows. In section II, a brief introduction to PLS and how it can be used in fault detection are reviewed. In Section III, the Kullback-Liebler divergence is briefly described. Section IV describes the proposed PLS-based KLD fault detection approach. Next, in Section V, we assess the proposed scheme and present some simulation results. Finally, some concluding remarks and future research directions are given in Section VI.

II. PLS-Based Process Monitoring

PLS is a well reputed multivariate statistical technique for dimensionality reduction of process data [9]. The key role of a linear PLS is based on its capability to deal with collinear data, and several variables in both the input (predictor) matrix $\mathbf{X}$ and output (response) matrix $\mathbf{Y}$ [10]. In its general form PLS finds the latent variables from the process data by capturing the largest variability in the data and achieves the maximum cross-correlation among the predictor and the response variables [10]. Given an input data matrix $\mathbf{X} \in R^{n\times m}$ having n observations and m variables, and an output data matrix $\mathbf{Y} \in R^{n\times p}$ consisting of p response variables, a PLS model is formally determined by two sets of linear equations: the inner model and the outer model [10]. The inner model represents the relations between the latent variables (LVs), and the outer model represents the relations linking the LVs and their associated observed variables [14]. More details on the PLS algorithms can be found in [10], [15].

The outer model, which links LVs, and the response and predictor matrices, can be expressed as [14]:

$$\begin{cases} \mathbf{X} = \widehat{\mathbf{X}} + \mathbf{E} = \sum_{i=1}^{l} \mathbf{t}_i \mathbf{P}_i^T + \mathbf{E} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} = \widehat{\mathbf{Y}} + \mathbf{F} = \sum_{i=1}^{l} \mathbf{u}_i \mathbf{q}_i^T + \mathbf{F} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} \end{cases} \tag{1}$$

where $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{Y}}$ represent modeling matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively, the matrices $\mathbf{T} \in R^{n\times l}$ and $\mathbf{U} \in R^{n\times q}$ consist of l kept LVs of the predictor and response data, respectively, the matrices $\mathbf{E} \in R^{n\times m}$ and $\mathbf{F} \in R^{n\times p}$ represent the approximation residuals of the predictor and response data, respectively, and the matrices $\mathbf{P} \in R^{m\times l}$ and $\mathbf{Q} \in R^{p\times q}$ represent the predictor and response loading matrices, respectively. The number of LVs, l, can be estimated by using cross-validation or some other techniques [16].

$$\mathbf{U} = \mathbf{T}\mathbf{B} + \mathbf{H}, \tag{2}$$

where $\mathbf{B}$ represents a regression matrix which consists of the model parameters linking the predictor and response LVs, and $\mathbf{H}$ represents a residual matrix. The response $\mathbf{Y}$ can now be expressed as: $\mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{Q}^T + \mathbf{F}^*$. Of course, the PLS method projects the data down to a number of LVs that explain most of the variation in both predictors and responses and then models the LVs by a linear regressions.

For PLS-based monitoring, two statistics, Hotelling's T^2 and Q or squared prediction error (SPE), are generally used [4]. The Hotelling's T^2 is a statistical metric that captures the behavior of retained LVs [17], and is defined as [17]:

$$T^2 = \hat{\mathbf{t}}^T \widehat{\Lambda}^{-1} \hat{\mathbf{t}}, \tag{3}$$

where $\widehat{\Lambda} = \frac{1}{n-1} \widehat{\mathbf{T}}^T \widehat{\mathbf{T}}$, is the covariance matrix of the l retained LVs. For new testing data, when the T^2 value of exceeds the control limit, it can be concluded that the process is out of control [17].

The Q statistic, on the other hand, shows deviations from normal operating conditions based on variations in the predictor variables that are not described by the PLS model [10]. Indeed, the $Q^{(X)}$ statistic quantifies the loss of fit with the PLS model developed and is defined as [18]:

$$Q^{(X)} = \left\| \mathbf{x} - \widehat{\mathbf{x}} \right\|_2^2, \tag{4}$$

where $\widehat{\mathbf{x}}$ represents the prediction of $\mathbf{x}$ by the PLS model. When a vector of new data is available, the Q statistic is calculated and compared with the threshold value Q_α given in [18]. A fault is detected if the Q statistics exceeds a threshold Q_α. The PLS-based fault detection algorithm is summarized next.

1) **Given**:
   - A training fault-free data set ($\mathbf{X}$ and $\mathbf{y}$) that represents the normal process operations and a testing data set (possibly faulty data),
2) **Data preprocessing**
   - Scale the data that is used for process model building, to zero mean and unit variance,
3) **Build the PLS model using the training data**
   - Select the number of latent variables using cross validation or any other model selection method,
   - Express the data matrix as a sum of approximate and residual matrices as shown in equation (1),
   - Compute the control limits for the statistical model (e.g., the Q_α statistic limits),
4) **Test the new data**

- Scale the new data with the mean and standard deviation obtained from the training data,
- Compute the residuals of the response variables, **F**,
- Compute the monitoring statistic (Q or T^2 statistics) for the new data using equation (4) or (3),

5) **Check for faults**

- Declare a fault when new data exceeds the control limits (e.g., $Q > Q_\alpha$).

In conventional PLS-based monitoring, two statistics, T^2 and Q, are generally used for FD [4]. Although the two methods have their advantages and disadvantages, both tend to fail to detect small faults [4], because they treat each observation individually and don't take into account information from past data. That makes them insensitive to small fault in the process variables and causes many missed detections. To overcome these limitations of the conventional PLS-based monitoring methods, we have developed an alternative fault detection approach, in which PLS is used as a modeling framework for fault detection using KLD metric. More details about KLD metric, and how it can be used in fault detection are presented next.

III. Kullback-Leibler Divergence-Based Monitoring Metric

The Kullback-Leibler information (KLI) also well-known in information theory as the relative entropy is an important statistical measure that can be used to quantify the dissimilarity, separability, distinguishability or closeness between two probability density functions (PDFs) [13], [19]. It plays a key role in various problems of statistical inference and data processing such as detection, estimation, compression, classification [20], [21].

Given two pdf's $p(x)$ and $q(x)$ which are assumed to have the same support, the KLI between p versus q is a measure of the information lost when p is used to approximate q, and defined as

$$I(p/\!\!/q) = \sum_{x\in\mathscr{X}} p(x)\ln\frac{p(x)}{q(x)}. \tag{5}$$

As a matter of fact, the KLI measure is not a distance or metric in the Euclidean sense, basically because the distance between two distributions generally is not symmetric function of p and q, i.e. $I(p/\!\!/q)$ is not equal to $I(q/\!\!/p)$), and the triangle inequality is not satisfied. Hence it must be interpreted as a pseudo-distance measure only. It is non-negative and null only when the two densities are equal. A familiar symmetrized version of the Information is called the Kullback *divergence*, which will be used in this paper as fault detection index. Therefore, the KL divergence (KLD) between $p(x)$ and $q(x)$ is given by:

$$J(p;q) = I(p/\!\!/q) + I(q/\!\!/p). \tag{6}$$

Assuming that $p(x)$ and $q(x)$ are two univariate normal distributions, that is, $p\sim\mathscr{N}(\mu_0,\sigma_0)$ and $q\sim\mathscr{N}(\mu_1,\sigma_1)$, the KLD can be simplified as follows:

$$J(p,q) = \frac{1}{2}\left(\frac{\sigma_1^2}{\sigma_0^2} + \frac{\sigma_0^2}{\sigma_1^2} + (\mu_1-\mu_0)^2\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma_1^2}\right) - 2\right) \tag{7}$$

If $\sigma_1^2 = \sigma_0^2$, the the Equation (7) can be rewritten as:

$$J(p,q) = \frac{(\mu_1-\mu_0)^2}{\sigma_0^2}. \tag{8}$$

Two similar distributions will have a small KLD close to zero, while very different distributions would have a larger KLD. It is this comparison operation which makes it well useful index to anomaly detection. The KLD can be used as a fault indicator by comparing the statistical similarity between the residuals distributions before and after anomaly. Therefore, it seems meaningful to adopt rather a PLS-based KLD technique for the statistical process monitoring. In the next section, the KLD metric will be integrated with PLS to extend its fault detection abilities for detecting small faults.

IV. Fault Detection Using PLS-Based KLD Technique

In this section, PLS modelling framework is integrated with KLD statistical index to develop a new fault detection scheme with a higher sensitivity to small faults in the data. Towards this end, the KLD is used to measure the divergence from the reference distribution (using fault-free data) of residuals of the responses variables obtained from the PLS model, in order to detect anomalies. When the inspected system is under normal operating conditions (no faults), the divergence measures between the current distribution of residuals and reference distributions is zero or close to zero in cases of modeling uncertainties and measurement noise. However, when a fault occurs, the KLD between the distributions before and after change deviate significantly from zero indicating the presence of a new condition that is significantly distinguishable from the normal faultless working mode. Thus, this work exploits the advantages of the KLD to improve fault detection over the conventional PLS-based methods. The detailed procedure will be exposed in the subsequent section.

A. PLS-Based KLD Monitoring Algorithm

As shown in Equation (1) of outer model, the output vector $\mathbf{y}$ can be written as the sum of an approximated vector $\widehat{\mathbf{y}}$ and a residual vector $\mathbf{F}$, i.e.,

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{F}. \tag{9}$$

The difference between the observed value of the input variable, y, and the predicted value, $\widehat{y}$, obtained from PLS model represent the residual of the output variable, $\mathbf{F} = [f_1,\ldots,f_t,\ldots,f_n]$ which can be used as an indicator to detect a possible fault. The residual $\mathbf{F}$ obtained from PLS model is assumed to be Gaussian. In this paper, the problem of detecting additive faults (or more specifically incipient faults) is addressed. It is assumed that the fault affect the mean parameter of residual distributions and the variance is supposed unchanged after the fault occurrence. The KLD measures the distance of probability distribution of current residual $p_0(f) \sim \mathcal{N}(\mu_1, \sigma_0^2)$ against a reference one $p_1(f) \sim \mathcal{N}(\mu_0, \sigma_0^2)$, where μ_0 and μ_1 are the means and $\sigma_0^2 > 0$ is the variance for $p_0(f)$ and $p_1(f)$. Then, the KLD decision function based on the residuals distributions of the response variable can be computed as follows: $J(p_0(f), p_1(f)) = \frac{(\mu_1-\mu_0)^2}{\sigma_0^2}$. Where μ_0 and σ_0 are the mean and the standard deviation of PLS-based residuals obtained with fault free data. The KLD-based test makes decision between the null hypothesis $\mathcal{H}_0$ (absence of faults) and alternative hypothesis $\mathcal{H}_1$ (existence of faults) by comparing between the decision statistic $J(p_0(f), p_1(f))$ and a given value of the threshold T_J.

$$J(p_0(f), p_1(f)) = \frac{(\mu_1 - \mu_0)^2}{\sigma_0^2} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} T_J. \quad (10)$$

For setting the detection threshold T_J, a simple approach based on the three-sigma rule was used.

$$T_J \;=\; \mu_0^{T_J} + L\sigma_0^{T_J} \quad (11)$$

where $\mu_0^{T_J}$ is the mean value, $\sigma_0^{T_J}$ is the standard deviation of the nominal KLD value obtained with fault-free data, and L is the width of the control limitswhich determines the confidence limits, usually specified in practice as 3 for a false alarm rate of 0.27%.

If the decision function $J(p_0(f), p_1(f))$ is larger than the threshold T_J, the KLD-based test decides for $\mathcal{H}_1$, otherwise $\mathcal{H}_0$ is assumed to be true. The steps of the PLS-based KLD anomaly detection algorithm are summarized in Figure 1.

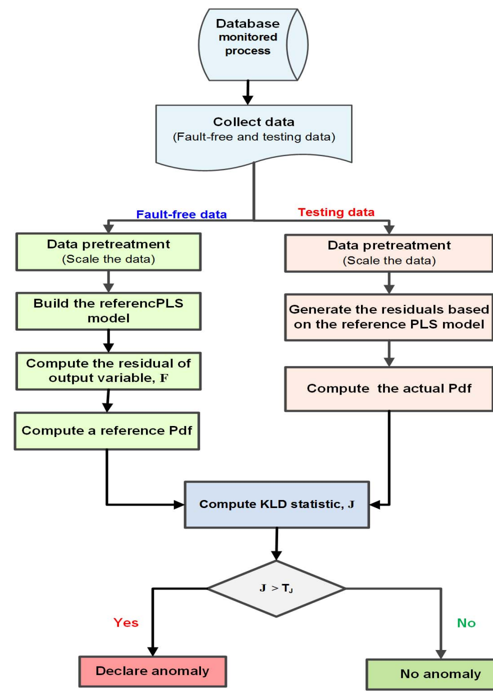

Fig. 1. A block diagram of the PLS-based KLD anomaly detection algorithm. KLD is used to compare the statistical distribution of a PLS-based residual against that of a reference. If the divergence, J , is greater than some prede?ned threshold, T_J, the presence of anomaly is declared.

In the next section, the performance of the PLS-based KLD fault detection method will be evaluated and compared to that of the conventional PLS fault detection scheme through simulated example using synthetic data.

V. Simulated Examples

A. Fault detection in synthetic data

In this section, the performance of the proposed PLS-based KLD fault detection algorithm is assessed through its utilization to detect faults in synthetic data sets which contained several different types of fault scenarios. We also conducted the same tests for the standard PLS method and compared the results with each other.

1) Data generation: The simulated data used in this example consists of six input variables and one output. The input variables are generated as follows:

$$\mathbf{x}_1 = \mathbf{u}_1 + \varepsilon_1;\quad \mathbf{x}_2 = \mathbf{u}_1 + \varepsilon_2;\quad \mathbf{x}_3 = \mathbf{u}_1 + \varepsilon_3;\quad \mathbf{x}_4 = \mathbf{u}_2 + \varepsilon_4;$$
$$\mathbf{x}_5 = 2\mathbf{u}_2 + \varepsilon_5;\quad \mathbf{x}_6 = 2\mathbf{x}_1 + 2\mathbf{x}_4 + \varepsilon_6; \quad (12)$$

where, ε_i, represent measurements errors, which follow a zero-mean Gaussian distribution having a standard deviation of 0.095. The first two input variables u_1 and u_2 represent a quad-chirp signal (sinusoidal waveform with quadratically increasing frequency) and a mishmash signal (this signal starts with low frequency oscillations but frequency increases as time goes on), respectively. The other input variables are computed as linear combinations of the first two inputs, which means that the input matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_6]$ is of rank 2. Then, the output variable is obtained as linear combination of the input variables as follows:

$$\mathbf{y} = \sum_{i=1}^{6} a_i \mathbf{x}_i, \quad (13)$$

where $a_i = \{1, 2, 1, 1.5, 0.5\}$, with $i \in [1, 6]$.

2) PLS model building: A first step of the modeling procedure, the model building step, was performed on the training dataset. Then, the testing or model application step was performed on the testing dataset. The training dataset was chosen to define normal operating conditions. One thousand process measurement dataset corresponding to the normal operating conditions were first generated by the above models, to perform a PLS model on $(\mathbf{X}; \mathbf{y})$. These data, are scaled to be zero mean with a unit variance, and then will be used to develop a PLS model.

An important step in PLS model building is the selection of the number of latent variables. In this study, the cross validation technique has been used to determine the number of latent variables for PLS model. To perform cross validation, the training dataset is divided into training and testing subsets. Then, the model is fitted to the training subset, and the prediction errors are calculated for the testing subset. For this example the first 500 faultless observations are considered as the training data and the whole data set as the test data. The optimal number of latent variables has been found to be 2.

Before applying the PLS-based KLD anomaly detection strategy, we need to check whether the residuals of the response variables follow Gaussian distribution to make sure that the data are well represented using a linear PLS model. Checking the normality of the residuals of response variable, F, which were not captured by the PLS model, can be done by visually checking the Henry's line and the histogram of residual vector, which are shown in Figure 2. Histogram and Henry's line, which are depicted in Figure 2, indicate that the normality assumption appears to be a reasonable one.

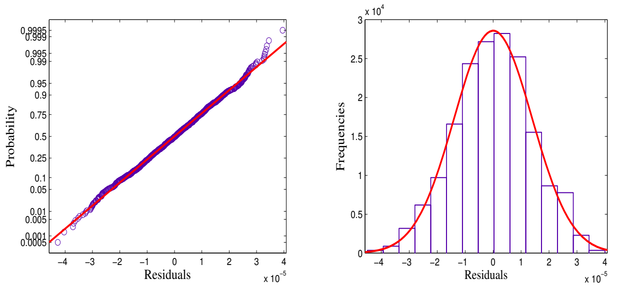

Fig. 2. Gaussian distribution test; Left: Henry's line. Right: Histograms showing the normality of the residuals.

Now, the performance of the PLS-based KLD anomaly detection strategy will be assessed. Before pre-processing the training data, additive anomalies were injected into the raw data. Two different cases of anomalies were simulated: abrupt anomaly and gradual anomaly. In the first case study, it is assumed that the testing data sets contains additive bias faults, (case A). In the second case study is assumed that the testing data set contains drift sensor fault (case B).

Case A: abrupt anomaly

In this case study, we investigated the problem of detecting abrupt anomaly. The testing data used to compare the various fault detection methods, which consist of 500 samples, are generated using the same model described earlier in Equations (12) and (13). To simulate a abrupt fault in the variable $\mathbf{x}_1$, an additive fault having a magnitude 5% of the total variation in $\mathbf{x}_1$ is introduced between samples 200 and 300.

The performances of the $Q^{(x)}$ and T^2 statistics are demonstrated in Figure 3, top and bottom panels, respectively. The dashed horizontal lines represent a 95% confidence limit used to identify the possible faults. These results show that the conventional PLS-based methods ($Q^{(x)}$ and T^2) are completely unable to detect this small simulated fault. This is because these conventional PLS-based fault detection metrics only take into account the information provided by the present data samples in the decision making process, which makes these metrics not very powerful in detecting small changes. The results of PLS-based KLD fault detection algorithm, however, which are illustrated in Figure 4, clearly show the capability of this proposed method in detecting this small fault without false alarms.

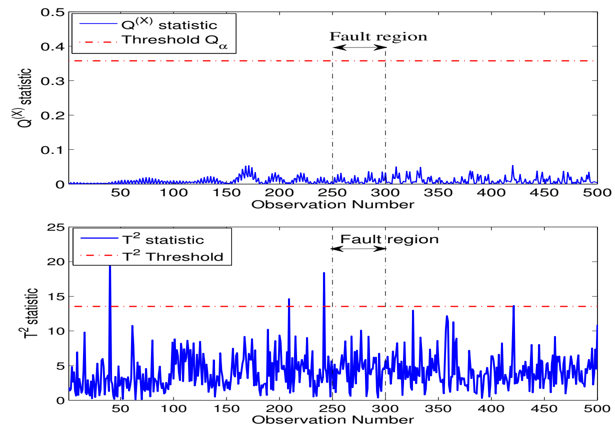

Fig. 3. The time evolution of the $Q^{(X)}$ statistic (top), and T^2 statistic (bottom) in the presence of a bias fault in $\mathbf{x}_1$ (Case A).

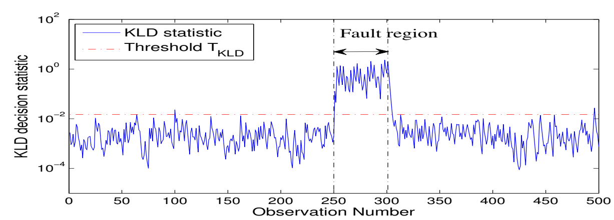

Fig. 4. The time evolution of the KLD decision statistic on a semi-logarithmic scale in the presence of a bias fault in $\mathbf{x}_1$ (Case A).

Case B: gradual anomaly

Gradual or incipient anomalies, for example a slow drift in a sensor, are more subtle and the impact is not so obvious. However, incipient anomalies if left unattended for a long period of time might degrade the required performance of the inspected system and might lead to catastrophic damages. The aim of this case study is to assess the potential of the proposed PLS-based KLD anomaly detection approach in detecting incipient or gradual anomaly. Towards this end, a slow increase in the input variable $\mathbf{x}_1$ with a slope of 0.001 was added to the test data starting at sample number 300 of the simulated testing data. In other words, the input variable $\mathbf{x}_1$ was linearly increased from sample 300 till the testing data by adding 0.001(k-300) to the $\mathbf{x}_1$ value of each sample in this range, where k is the sample number. The $Q^{(x)}$ and T^2 statistics for this case are plotted in Figure 5, top and bottom panels, respectively. The result of the $Q^{(x)}$ and T^2 statistics show that faults remained undetectable by applying the conventional PLS statistic. In other words, The conventional PLS approach fails to detect this small fault. In contrast to the conventional PLS, the results of KLD scheme, which are shown in Figure 6, clearly indicate that the proposed strategy can successfully detect this fault. The KLD statistic was gradually increased as the fault was slowly developed, and began to violate the threshold value when the fault magnitude become sufficiently large enough to be detected by given model.

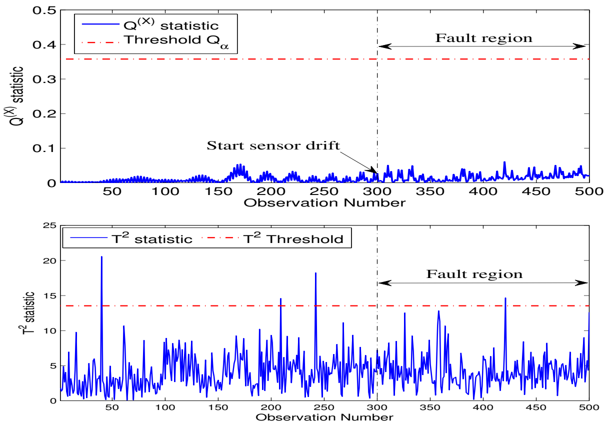

Fig. 5. The evolution of the Q statistic (top), and T^2 statistic (bottom) in the presence of drift fault with slope 0.001 in $\mathbf{x}_1$ (Case B).

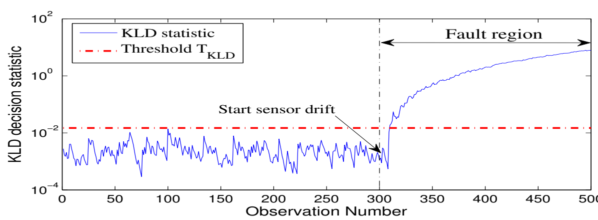

Fig. 6. The evolution of the KLD decision statistic in the presence of drift fault with slope 0.001 in $\mathbf{x}_1$ (Case B).

In summary, the proposed method clearly outperforms the $Q^{(x)}$ and T^2 methods. It successfully detects the small bias faults without false alarms and detects the drift faults correctly and more quickly. These results are encouraging especially when it is of interest to detect faults with small magnitudes.

VI. Conclusion

Early detection of small fault is of crucial interest in order to avoid the economic or even catastrophic consequences that can result from an accumulation of such small anomalies. In this paper, an incipient anomaly detection technique based on KLD has been developed. PLS was used as the modelling framework of the proposed PLS-based KLD anomaly detection methodology. Then, the KLD was used as monitoring index was applied to the output residuals of PLS model to detect anomalies when the data did not fit the reference PLS model. To assess the fault detection abilities of the proposed PLS-based KLD anomaly detection method, two case studies were considered, one involving an abrupt fault, the second one involving a gradual or drifting fault. The simulation results of all the cases clearly show the effectiveness of the proposed algorithm anomaly detection methodology over conventional PLS methods, especially small or moderate anomalies. It has been shown that satisfactory detection results were obtained using the proposed method, especially for detecting small fault.

References

[1] R. Isermann, "Model-based fault-detection and diagnosis : status and applications," *Annual Reviews in Control*, vol. 29, pp. 71–85, 2005.

[2] L. Chaing, E. Russel, and R. Braatz, "Fault detection and diagnosis in industrial systems," *Springer, London*, 2001.

[3] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, and K. Yin, "A review of process fault detection and diagnosis part III: Process history based methods," *Computers and Chemical Engineering*, vol. 27, pp. 327–346, 2003.

[4] D. C. Montgomery, "Introduction to statistical quality control," *John Wiley& Sons, New York*, 2005.

[5] F. Harrou and M. Nounou, "Monitoring linear antenna arrays using an exponentially weighted moving average-based fault detection scheme," *Systems Science & Control Engineering: An Open Access Journal*, vol. 2, no. 1, pp. 433–443, 2014.

[6] F. Harrou, F. Kadri, S. Chaabane, C. Tahon, and Y. Sun, "Improved principal component analysis for anomaly detection: Application to an emergency department," *Computers & Industrial Engineering*, vol. 88, pp. 63–77, 2015.

[7] F. Harrou, M. N. Nounou, H. N. Nounou, and M. Madakyaru, "PLS-based EWMA fault detection strategy for process monitoring," *Journal of Loss Prevention in the Process Industries*, vol. 36, no. 1, pp. 108–119, 2015.

[8] F. Harrou, M. Nounou, H. Nounou, and M. Madakyaru, "Statistical fault detection using PCA-based GLR hypothesis testing," *Journal of Loss Prevention in the Process Industries*, vol. 26, no. 1, pp. 129–139, 2013.

[9] S. Wold, H. Ruhe, H. Wold, and W. D. III, "The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses," *SIAM Journal on Scientific and Statistical Computing*, vol. 5, no. 3, pp. 735–743, 1984.

[10] P. Geladi and B. Kowalski, "Partial least-squares regression: a tutorial," *Analytica chimica acta*, vol. 185, pp. 1–17, 1986.

[11] 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.

[12] G. Li, X. Zeng, J. Yang, and M. Yang, "Partial least squares based dimension reduction with gene selection for tumor classification," in *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, BIBE 2007.* IEEE, 2007, pp. 1439–1444.

[13] S. Kullback and R. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, pp. 79–86, 1951.

[14] T. Kourti and J. MacGregor, "Process analysis, monitoring and diagnosis using multivariate projection methods: A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 28, no. 3, pp. 3–21, 1995.

[15] H. Wold, "Nonlinear estimation by iterative least squares procedures," *Research papers in statistics*, vol. 12, no. 5, pp. 134–139, 1966.

[16] B. Li, J. Morris, and E. Martin, "Model selection for partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 64, no. 1, pp. 79–89, 2002.

[17] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441, 1933.

[18] J. Jackson and G. Mudholkar, "Control procedures for residuals associated with principal component analysis," *Technometrics*, vol. 21, pp. 341–349, 1979.

[19] M. Basseville and I. Nikiforov, *Detection of Abrupt change: Theory and Application.* Prentice Hall, Information and System Sciences Series, 1993.

[20] F. Harrou, L. Fillatre, and I. Nikiforov, "Anomaly detection/detectability for a linear model with a bounded nuisance parameter," *Annual Reviews in Control*, vol. 38, no. 1, pp. 32–44, 2014.

[21] F. Liese and K.-J. Miescke, *Statistical decision theory: estimation, testing, and selection.* Springer Science & Business Media, 2008.