Improved Anomaly Detection Using Multi-scale PLS and Generalized Likelihood Ratio Test

(1)Muddu Madakyaru,
(1)Department of Chemical Engineering
Manipal Institute of Technology
Manipal University, India
muddu.m@manipal.edu

(2)Fouzi Harrou, *Member, IEEE*, (2)Ying Sun
(2)King Abdullah University of Science and Technology (KAUST)
Computer, Electrical and Mathematical Sciences and Engineering Division
Thuwal, 23955-6900, Saudi Arabia
fouzi.harrou@kaust.edu

***Abstract*—Process monitoring has a central role in the process industry to enhance productivity, efficiency, and safety, and to avoid expensive maintenance. In this paper, a statistical approach that exploit the advantages of multiscale PLS models (MSPLS) and those of a generalized likelihood ratio (GLR) test to better detect anomalies is proposed. Specifically, to consider the multivariate and multi-scale nature of process dynamics, a MSPLS algorithm combining PLS and wavelet analysis is used as modeling framework. Then, GLR hypothesis testing is applied using the uncorrelated residuals obtained from MSPLS model to improve the anomaly detection abilities of these latent variable based fault detection methods even further. Applications to a simulated distillation column data are used to evaluate the proposed MSPLS-GLR algorithm.**

I. Introduction

Engineering systems require monitoring approaches to detect abnormalities and sustain its normal operation. In such framework, data-based techniques provide efficient tools for extracting useful feature for design of monitoring schemes based on empirical models derived from the available process data [1]–[4]. Such methods require a minimal a prior knowledge about process physics, but depends on the availability of quality input data. Principal component analysis (PCA) and partial least squares (PLS) are two basic methods of multivariate analysis and reputed as powerful tools for monitoring multivariate processes with highly correlated process data [5], [6]. By extracting useful data from the original dataset using PLS modeling and then using monitoring indices such as T^2 and Q statistics faults in the monitored process can be detected. However, PLS-based monitoring indices such as T^2 and Q statistics often are relatively inefficient at detecting incipient faults [7], [8].

However, the presence of measurement errors (noise) in the data and model uncertainties degrade the quality of fault detection (FD) techniques. In addition, most of industrial process data generally have multiscale properties, signifying that they include features and noise occurring at different contributions over both time and frequency. Nevertheless, majority FD approaches are based on time-domain data such PCA and PLS approaches (operates on a single time scale), and thus they not take into consideration the multiscale characteristics of the data. Multiscale representation of data using wavelets, which is a powerful feature extraction tool, has been demonstrating a good capacity to separate efficiently deterministic and stochastic features [9]. Wavelet based-multiscale representation of data has been used extensively in literature to ameliorate the effectiveness and robustness of fault detection strategies [9], [10]. Regarding multiscale PLS (MSPLS) modeling and monitoring, in [11], [12] the author employed multiscale representation to develop multiscale MSPLS in order to enhance the accuracy of PLS model. In fact, MSPLS model is able to remove the autocorrelations of variables by wavelet analysis and to remove correlations between variables with PLS transformation [11]. The author in [11] demonstrates that multiscale representation of data improved the FD abilities of conventional PLS. Thus, combining the advantages of MSPLS with those of generalized likelihood ratio (GLR) hypothesis testing should provide even further improvements in fault detection. Likewise, GLR hypothesis testing which is very popular in the framework of model-based fault detection, has demonstrated good detection capacity for a fixed false alarm probability [7], [13], [14].

The rest of this paper is organized as follows. Section II gives a brief overview of the PLS model. In Section III, the multiscale PLS approach is briefly reviewed, and Section IV introduces the GLR hypothesis testing and its use in anomaly detection. Next, the concept of marrying MSPLS modeling and GLR test is presented in Section V. Section VI applies the proposed MSPLS-GLR procedure to a simulated distillation column process. Finally, Section VII concludes this paper.

II. PLS modeling

Consider a pair of datasets $\mathbf{X} \in \mathbb{R}^{N\times M}$ and $\mathbf{Y} \in \mathbb{R}^{N\times 1}$, where $\mathbf{X}$, $\mathbf{Y}$ are the input and output variables, respectively. After the data standardization by first subtracting the sample mean of the training data and then dividing by the sample standard deviation of the training data, PLS projects $\mathbf{X}$ and $\mathbf{Y}$ on to a lower dimension subspace defined by number of latent variable $[z_1, z_2, \dots, z_l]$ as follows:

$$\begin{cases} \mathbf{X} = \mathbf{Z}\mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} = \mathbf{Z}\mathbf{Q}^T + \mathbf{F} \end{cases} \quad (1)$$

Where $\mathbf{Z} \in \mathbb{R}^{N \times l}$ (l is the number of latent variable) is the score matrix represents the projection of the variables on the subspace, $\mathbf{P} \in \mathbb{R}^{M \times l}$ represents the loading matrix for $\mathbf{X}$ and $\mathbf{Q} \in \mathbb{R}^{1 \times l}$ represents the loading matrix for $\mathbf{Y}$. $\mathbf{E}$ and $\mathbf{F}$ represents the model residue of input and output respectively. PLS calculates the input loading vectors, $\mathbf{P}_i$, so that the covariance between the estimated latent variable $\hat{\mathbf{Z}}_i$ and model output, $\mathbf{Y}$, i.e., [15]:

$$\hat{\mathbf{P}}_i = arg \max_{\mathbf{P}_i} \quad cov(\mathbf{Z}_i, \mathbf{Y}) \tag{2}$$

$$\text{such that} \quad \mathbf{P}_i^T \ \mathbf{P}_i = 1; \quad \mathbf{Z}_i = \mathbf{X}\mathbf{P}_i$$

where, $i = 1, \ldots, l, \ l \leq m$. Various algorithms have been proposed to compute PLS-based latent variables [15], [16], the most used is the Non-linear iterative partial least squares (NIPALS) algorithm .

III. Multiscale PLS modeling

In this section, the multiscale will be merged with PLS model to improve the prediction quality of the PLS model and latter utilized for fault detection methods. We introduce multiscale representation of data and describe their advantages when applied to fault detection techniques.

A. Wavelet-based multiscale representation

Using a discrete wavelet transform, an original signal space, x, can be decomposed into two sub-spaces: an approximation subspace, A, and detailed subspaces, D. The scale function $\phi_{j,k}(t) = \sqrt{2^{-j}}\phi(2^{-j}t - k), k \in \mathbb{Z})$ and wavelet functions $\psi_{j,k}(t) = \sqrt{2^{-j}}\psi(2^{-j}t - k), j = 1, \ldots, J, k \in \mathbb{Z}$, where the coarsest scale J normally termed the decomposition level, span the approximation and detailed subspaces, respectively. Any signal can be represented by a summation of all scaled and detailed signals as follows [17]:

$$x(t) = \overbrace{\sum_{k=1}^{n2^{-J}} a_{Jk}\phi_{Jk}(t)}^{A_J(t)} + \sum_{j=1}^{J} \overbrace{\sum_{k=1}^{n2^{-j}} d_{jk}\psi_{jk}(t)}^{D_j(t)} . \tag{3}$$

where j, k, J and n represent the dilation parameter, translation parameter, number of scales, and number observations in the original signal, respectively [18]–[20]. d_{jk} and a_{Jk} represent the scaling and the wavelet coefficients, respectively, and $A_J(t)$ and $D_j(t), (j = 1, 2, \ldots, J)$ are the approximated signal the detail signals, respectively.

In other words, the detailed signal $D_j(t)$ at scale j can be obtained by passing the original and scaled signals through a high-pass filter (g), and the scaled signals are generated by passing the original and scaled signals through a low-pass filter (h) [21]. A signal can be described at multiple resolutions by decomposing it on a family of wavelets and scaling functions. For example, consider the series time measurements of the feature indicator shown in Figure 1. The signals in Figures 1(b, d and f) are at increasingly coarser scales compared to the original signal in Figure 1(a).

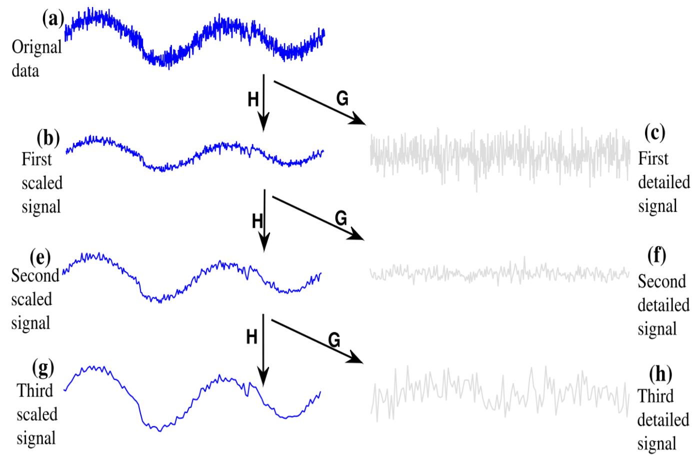

Fig. 1. Illustration of data representation at multiple scales of a heavy-sine signal.

Multiscale representation is an effective method for dealing with autocorrelated or non-Gaussian data [10]. Some of the advantages of multiscale filtering in PLS model are pointed out next [10]: (1) The ability to naturally denoise signals. (2) Since in most cases the resulting wavelet coefficients are orthogonal the coefficients become decorrelated even if the input set of data happens to be autocorrelated. (3) Wavelet coefficients follow a normal distribution regardless of the distribution of the input set of data. (4) Wavelet coefficients are stationary and can therefore be used for both stationary and non-stationary input signals. These advantages will be exploited to improve the quality of PLS models via the development of an algorithm that merge multiscale de-noising and PLS models. Before discussing the multiscale PLS modeling, a brief description of multiscale de-noising is presented.

B. Multiscale data filtering algorithm

A wavelet-based de-nosing algorithm comprises the following three main steps [22]:

(1) Decompose the original signal at multiple scales via wavelet transform to obtain wavelet coefficient series in different level;
(2) Select thresholds for each level and remove the wavelet coefficients that are below a threshold value;
(3) Inverse wavelet transform based on the detail coefficients to obtain a de-noised signal.

Several threshold selection criterion have been proposed including Fixed Threshold [23], Rigorous Sure threshold, Heursure threshold and Min-max threshold. The simplest wavelet threshold method, which is proposed in [23], uses the same threshold to deal with coefficients in the expansion.

C. Multiscale PLS (MSPLS) modeling algorithm

The motive behind the integrated multiscale PLS modeling is to amalgamate the benefit of multiscale denoising and PLS modeling to improve model quality and thus improving the fault detection methods. Let the input data matrix $\mathbf{X}$ and the output data matrix $\mathbf{y}$, and the denoised data via multiscale filtering at a scale (j) be $\mathbf{X}_j$ and $\mathbf{y}_j$, then the PLS model, which is computed using these denoised data, can be expressed as,

$$\mathbf{y}_j = \mathbf{T}_j \mathbf{B}_j \mathbf{Q}_j^T - \mathbf{F}_j, \tag{4}$$

where, $\mathbf{X}_j \in \mathbb{R}^{n \times m}$ is the filtered input data matrix at scale (j), $\mathbf{y}_j \in \mathbb{R}^{n \times 1}$ is the filtered output vector at scale (j), $\mathbf{F} \in \mathbb{R}^{m \times p}$ is the residuals of output matrices at scale (j).

However, filtering the input and output data a priori without taking the relationship between these two data sets into account may result in the removal of features that are important to the model. Thus, multiscale filtering needs to be integrated with PLS model for proper noise-removal. One way to accomplish this integration between multiscale filtering and PLS modeling is using the following MSPLS modeling algorithm [24]:

TABLE I
MSPLS MODELING FRAMEWORK.

(1) Split the data into two sets: training and testing.
(2) Pre-processing of the input\output data is required to ensure that all variable data is set to zero mean and unit variance
(3) Denoise the training data at different scales (decomposition depths) via the denoising algorithm presented in Section III-B.
(4) Construct a PLS model based on the denoised data at each scale. The number of LVs is determined using cross-validation
(5) Use the estimated model from each scale to predict the output for the testing data, and compute the cross validation mean square error.
(6) Select the PLS with the least cross validation mean square error as the MSPLS model.

After a MSPLS model is builded, various methods for anomaly detection can be applied. Two monitoring statistics, the T^2 and Q statistics, are usually utilized for FD purpose [25]. First, Hoteling T^2 statistics indicates the variation within the process model in LVs space. The other is the Q statistic, also known as the squared prediction error (SPE), which monitors how well the data conforms to the model. Although the two methods have their advantages and disadvantages, both tend to fail to detect small anomalies [26]. Motivated by the powerful of GLR test for detecting additive shifts in the process mean [13], we propose an innovative MSPLS-based GLR faut detection method for multivariate processes. In the next section, we briefly describe the GLR test.

IV. GLR TEST-BASED FAULT DETECTION

Assume that we have a measured vector $Y = [y_1, y_2, \dots, y_n] \in \mathbb{R}^n$ distributed according to one of the two following Normal distributions, $\mathcal{N}(0, \sigma^2 I_n)$ or $\mathcal{N}(\theta \neq 0, \sigma^2 I_n)$, where θ is the mean vector (which is the value of the anomaly) and $\sigma^2 > 0$ is the variance, which is supposed to be known. The GLR test decides between the null hypothesis $\mathcal{H}_0 = \{Y \sim \mathcal{N}(0, \sigma^2 I_n)\}$ and the alternative hypothesis $\mathcal{H}_1 = \{Y \sim \mathcal{N}(\theta, \sigma^2 I_n)\}$ by comparing between the generalized likelihood ratio, $\mathcal{L}(Y)$, and a given value of the threshold, $h(\alpha)$. The likelihood ratio test statistic, $\mathcal{L}(Y)$, is given as

$$\mathcal{L}(Y) = 2 \log \frac{\underset{\theta \in \mathbb{R}^n}{sup} f_\theta(Y)}{f_{\theta=0}(Y)} \frac{1}{\sigma^2} \left\{ \|Y\|_2^2 \right\} \gtrless^{\mathcal{H}_0}_{\mathcal{H}_1} \quad h(\alpha). \tag{5}$$

Typically, the threshold $h(\alpha)$ is chosen to achieve a desired probability of a false alarm, predefined a priori [27].

$$\mathbb{P}_0 \left(\mathcal{L}(Y) \geq h(\alpha) \right) = \int_h^\infty f_0(y) dy = 1 - F_{\chi_1^2}(h) = \alpha. \tag{6}$$

Notice that $Y_t \sim \mathcal{N}(\theta, \sigma^2)\}$ under $\mathcal{H}_0$ and consequently $\mathcal{L}$ has a central χ^2 distribution with one degree of freedom. Moreover, $\mathcal{H}_0$ can be rejected at the significance level α if the observed value of $\mathcal{L}(Y)$ is larger than $(1-\alpha)$-th quantile of the χ_1^2 distribution. The power function is given by:

$$\beta_{\delta^*}(c^2) \quad = \quad \mathbb{P}_\theta(\mathcal{L}(Y) \geq h(\alpha)) = 1 - F_{1,\lambda(\theta)}(h),$$

where, $F_{1,\lambda}(Y)$ is the non-central $\chi^2(1, \lambda)$ distribution with one degrees of freedom and non-centrality parameter $\lambda(\theta) = \frac{1}{\sigma^2} \parallel P_H^\perp \theta \parallel_2^2$. The GLR test will be merged with MSPLS to enhance its fault detection ability.

V. MSPLS-BASED GLR FAULT DETECTION SCHEME

In the proposed MSPLS-GLR monitoring approach, the GLR test is applied to the residuals of the responses variables obtained from the MSPLS model. As given in Equation (4), the output vector $\mathbf{y}$ can be written as the sum of a predicted vector $\widehat{\mathbf{y}}$ and a residual vector $\mathbf{F}$, i.e.,

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{F}. \tag{7}$$

The residual of the output variable, $\mathbf{F} = [f_1, \dots, f_t, \dots, f_n]$, which is the diference between the observed value of the output variable, y, and the predicted value, $\widehat{y}$, obtained from MSPLS model, is potential indicator for fault detection. In nominal conditions, no abnormalities happen in the monitored process; thus, the value of residuals fluctuates around zero due to measurement noise. A significant departure from zero of residuals reveals important deviations from normal behavior, indicating that the inspected process is running under abnormal conditions. Thus, the fault-detection problem can be addressed as as a binary hypothesis testing problem, considering two hypotheses: the null hypothesis $\mathcal{H}_0$, where $\mathbf{F}$ is fault-free and the alternative hypothesis $\mathcal{H}_1$, where $\mathbf{F}$ contains a fault. The GLR-based test is used to make decisions between the null hypothesis $\mathcal{H}_0$, (absence of anomalies) and the alternative hypothesis $\mathcal{H}_1$, (presence of anomalies). In such cases, to know whether the process is under control, it is natural to consider testing the following hypotheses:

$$\begin{cases} \mathcal{H}_0 = \{\mathbf{F} \sim \mathcal{N}(0, \sigma^2 I_n)\}, & \text{(null hypothesis);} \\ \mathcal{H}_1 = \{\mathbf{F} \sim \mathcal{N}(\theta, \sigma^2 I_n), & \text{(alternative one).} \end{cases} \tag{8}$$

To test whether $\mathcal{H}_0$ should be rejected in favor of $\mathcal{H}_1$, we use the GLR test presented in Section IV. An algorithm that outlines the proposed MSPLS-based GLR fault-detection method is summarized next.

TABLE II
MSPLS-BASED GLR FAULT DETECTION ALGORITHM.

1) **Given:**
   - Collect the training data set ($\mathbf{X}$ and $\mathbf{y}$) representative of a normal condition. This is necessary to build the reference MSPLS model and to set the control limits.
   - A predefined false alarm probability, α_0,
2) **Data pre-processing:**
   - Auto-scale the data by removing the data mean and scaling the variance to unity.
3) **MSPLS training phase:**
   (1) Construct a MSPLS model using the training data.
   (2) Determine the number of LVs using cross-validation technique or any other model selection method,
   (3) Express the output matrix as a sum of predicted and residual matrices as given in equation (4),
4) **MSPLS-based monitoring phase:**
   (1) For a new sample data, apply the same scaling used in training phase.
   (2) Compute the model output residuals, $\mathbf{F}$,
   (3) Compute the GLR threshold, $h(\alpha)$
   (4) Compute the GLR decision function, $\mathcal{L}(\mathbf{F})$, and check whether there is any violation of its threshold, $h(\alpha)$.

VI. MONITORING A SIMULATED DISTILLATION COLUMN

A. Process description

The method is tested using a distillation column process simulated by Aspen (see [28] for details) with added zero-mean Gaussian noise, where the predictor variables consist of ten temperatures (Tc) in different stages of the monitored column, feed flow rates and reflux stream, and the composition of the light component in the distillate stream represents the response variable. The Aspen simulator is used to generate 1024 data samples to be used in constructing the reference MSPLS model. Figure 2 shows the dynamic input-output data of the distillation column around the nominal opearting condition which then added noise of Signal-to-Noise Ration of 10. These data are used to develop model development and testing purpose. The first 512 data set are used for training the PLS model and Latter 512 data set used for testing purpose. Using the cross-validation technique, three LVs were needed for the PLS model.

To evaluate the performance of the inferential model, four numerical criteria were used: R^2 and the root mean square error (RMSE). These were calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}, \quad (9)$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - mean(Y))^2}, \quad (10)$$

where y_t are the measured values, $\hat{y}_t$ are the corresponding predicted values by the PLS model and n is the number of samples. The constructed MSPLS model provides a good predictive quality, with the $R^2 = 0.96$ and the low RMSE of 0.003.

A PLS model is fitted to the scaled training dataset, and the goodness of fit is shown in Figure 3. Figure 3 show the

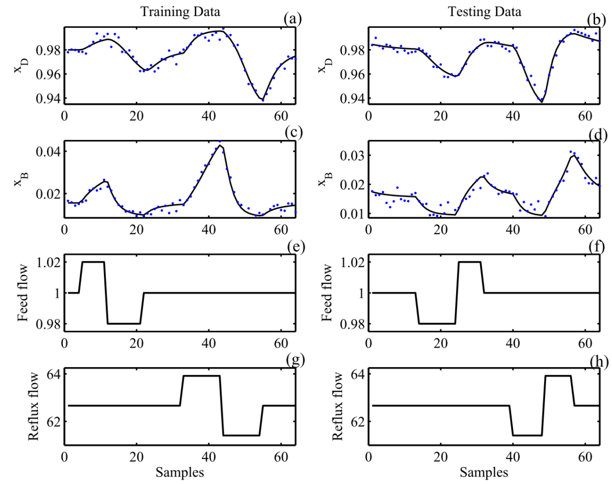

Fig. 2. Simulation of Distillation Column: Variation of input-output data with noise SNR=10 (solid red line:noise free data; blue dots- Noisy data)

scatter plot of observed versus predicted values of the testing data set obtained from the selected MSPLS model, and the regression line. It shows that the points are distributed along the regression line the slope of the regression line between observed and predicted values is not significantly different from 1 and the y-intercept is not significantly different from 0. Therefore, the models were successful in accounting for most of the significant autocorrelations present in the data, and there is no indication of a curvature or other anomalies. According to Figure 3, it can be seen that the scatter plots of observed and predicted data indicate a reasonable performance of the constructed MSPLS model.

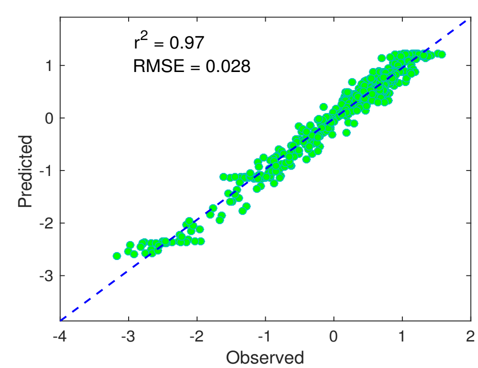

Fig. 3. Scatter plots of predicted and observed training data.

After the model is identified, it is used to monitor the abnormal events (faults) in the distillation column process that may lead the process to depart from its normal state.

B. Detection results

After a process model has been successfully identified, we can proceed with fault detection. We then introduce two types of anomalies into the testing data sets and compare the performance of different anomaly detection techniques (i.e., MSPLS-GLR and PLS-GLR charts). In the first case study, it is assumed that the testing data sets contain additive bias anomlies, case (i). In the second case it is assumed that the testing data set contains drift sensor anomalies, case (ii).

Case (i): Abrupt anomaly - Bias sensor anomaly: The testing data with SNR=10 are generated for the purpose of evaluation of MSPLS and PLS monitoring performances. In this case study, a bias anomaly, which is 2% of the total variation in temperature Tc_3, was incorporated into the temperature sensor measurements Tc_3 between samples 100 and 150. The performances of the PLS-GLR test and MSPLS-GLR test are demonstrated in Figure 4(a)-(b), respectively. From figure 4(a) it can be seen that the PLS-GLRT chart is capable to detect this abrupt anomaly but with several false alarms. Figure 4(b), clearly show the capability of this proposed MSPLS-GLR monitoring method, in detecting this small anomaly without false alarms.

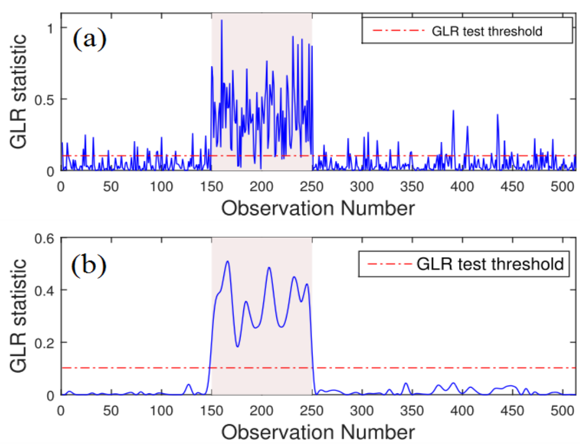

Fig. 4. Monitoring results of PLS-GLR chart (a), and MSPLS-GLR chart (b) in the presence of a bias anomaly in the temperature sensor measurements 'Tc_3' (Case (i)).

Case (ii): Gradual anomaly - Slow drift sensor anomaly: This case is aimed to assess the potential of the proposed MSPLS-based GLR anomaly detection scheme to detect a slow drift anomaly. A slow drifting sensor anomaly with a slope of 0.01 was added to the temperature sensor Tc_3 starting at sample 250 lasting until the end of the testing data. Monitoring results of PLS-GLRT and MSPLS-GLRT statistics are shown in Figure 5(a)-(b). Figure 5(a) illustrates that PLS-GLRT chart detected anomaly at sample 300, but with several region of false alarms. The MSPLS-GLRT chart increased linearly from sample 250, exceeding the control limits at signal 295 (see Figure 5(b)). The superiority of the new MSPLS-GLRT chart over the conventional PLS-GLRT chart is verified again.

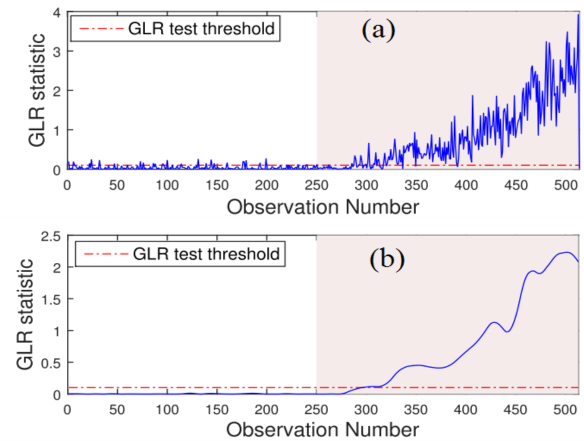

Fig. 5. Monitoring results of PLS-GLR chart(a), and MSPLS-GLR chart (b) in the presence drift sensor anomaly in 'Tc_3' (Case (ii)).

VII. Conclusion

Data observed from chemical processes are usually noisy and correlated in time, which makes the anomaly detection more difficult as the presence of noise degrades anomaly detection quality and most methods are developed for independent observations. This paper proposes a statistical method to monitor multivariate input output systems, which is based on MSPLS algorithm and GLR test. MSPLS has been used in this work as a modeling framework for fault detection using GLR hypothesis testing. The GLR test is applied on the uncorrelated residuals obtained from the MSPLS model. Data of the simulated distillation column is used to validate the advantages of the MSPLS-based GLR fault detection method. Results show that the combined use of MSPLS models and GLR hypothesis testing can achieve better fault detection efficiency than the conventional PLS based GLRT.

Acknowledgement

This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No: OSR-2015-CRG4-2582.

References

[1] S. Yin, S. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014.

[2] F. Harrou, Y. Sun, and S. Khadraoui, "Amalgamation of anomaly-detection indices for enhanced process monitoring," *Journal of Loss Prevention in the Process Industries*, vol. 40, pp. 365–377, 2016.

[3] F. Harrou, M. Madakyaru, Y. Sun, and S. Khadraoui, "Improved detection of incipient anomalies via multivariate memory monitoring charts: Application to an air flow heating system," *Applied Thermal Engineering*, vol. 109, pp. 65–74, 2016.

[4] F. Harrou, Y. Sun, and M. Madakyaru, "Kullback-leibler distance-based enhanced detection of incipient anomalies," *Journal of Loss Prevention in the Process Industries*, vol. 44, pp. 73–87, 2016.

[5] W. Liang and L. Zhang, "A wave change analysis (WCA) method for pipeline leak detection using gaussian mixture model," *Journal of Loss Prevention in the Process Industries*, vol. 25, no. 1, pp. 60–69, 2012.

[6] L. Chiang, R. Braatz, and E. Russell, *Fault detection and diagnosis in industrial systems*. Springer Science & Business Media, 2001.

[7] 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.

[8] ——, "PLS-based EWMA fault detection strategy for process monitoring," *Journal of Loss Prevention in the Process Industries*, vol. 36, pp. 108–119, 2015.

[9] B. Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," *AIChE journal*, vol. 44, no. 7, pp. 1596–1610, 1998.
[10] R. Ganesan, T. Das, and V. Venkataraman, "Wavelet-based multiscale statistical process monitoring: A literature review," *IIE transactions*, vol. 36, no. 9, pp. 787–806, 2004.
[11] H. Lee, M. Lee, and J. Park, "Multi-scale extension of PLS algorithm for advanced on-line process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 98, no. 2, pp. 201–212, 2009.
[12] M. S. Roodbali and M. Shahbazian, "Multiscale PLS modeling for industrial process monitoring," *International Journal of Computer Applications*, vol. 26, no. 6, pp. 26–33, 2011.
[13] 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.
[14] M. Basseville and I. Nikiforov, *Detection of abrupt changes: theory and application*. Prentice Hall Englewood Cliffs, 1993, vol. 104.
[15] Y. Hiroyuki, Y. B. Hideki, F. C. E. O. Hiromu, and F. Hideki, "Canonical correlation analysis for multivariate regression and its application to metabolic fingerprinting," *Biochemical Engineering Journal*, vol. 40, pp. 199–204, 2008.
[16] J. Shao, "Linear model selection by cross-validation," *Journal of the American statistical Association*, vol. 88, no. 422, pp. 486–494, 1993.
[17] R. Gao and R. Yan, *Wavelets: Theory and applications for manufacturing*. Springer Science & Business Media, 2010.
[18] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM review*, vol. 31, no. 4, pp. 614–627, 1989.
[19] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on pure and applied mathematics*, vol. 41, no. 7, pp. 909–996, 1988.
[20] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
[21] M. Sheriff, F. Harrou, and M. Nounou, "Univariate process monitoring using multiscale Shewhart charts," in *International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, 2014, pp. 435–440.
[22] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptotia?" *J. Roy. Stat. Soc. B*, vol. 57, no. 301, 1995.
[23] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
[24] M. Madakyaru, M. N. Nounou, and H. N.Nounou, "Integrated multiscale latent variable regression and application to distillation coloumn," *Modelling and Simulation in Engineering*, vol. 2013, pp. 1–7, 2013.
[25] S. Qin, "Statistical process monitoring: Basics and beyond," *Journal of Chemometrics*, vol. 17, no. 8/9, pp. 480–502, 2003.
[26] D. C. Montgomery, "Introduction to statistical quality control," *John Wiley& Sons, New York*, 2005.
[27] E. Lehmann, *Testing Statistical Hypotheses*. New York: Chapman and Hall, 1996.
[28] M. Madakyaru, M. Nounou, and H. Nounou, "Enhanced modeling of distillation columns using integrated multiscale latent variable regression," in *IEEE Symposium on Computational Intelligence in Control and Automation (CICA)*. IEEE, 2013, pp. 73–80.