Accepted Manuscript

Title: Improved Data-based Fault Detection Strategy and Application to Distillation Columns

Author: Muddu Madakyaru Fouzi Harrou Ying Sun

PII: S0957-5820(17)30022-8
DOI: <http://dx.doi.org/doi:10.1016/j.psep.2017.01.017>
Reference: PSEP 957

To appear in: *Process Safety and Environment Protection*

Received date: 5-10-2016
Revised date: 29-12-2016
Accepted date: 19-1-2017

Please cite this article as: Muddu Madakyaru, Fouzi Harrou, Ying Sun, Improved Data-based Fault Detection Strategy and Application to Distillation Columns, *Process Safety and Environment Protection* (2017), <http://dx.doi.org/10.1016/j.psep.2017.01.017>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Improved Data-based Fault Detection Strategy and Application to Distillation Columns

Highlights:

1. Developed a novel data-based fault detection (FD) method.
2. Combined the wavelet-based multiscale PLS modeling and GLR test to the FD.
3. A distillation column is used to evaluate the performances the developed method.
4. The detection results show the superior performance of the new combined strategy.

Improved Data-based Fault Detection Strategy and Application to Distillation Columns

Muddu Madakyaru^a, Fouzi Harrou^b, Ying Sun^b

^aDepartment of Chemical Engineering, Manipal Institute of Technology, Manipal University, Manipal, India

^bKing Abdullah University of Science and Technology (KAUST)

Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, Thuwal 23955-6900, Saudi Arabia
Tel: +966.12808.0602, Fax: +966.12802.1386, E-mail: fouzi.harrou@kaust.edu.sa

Abstract

Chemical and petrochemical processes require continuous monitoring to detect abnormal events and to sustain normal operations. Furthermore, process monitoring enhances productivity, efficiency, and safety in process industries. Here, we propose an innovative statistical approach that exploits the advantages of multiscale partial least squares (MSPLS) models and generalized likelihood ratio (GLR) tests for fault detection in processes. Specifically, we combine an MSPLS algorithm with wavelet analysis to create our modeling framework. Then, we use GLR hypothesis testing based on the uncorrelated residuals obtained from the MSPLS model to improve fault detection. We use simulated distillation column data to evaluate the MSPLS-based GLR chart. Results show that our MSPLS-based GLR method is more powerful than the PLS-based Q and GLR method and MSPLS-based Q method, especially in early detection of small faults with abrupt or incipient behavior.

Keywords: Multi-scale PLS models, GLR hypothesis testing, Data uncertainty, Process monitoring, Distillation Columns.

1. Introduction

Modern automated industrial processes rely on the precise control of process conditions. Detection of anomalies or deviations in such processes is essential. Furthermore, to improve the reliability, safety and efficiency of advanced process control methods, fault detection and fault diagnosis have become important in numerous technical processes. For example, chemical processes require monitoring approaches that can detect abnormalities while sustaining normal operations. Increasing attention to fault detection and safety has led to the development of several fault detection techniques that can be grouped into two main families: model-based approaches and data-based approaches (Venkatasubramanian et al., 2003; Yin and Zhu, 2015; Harrou et al., 2014; Yin et al., 2012, 2014; Gao et al., 2015). The merits of both model-based and data-based process-monitoring techniques have been demonstrated in practice over the past four decades. In model-based approaches, a residual signal is generated from a mathematical model of a system and then

used as an indicator of a fault (Venkatasubramanian et al., 2003; Isermann, 2006). The most commonly used analytical model-based approaches for residual signal generation include the observer-based approach, the parity space approach and the parameter estimation-based approach (Venkatasubramanian et al., 2003; Isermann, 2006). Unfortunately, deriving accurate models of monitored systems, especially complex industrial processes including chemical and environmental processes can be difficult. Also, modeling of complex systems can be very time consuming. In the absence of an explicit model and if measurement signals are the only available resource for process monitoring, data-driven implicit models are a suitable alternative. Unlike the model-based approaches, data-based techniques provide efficient tools for extracting useful features for the design of monitoring schemes based on empirical models derived from the available process data (Yin et al., 2012, 2014; Venkatasubramanian et al., 2003; Harrou et al., 2016b,a; Zhao et al., 2013). Such methods require minimal a priori knowledge of process physics, but they depend on the availability of quality input data (Qin, 2012). Multivariate Statistical Process Control (MSPC) is one such data-based technique. MSPC and its associated statistical techniques are increasingly used in the control of continuous and batch processes in process industries.

Multivariate statistical process monitoring can provide early warnings of abnormal changes in process operations. Principal component analysis (PCA) and partial least squares (PLS) are two basic methods of multivariate analysis and are reputed to be powerful tools for monitoring multivariate processes with highly correlated process data. They have been extensively applied in the field of chemometrics (Liang and Zhang, 2012; Abdi and Williams, 2010; Chiang et al., 2001). Some chemical processes, such as distillation, are usually modeled by two sets of variables, inputs and output). PLS regression is widely used to model multivariate input-output process data (Wold et al., 1984; Yin et al., 2015; Madakyaru et al., 2012). Unlike PCA, PLS finds an optimum pair of latent variables both from the predictor (input) and predicted (output) variables that have the largest covariance (Geladi and Kowalski, 1986; Harrou et al., 2015). Extracting useful data with PLS modeling and then using monitoring indices lead to detection of faults in the monitored process (Harrou et al., 2013c). Several PLS variants have been proposed to overcome the shortcomings of the classical PLS, such as multiway PLS (Nomikos and MacGregor, 1995), multi-block PLS (MacGregor et al., 1994), dynamic PLS (Lee et al., 2004) as well as kernel PLS (Jia and Zhang, 2016). Very recently, an improvement in the PLS method that reduces of reducing the number of required latent variables to achieve a reduction of the computational load compared with the conventional PLS method was reported (Yin et al., 2016a).

However, the presence of measurement errors (noise) in the data and model uncertainties degrade the quality of fault detection techniques. In addition, most chemical process data generally include features and noise occurring over both time and frequency. Nevertheless, the majority of fault detection approaches, including PCA and PLS, are based on time-domain data (operating on a single time scale), and thus

they do not take into consideration the multiscale characteristics of the data. As a consequence model-based and model-free data denoising methods are used for data filtering. For example, extended Kalman filtering and particle filtering are utilized to denoise collected data for fault diagnosis (Yin et al., 2016b; Yin and Zhu, 2015). When a filter is not available, multiscale representation of data using wavelets, which is a powerful feature-extraction tool, has been found to separate efficiently deterministic and stochastic features (Bakshi, 1998a). Wavelet-based multiscale representation of data has been used extensively in the literature to ameliorate the effectiveness and robustness of fault detection strategies (Bakshi, 1998a; Yoon and MacGregor, 2004; Ganesan et al., 2004; Li and Yao, 2005).

The detection of incipient anomalies is crucial to maintaining the normal operations of a system by providing early fault warnings. The problem is that incipient anomalies are often too weak to be detected by conventional monitoring methods. However, conventional MSPLS-based monitoring indices such as T^2 and Q charts cannot detect small changes in process data (Harrou et al., 2016a). Combining the advantages of MSPLS with those of generalized likelihood ratio (GLR) hypothesis testing should improve fault detection. GLR hypothesis testing, which is very popular in the framework of model-based fault detection, has demonstrated good fault detection capacity (Harrou et al., 2014, 2013c; Basseville and Nikiforov, 1993). Here, we draw on wavelet-based multiscale representation of data to improve a PLS-based hypothesis testing fault detection method. Specifically, to consider the multivariate and multi-scale nature of process dynamics, we use a MSPLS algorithm combining PLS and wavelet analysis as the modeling framework. Then, we apply GLR hypothesis testing using residuals obtained from the MSPLS model to improve the fault detection abilities of the latent variable-based fault detection method. Results from simulated distillation column data show that the MSPLS-GLR approach can achieve better fault detection efficiency than a PLS-based GLR approach.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of the PLS model. In Section 3, the multiscale PLS approach is briefly reviewed, and Section 4 introduces GLR hypothesis testing and its use in anomaly detection. Next, the concept of combining MSPLS modeling with the GLR test is presented in Section 5. Section 6 applies the proposed MSPLS-GLR procedure to a simulated distillation column process. Finally, Section 7 concludes this paper.

2. PLS modeling

PLS is a basic multivariate projection method used in multivariate statistic process monitoring (Höskuldsson, 1988). The purpose of PLS is to analyze relationships between input data, $\mathbf{X}$, and output data, $\mathbf{Y}$. Specifically, PLS finds an optimum pair of latent variables in both $\mathbf{X}$ and $\mathbf{Y}$ such that these transformed variables have the largest covariance (Yin et al., 2012; Harrou et al., 2015). PLS has been widely used in economics, sociology and chemometrics.

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 data standardization by first subtracting the sample mean of the training data and then dividing by the standard deviation of the training data, PLS projects $\mathbf{X}$ and $\mathbf{Y}$ on to a lower dimension subspace defined by the number of the 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 the latent variable) is the score matrix representing the projection of the variables on the subspace, $\mathbf{P} \in \mathbb{R}^{M \times l}$ is the loading matrix for $\mathbf{X}$ and $\mathbf{Q} \in \mathbb{R}^{1 \times l}$ is the loading matrix for $\mathbf{Y}$. $\mathbf{E}$ and $\mathbf{F}$ are the residue of the 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., (Hiroyuki et al., 2008),

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

can be maximized with constraint

$$\mathbf{P}_i^T \mathbf{P}_i = 1; \quad \mathbf{Z}_i = \mathbf{X}\mathbf{P}_i,$$

where $i = 1, \dots, l$, $l \leq m$. Various algorithms have been proposed to compute PLS-based latent variables (Hiroyuki et al., 2008; Shao, 1993). The Non-linear iterative partial least squares (NIPALS) algorithm is the most popular. We refer the reader to the chemometrics literature (Godoy et al., 2013; Höskuldsson, 1988) for details on PLS algorithms.

3. Multiscale PLS modeling

As noted earlier, data collected from engineering processes are usually noisy and correlated in time, which makes fault detection more difficult because the presence of noise degrades detection quality and most methods are developed for independent observations. Wavelet-based multiscale modeling of data is an efficient tool for extracting features and is well suited to denoising and decorrelating time series data (Ganesan et al., 2004). Here, we merge multiscale modeling with PLS to improve the prediction quality of the PLS model. We introduce a multiscale representation of the data and describe its advantage when applied to fault detection techniques.

3.1. Wavelet-based multiscale representation

Multiresolution time-series decomposition was initially applied by Mallat, who used orthogonal wavelets during data compression for image decoding (Mallat, 1989). Wavelets are a family of basis functions that can be expressed as the following localized in both time and frequency (Bakshi, 1998b):

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad (3)$$

4

where a is the dilation parameter, b is the translation parameter (Gao and Yan, 2010) and $\psi(t)$ is the mother wavelet. Figure 1 illustrates the translation and dilation mechanism of a mother wavelet, which shows the feature in time and frequency. Both these parameters are commonly discretized dyadically as $a = 2^m$, $b = 2^m k$, $(m, k) \in \mathbb{Z}^2$. The family of wavelets can be represented as $\psi_{mn}(t) = 2^{-\frac{m}{2}} \psi(2^{-m}t - m)$. Here, $\psi(t)$ is the mother wavelet and m and k are the respective dilation and translation parameters. Different families of basis functions are created based on their convolution with different filters, such as the Haar scaling function and the Daubechies filters (Gao and Yan, 2010; Zhou et al., 2006; Daubechies, 1988). Parameters that are discretized dyadically force downsampling, reducing the number of parameters dyadically with every decomposition; However, dyadically discretized wavelets force samples at non-dyadic locations to become decomposed only after a certain time delay.

Figure 1: Representation of translation via the time constant, t , and dilation via the scaling factor, a .

Based on a discrete wavelet transform, an original signal space, S , can be decomposed into two sub-spaces: an approximation subspace, S_a , and a detailed subspace, S_d . The scale function, $\phi_{j,k}(t) = \sqrt{2^{-j}}\phi(2^{-j}t - k)$, $k \in \mathbb{Z}$, and the wavelet functions, $\psi_{j,k}(t) = \sqrt{2^{-j}}\psi(2^{-j}t - k)$, $j = 1, \dots, J$, $k \in \mathbb{Z}$, where the coarsest scale, J , is 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 (Gao and Yan, 2010):

$$x(t) = \underbrace{\sum_{k=1}^{n2^{-J}} a_{Jk} \phi_{Jk}(t)}_{A_J(t)} + \sum_{j=1}^J \underbrace{\sum_{k=1}^{n2^{-j}} d_{jk} \psi_{jk}(t)}_{D_j(t)}, \quad (4)$$

where j , k , J and n represent the dilation parameter, translation parameter, number of scales, and number of observations in the original signal, respectively (Strang, 1989; Daubechies, 1988; Mallat, 1989). d_{jk} and a_{Jk} represent the scaling and the wavelet coefficients, respectively, and $A_J(t)$ and $D_j(t)$, ($j = 1, 2, \dots, J$) are the approximated signal and the detail signal, respectively.

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) (Sheriff et al., 2014). 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 2. The signals in Figures 2(b, d and f) are at increasingly coarser scales compared with the original signal in Figure 2(a).

Figure 2: Illustration of a data representation at multiple scales of a heavy-sine signal.

Multiscale representation is an effective method for dealing with autocorrelated or non-Gaussian data (Ganesan et al., 2004). In the next section, we highlight the advantages of multiscale representation.

3.2. Advantages of multiscale representation in PLS modeling

Since practical process data, including chemical and environmental process data, generally have multiscale properties, modeling such data requires a multiscale modeling approach that exploits the advantages of multiscale denoising to enhance the prediction quality of the PLS model. Some of the advantages of multiscale filtering in the PLS model are as follows (Madakyaru et al., 2013b).

- One of the factors affecting the performance of anomaly detection methods such PLS is the presence of autocorrelated measurement noise, which can be introduced by modeling errors. In this regard, one important advantage of multiscale representation is that the wavelet coefficients of autocorrelated data are approximately decorrelated at multiple scales (Aradhye et al., 2003; Ganesan et al., 2004). The decorrelation of noise at multiple scales will improve the effectiveness of fault detection methods if applied using detailed signals at multiple scales.

- Non-Gaussian errors can be introduced by malfunctioning sensors that introduce bias or skewed randomness to the measurements or by modeling errors that leave non-modeled process variations in the model residuals. The multiscale decomposition achieved using wavelet analysis also causes the distribution of data to become more Gaussian on multiscale levels (Ganesan et al., 2004). Since the performance of various anomaly detection techniques can deteriorate as the data deviate from normality, this property of multiscale representation should help enhance the performance of these techniques in the case of non-Gaussian data.
- One of the biggest advantages of multiscale representation is its capacity to distinguish measurement noise from useful data features (Harrou et al., 2013b; Madakyaru et al., 2013b) by applying low- and high-pass filters to the data during multiscale decomposition. This allows the separation of features at different resolutions or frequencies, which makes multiscale representation a better tool for filtering or denoising noisy data than traditional linear filters, like the mean filter and the exponentially weighted moving average (EWMA) filter (Sheriff et al., 2014). The ability of multiscale representation to separate noise has been used not only to improve data filtering, but also to improve the prediction accuracy of several empirical modeling methods and the accuracy of state estimators.

Here, we exploit these to improve the quality of PLS models via the development of an algorithm that merges multiscale denoising and PLS models. Before discussing multiscale PLS modeling, we present a brief description of multiscale denoising.

3.3. A multiscale data filtering algorithm

Multiscale denoising via wavelets is based on the observation that random errors in a signal are present over all wavelet coefficients while deterministic changes are captured in a small number of relatively large coefficients (Donoho et al., 1993, 1995; Bakshi, 1999). Wavelet-based denoising algorithms comprise the following three main steps (Donoho et al., 1995):

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

Several threshold selection criteria have been proposed including fixed threshold (Donoho and Johnstone, 1994), rigorous sure threshold and min-max threshold. The simplest wavelet threshold method, which was proposed by (Donoho and Johnstone, 1994), uses the same threshold to deal with coefficients in the expansion. It determines the threshold by the following form:

$$t_j = \sigma_j \sqrt{2 \log n}, \quad (5)$$

where n denotes the length of the analyzed signal and σ_j is the standard deviation of the errors at scale j , which can be estimated from the wavelet coefficients at that scale by equation 6,

$$\sigma_j = \frac{\text{median}(|d_{jk}|)}{0.6745}. \quad (6)$$

3.4. A multiscale PLS (MSPLS) modeling algorithm

Multiscale PLS modeling amalgamates the benefits of multiscale denoising and PLS modeling to improve model prediction thus improving fault detection. 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,

$$\mathbf{y}_j = \mathbf{T}_j \mathbf{B}_j \mathbf{Q}_j^T - \mathbf{F}_j, \quad (7)$$

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 residual of the 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 the PLS model for proper removal of noise. One way to accomplish this integration between multiscale filtering and PLS modeling is to use the MSPLS modeling algorithm that is presented in Table 2 and schematically illustrated in Figure (3) (Madakyaru et al., 2013b):

Table 1: MSPLS modeling framework.

-
- (1) Split the data into two sets: training and testing.
 - (2) Pre-process the input\output data 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 3.3.
 - (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-validated mean square error.
 - (6) Select the PLS with the smallest cross-validated mean square error as the MSPLS model.
-

Figure 3: A flow diagram of the integrated MSPLS modeling algorithm.

After a model is obtained using the MSPLS method, various methods for fault detection can be applied. We first run MSPLS on normal operating data (training data) to enable us to obtain a pair of models. Abnormal events are detected if the measurements deviate from the region of normal operation in the latent variable space or in the residual space. In MSPLS-based monitoring, two monitoring statistics, the T^2 and Q statistics, are usually utilized for fault detection purposes (Qin, 2003). First, the Hotelling T^2 statistic indicates the variation within the process model in the LVs space. The Q statistic, also known as the Squared Prediction Error (SPE), monitors how well the data conform to the model. Although the two methods have their advantages and disadvantages, both tend to fail to detect small faults (Harrou et al., 2013c). Here, we use only the Q -based chart as a benchmark for fault detection with PLS and MSPLS. Motivated by the power of the GLR test for detecting additive shifts in the process mean (Harrou et al., 2014; Montgomery, 2005), we propose an innovative MSPLS-based GLR fault-detection method for multivariate processes. In the next section, we briefly describe the GLR test.

4. Generalized likelihood ratio test-based fault detection

Detecting a particular fault that occurs in a monitored process requires checking whether the current measurements are statistically different from the a priori known faultless measurements (i.e., measurements without anomalies). Fault detection, which is a binary decision making process, consists of identifying fault from non-fault events based on some relevant data features. In this work, we present a fault detection algorithm based on the uncorrelated residuals obtained from the MSPLS model. To make a decision regarding

the process performance we compute a GLR decision statistic using output residuals from the MSPLS model. We compare GLR statistic with a threshold value. If the GLR statistic is below the threshold value, then we conclude that the process is under control. Otherwise, a fault signal is given. In the following, we briefly discuss the basic idea of the GLR test and how it can be used in fault detection.

4.1. GLR hypothesis testing

A general methodology for deriving a testing procedure for a composite hypothesis-testing problem is the GLR described here. GLR hypothesis testing is a well-known algorithm for statistical decision-making. It chooses between two composite hypotheses (Harrou et al., 2013c, 2009; Lehmann, 1996; Harrou et al., 2013a). In binary hypothesis testing, when hypotheses are composite or the corresponding data probability density functions contain unknown parameters, the GLR test is a popular means for deciding between two possibilities. Specifically, it is based on the maximization of the likelihood ratio function over all possible faults (Lenz et al., 2012), which usually makes it applicable to most parametric hypothesis-testing problems.

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

$$\begin{aligned} \mathcal{L}(Y) &= 2 \log \frac{\sup_{\theta \in \mathbb{R}^n} f_{\theta}(Y)}{f_{\theta=0}(Y)} \\ &= 2 \log \left\{ \sup_{\theta} \exp \left\{ -\frac{\|Y - \theta\|_2^2}{2\sigma^2} \right\} / \exp \left\{ -\frac{\|Y\|_2^2}{2\sigma^2} \right\} \right\}, \end{aligned} \quad (8)$$

where $\|\cdot\|_2$ is the Euclidean norm, $f_{\theta}(Y) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \|Y - \theta\|_2^2 \right\}$ is the probability distribution function (PDF) of Y . Rewriting equation (8) gives:

$$\begin{aligned} \mathcal{L}(Y) &= \frac{1}{\sigma^2} \left\{ \min_{\theta} \|Y - \theta\|_2^2 + \|Y\|_2^2 \right\} \\ &= \frac{1}{\sigma^2} \left\{ \|Y - \hat{\theta}\|_2^2 + \|Y\|_2^2 \right\}. \end{aligned} \quad (9)$$

In equation (9), we obtain the maximum estimate of θ as $\hat{\theta} = \arg \min_{\theta} \|Y - \theta\|_2^2 = Y$. Substituting $\hat{\theta}$ into equation (9), we get

$$\mathcal{L}(Y) = \frac{1}{\sigma^2} \left\{ \|Y\|_2^2 \right\}. \quad (10)$$

When $\mathcal{L}$ exceeds the threshold $h(\alpha)$, the GLR test chooses $\mathcal{H}_1$; otherwise $\mathcal{H}_0$ is selected.

$$\mathcal{L}(Y) = 2 \log \frac{\sup_{\theta \in \mathbb{R}^n} f_{\theta}(Y)}{f_{\theta=0}(Y)} = \frac{1}{\sigma^2} \left\{ \|Y\|_2^2 \right\} \begin{cases} \geq \mathcal{H}_0 \\ \leq \mathcal{H}_1 \end{cases} h(\alpha). \quad (11)$$

Typically, the threshold, $h(\alpha)$, is chosen to achieve a desired probability of a false alarm, predefined a priori (Lehmann, 1996).

$$\mathbb{P}_0(\mathcal{L}(Y) \geq h(\alpha)) = \int_h^\infty f_0(y)dy = 1 - F_{\chi_1^2}(h) = \alpha. \quad (12)$$

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 the $(1 - \alpha)$ -th quantile of the χ_1^2 distribution. The power function is given by:

$$\begin{aligned} \beta_{\delta^*}(c^2) &= \mathbb{P}_\theta(\delta^*(Y) = \mathcal{H}_1) = \mathbb{P}_\theta(\mathcal{L}(Y) \geq h(\alpha)) \\ &= \int_h^\infty f_\lambda(y)dy = 1 - F_{1, \lambda(\theta)}(h), \end{aligned}$$

where $F_{1, \lambda}(Y)$ is the non-central $\chi^2(1, \lambda)$ distribution with one degree of freedom and non-centrality parameter $\lambda(\theta) = \frac{1}{\sigma^2} \|P_H^\perp \theta\|_2^2$.

The goal of this study is to exploit the advantages of the GLR test and those of MSPLS modeling to reach improved detection performance compared to the conventional PLS-based GLR method. We thus merge the GLR test with MSPLS to enhance its fault detection ability.

5. The MSPLS-based GLR fault-detection scheme

In this section, MSPLS is coupled with GLR hypothesis testing to design an innovative fault detection scheme with improved detection abilities. In general, we obtain the model first and then perform the fault detection procedure accordingly. MSPLS indicates the capabilities of the modeling and monitoring process at different frequency bands. MSPLS using wavelets is used for data denoising and for reducing autocorrelation in the data. After the reference MSPLS model is identified, it is used to monitor the abnormal events (faults) in the process that may lead the process to depart from its normal state. Specifically, the residuals of the response variables are used as indicators of faults. Combining the advantages of MSPLS with those of the GLR monitoring scheme should result in an improved fault detection system.

In this approach, the GLR test is applied to the residuals of the response variables obtained from the MSPLS model. As given in Equation (7), the output vector, $\mathbf{y}$, can be written as the sum of a predicted vector, $\hat{\mathbf{y}}$, and a residual vector, $\mathbf{F}$, i.e.,

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{F}. \quad (13)$$

The residual of the output variable, $\mathbf{F} = [f_1, \dots, f_t, \dots, f_n]$, which is the difference between the observed value of the output variable, y , and the predicted value, $\hat{y}$, obtained from the MSPLS model is a potential fault indicator. Under nominal conditions, no abnormalities occur in the monitored process; thus, the value

of residuals fluctuates around zero due to measurement noise. A significant departure from zero residual 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 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, we test the following hypotheses:

$$\begin{cases} \mathcal{H}_0 = \{\mathbf{F} \sim \mathcal{N}(0, \sigma^2 I_n)\}, & (\text{null hypothesis}); \\ \mathcal{H}_1 = \{\mathbf{F} \sim \mathcal{N}(\theta, \sigma^2 I_n)\}, & (\text{alternative hypothesis}). \end{cases} \quad (14)$$

To test whether $\mathcal{H}_0$ should be rejected in favor of $\mathcal{H}_1$, we use the GLR test presented in Section 4.

Table 2: MSPLS-based GLR fault detection algorithm.

1. **Given:**

- 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 (7),

4. **MSPLS-based monitoring phase:**

- (1) For a new sample data, apply the same scaling used in the 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, $h(\alpha)$.
-

6. Monitoring a simulated distillation column

In this section, the ability of the proposed MSPLS-GLR technique to detect faults is applied to simulated data and the results are compared with those obtained using the traditional PLS-GLR method. In all monitoring charts, the red-shaded area is the region where the fault is injected to the test data while the 95% control limits are plotted by the horizontal dashed line.

6.1. Process description

The method is tested using a distillation column process simulated by Aspen (see Madakyaru et al. (2013a) for details) with added zero-mean Gaussian noise, where the predictor variables consist of ten temperatures (T_c) 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. A distillation process, which is one of the most common operations in the chemical industry, is schematically shown in Figure 4. The Aspen simulator is used to generate 1024 data samples to be used in constructing the reference MSPLS model. The parameters of the nominal steady-state operating conditions used in the distillation column simulation are given in Table 3. Figure 5 shows the dynamic input-output data of the distillation column around the nominal operating condition to which noise of Signal-to-Noise Ratio (SNR) of 10 is added. These data are used for model development and testing purposes. The first 512 data points are used for training the PLS model and the latter 512 data points are used for testing purposes. Based on cross-validation technique, three LVs are needed for the MSPLS model.

Table 3: Parameters for the simulated distillation column.

Process Variable	Value	Process Variable	Value
<u>Feed</u>			
F	1 kg mole /sec	P	1.7022×10^6 Pa
T	322 K	x_D	0.979
P	1.7225×10^6 Pa	<u>Reboiler Drum</u>	
z_F	0.4	B	0.5979 kg mole /sec
<u>Reflux Drum</u>		Q	2.7385×10^7 Watts
D	0.40206 kg mole /sec	T	366 K
T	325 K	P	1.72362×10^6 Pa
Reflux	62.6602 kg /sec	x_B	0.01

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad (15)$$

Figure 4: A distillation column.

Figure 5: Simulation of a distillation column: Variation of input-output data with SNR=10 (solid red line: noise-free data; blue dots: noisy data).

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \text{mean}(Y))^2}, \quad (16)$$

where y_t are the measured values, $\hat{y}_t$ are the corresponding predicted values by the MSPLS model and n is the number of samples.

A MSPLS model is fitted to the scaled training dataset. The goodness-of-fit measurements are shown in Figure 6. The right-hand side of Figure 6 shows the scatter plot of observed versus predicted values of the testing dataset obtained from the selected MSPLS model and the regression line. The points are distributed along the regression line, i.e., for the studied time series, 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. The models were thus 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 6 the scatter plots of observed and predicted data indicate a reasonable performance of the selected models, where y_t are

Figure 6: (Left panel): Plot of observed and predicted training data via the MSPLS model for the case where SNR=10. (Right panel): Scatter plots of predicted and observed training data.

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 predicts well with $R^2 = 0.97$ and a low RMSE of 0.028.

In general, we have to obtain the model first and then perform fault detection procedures accordingly. Before applying the MSPLS-GLR chart to fault detection, we need to check whether the residuals of the response variables follow a Gaussian distribution to ensure that the data are well represented using a linear MSPLS model. We can use a Q-Q plot to evaluate the normality of the residuals of the response variables. Another straightforward method to get a rough idea of the normality of residuals is to plot a histogram of the residuals. The Q-Q plots in Figure 7 evaluates the normality of the output residuals. The curve of the left of shows that the output residual obtained from MSPLS is normal; the curve of the right confirms that the residuals are normal. The histogram on the right panel of Figure 7 indicates that the normality assumption appears to be a reasonable one.

Next, we check for the absence of autocorrelation in the residuals of the MSPLS model, which is assumed to be uncorrelated. If this assumption is satisfied, the autocorrelation function of the residuals will have no significant spikes at any non-zero lags. Figure 8 indicates that the residuals of the response variables are not significantly correlated.

Figure 7: Tests for normality of residuals (left panel). Histogram showing the normality of the residuals. The histogram of residuals obtained from the MSPLS model has an approximately normal distribution (solid curve). (Right panel) Henry's line.

Figure 8: Autocorrelation function of the output residuals of MSPLS model.

6.2. Detection results

After the model is identified, it is used to monitor abnormal events (faults) in the distillation column process that may lead the process to depart from its normal state. Three different cases of faults were simulated to assess the proposed algorithms: an abrupt fault, an intermittent fault and a drift fault. We compared the results with results from a PLS-GLR monitoring scheme.

The false alarm rate (FAR) and miss detected rate (MDR) are often used to assess the efficiency of different monitoring techniques. MDR captures the number of faults that are wrongly judged as normal observations (missed detection) over the total number of faults. The FAR captures the number of normal observations that are wrongly judged as a fault (false alarm) over the total number of fault-free data.

$$FAR = \frac{\text{number of false alarm incidents}}{\text{number of fault free observations}} \quad (17)$$

$$MDR = \frac{\text{number of missed detection}}{\text{total number of faults}} \quad (18)$$

The smaller the values of MDR and FAR, the better the performance of the corresponding monitoring method.

SNR is typically defined as the ratio of the signal power to the average noise power, and it is typically measured in dBs. Here, SNR is measured as the ratio of the standard deviation of the signal's excitation to the standard deviation of the noise. More specifically, after the error-free data set is generated, measurement errors are added to the true values at each time instant to simulate the measured values. The measurement error at each time instant is a random normally distributed vector with a zero-mean and a specified standard deviation (noise level). Changing the noise levels means changing the measurement accuracy and, consequently, changing the SNR.

6.2.1. Case (i): Abrupt anomaly - Bias sensor anomaly

In this case study, the detection of an abrupt fault in the temperature sensor is investigated. Two sets of testing data were generated with high and low SNR and then MSPLS and PLS were applied to them. Towards this end, a bias anomaly, which is 2% of the total variation in temperature, T_{c3} , was incorporated into the temperature sensor measurements, T_{c3} , between samples 100 and 150. For data with $SNR = 30$, the performances of the PLS and MSPLS-based Q and GLR charts are shown in Figure 9(a)-(d). PLS- Q is ineffective, as expected, when the magnitude of the fault is relatively small (see Figure 9(a)). Figure 9(c) shows that the MSPLS- Q chart did not detect this simulated mean shift. In this example, the PLS-GLR test resulted in $FAR=10.43\%$ and $MDR=9\%$. MSPLS-GLR correctly detected this bias faults without false alarms (see Table 4). Results of the PLS and MSPLS-based Q and GLR charts in the case of $SNR=5$ are shown in Figure 9(a)-(d). This example shows that the PLS and MSPLS-based Q chart did not detect the mean shift in this case (see Figure 9(a) and (c)). Figure 10(a) shows that the PLS-GLRT is capable of detecting this fault but with the expense of a lot more false alarms and missed detections (i.e., $FAR=46.84\%$ and $MDR=33\%$). The plot in Figure 10(b), clearly shows the capability of our proposed MSPLS-GLR monitoring method to detect small anomalies without false alarms. In the last example with $SNR=5$, the MSPLS-GLR chart provides good monitoring performance, and the PLS-GLR chart yields a high missed detection and false alarm rates. This finding can be attributed to the noisy data that can affect the residuals. With low SNR values, it becomes more difficult for the PLS model to accurately predict output.

Of course, for data with high SNR values, PLS-GLRT provides an acceptable detection performance but it becomes inefficient for data with low SNR values (see Table 4). On the other hand, our MSPLS-GLR method gives good results no matter whether the SNR values were high or low (see Table 4). This is because MSPLS modeling provides an optimal solution for decoupling the signal from the noise. It is a great improvement over conventional PLS. Table 4 shows that when SNR values increase, the MDR and

Figure 9: Monitoring results of PLS-Q chart (a), PLS-GLR chart (b), MSPLS-Q chart (c) and MSPLS-GLR chart (d) in the presence of a bias anomaly in the temperature sensor measurements ' T_{c3} ' with SNR=30 (Case (i), first example).

Figure 10: Monitoring results of PLS-Q chart (a), PLS-GLR chart (b), MSPLS-Q chart (c) and MSPLS-GLR chart (d) in the presence of a bias anomaly in the temperature sensor measurements ' T_{c3} ' with SNR=5 (Case (i), third example).

FAR increase quickly and the performance of the PLS-GLR chart decreases. In addition, compared with the PLS-based Q , the GLR chart and the MSPLS- Q chart, MSPLS-GLR significantly increased the detection efficiency of incipient faults. The superiority of the MSPLS-GLR method over other charts can be confirmed by the results in Table 4, showing that the MSPLS-GLR method has the lowest FAR and MDR compared to the PLS-GLR monitoring method.

Table 4: Percentage of false alarms and missed detections of PLS-GLR and MSPLS-GLR monitoring statistics (Case (i)).

Statistical chart	Case (i) - Example 1		Case (i) - Example 2	
	FAR	MDR	FAR	MDR
PLS-Q	100	100	0.48	97
PLS-GLRT	10.43	9	46.84	33
MSPLS-Q	100	100	100	100
MSPLS-GLRT	0	0	0	6

6.2.2. Case (ii): Intermittent anomalies - intermittent bias sensor anomaly

In this case study, the performance of the MSPLS-GLR method is assessed and compared to that of the PLS-GLR method by their ability to detect intermittent anomalies in a simulated distillation column. Two sets of training data were generated with high and low SNR (i.e., SNR=30 and SNR=5) and then a small bias level of 2% of the total variation in temperature, T_{c3} , was introduced between the sample intervals [150, 250] and [350, 400]. Application of the PLS and MSPLS-based Q and GLR charts to the testing data with SNR=30 is shown in Figure 11(a)-(d). The PLS and MSPLS-based Q charts exhibit very poor fault detection performance (see Figure 11(a) and (c) and Table 5). This example shows that the Q chart is not sensitive to small faults. The PLS-GLR chart is shown in Figure 11(b). The PLS-GLR statistic clearly violate the control limit and thus the ability of this chart to detect this anomaly, but it resulted in several false alarms and missed detections (i.e., FAR=13.14% and MDR=15.5%). Figure 11(d) shows that the MSPLS-GLR method performs well without false alarms but with a few missed detections (MDR=2.5%).

Figure 11: Monitoring results of PLS-Q chart(a), PLS-GLR chart (b), MSPLS-Q chart (c) and MSPLS-GLR chart (d) in the presence drift sensor anomaly in ' T_{c3} ' with SNR=30 (Case (ii), first example).

Monitoring results of the case with SNR=5 are given in Figure 12(a)-(d). Figure 12(a) and (c) show that the PLS and MSPLS-based Q charts cannot detect intermittent faults effectively (see Table 5). The PLS-GLR method resulted in a FAR and a MDR of 46.47% and 30.5%, respectively. The MSPLS-GLR method resulted in a lower FAR and MDR of 0.64% and 5.5%, respectively. Compared with all other charts, MSPLS-GLR chart had significantly increased detection efficiency of incipient faults (see Table 5). The results show that our proposed method provides favorable performance for the detection of intermittent faults (Table 5). As a matter of the fact, the results of this case study confirm that merging MSPLS modeling with GLR testing enhances the ability to detect incipient faults compared with conventional PLS-based charts.

Figure 12: Monitoring results of PLS-Q chart(a), PLS-GLR chart (b), MSPLS-Q chart (c) and MSPLS-GLR chart (d) in the presence drift sensor anomaly in 'Tc3' with SNR=5 (Case (ii), second example).

Table 5: Percentage of false alarms and missed detections of PLS-GLR and MSPLS-GLR monitoring statistics (Case (ii)).

Statistical chart	Case (ii) - Example 1		Case (ii) - Example 2	
	FAR	MDR	FAR	MDR
PLS-Q	0	96	0	95
PLS-GLRT	13.14	15.5	46.47	30.5
MSPLS-Q	0	71.5	0	87
MSPLS-GLRT	0	2.5	0.64	5.5

6.2.3. Case (iii): Gradual anomaly - Slow drift sensor anomaly

This case is aimed to assess the potential of the MSPLS-based GLR anomaly detection scheme to detect a slow drift anomaly. Two sets of testing data were generated with high and low SNR respectively (i.e., SNR=30 and SNR=5). A slow drifting sensor anomaly with a slope of 0.01 was added to the temperature

sensor, T_{c3} , starting at sample 250 and lasting until the end of the testing data. For the testing data with SNR=30, monitoring results of PLS and MSPLS-based Q and GLRT statistics are shown in Figure 13(a)-(d). The PLS- Q chart is presented in Figure 13(a), in which we can see that a signal is first given at sample 313 with a significant false alarm rate (i.e., FAR=22.4%). Figure 13(b) shows that the PLS-GLRT chart detected an anomaly at sample 300, but with several false alarms (i.e., FAR=13.2%). Figure 13(c) shows the monitoring results of the MSPLS- Q chart. A signal is first detected at sample 323. The new chart, MSPLS-GLRT, increased linearly from sample 250, exceeding the control limits at signal 295 (see Figures 13(b)). The superiority of the MSPLS-GLRT chart over the PLS-based Q and GLRT and MPLS-based- Q chart is verified again, both in its sensitivity and detection rate.

Figure 13: Monitoring results of PLS- Q chart(a), PLS-GLR chart (b), MSPLS- Q chart (c) and MSPLS-GLR chart (d) in the presence drift sensor anomaly in ' T_{c3} ' with SNR=30 (Case (iii), first example).

For testing data with SNR=5, results of the PLS and MSPLS-based Q and GLR charts are shown in Figure 14(a)-(d). Figure 14(a) shows that the PLS- Q method is indeed ineffective in detecting small and persistent mean shifts when the SNR is small. It can be observed that the performance of the PLS-GLR method can be significantly affected by data with low SNR and can yield a high false alarm rate (FAR=51.6%), see Figure 14(b). The MSPLS- Q chart is shown in Figure 14(c), which indicates the first signal at the sample 364. The results of the MSPLS-GLR method which are given in Figure 14(d), clearly show the capability of this chart in detecting this small anomaly without false alarms. The plot shows that the MSPLS-GLR chart detects the first signal at the 304th observation. Therefore, fewer extra observations are needed for the MSPLS-GLR chart to detect a signal compared to the other charts. This case study clearly shows the superiority of the MSPLS-GLR method over the PLS-GLR method.

In summary, the idea behind the MSPLS fault detection method is to construct multiple PLS models

Figure 14: Monitoring results of PLS-Q chart(a), PLS-GLR chart (b), MSPLS-Q chart (c) and MSPLS-GLR chart (d) in the presence of a drift sensor anomaly in ' T_{c3} ' with SNR=5 (Case (iii), second example).

using wavelet coefficients (detail signals) at different scales, and then use these models in process monitoring. Thus, the developed MSPLS-based GLR process monitoring algorithm is more sensitive to anomalies than the conventional PLS approach because wavelet representation is an efficient denoising tool and wavelet coefficients are less autocorrelated than the actual process data. From this case study, it can be seen that using multiscale representation to develop a MSPLS modeling algorithm and then merging it with GLR hypothesis testing improve the fault detection abilities of PLS-based hypothesis testing fault detection methods even further.

7. Conclusion

Statistical process control is an important statistical tool for monitoring chemical processes. Data observed from chemical processes are usually noisy and correlated in time, which makes the fault detection more difficult as the presence of noise degrades fault detection quality and most methods are developed for independent observations. Multiscale representation of data using wavelets is a powerful feature-extraction tool that is well suited to denoising and decorrelating time series data. This paper proposes an innovative statistical method to monitor multivariate input output systems, which is based on a multiscale PLS (MSPLS) algorithm and generalized likelihood ratio (GLR) test. MSPLS has been used in this work as a modeling framework for fault detection using GLR hypothesis testing. In addition, MSPLS using wavelets is used for data denoising and reducing autocorrelation in the data. The GLR test is applied on the uncorrelated residuals obtained from the MSPLS model. Data of the simulated distillation column are 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 PLS-based GLR method.

This work can be extended to handle nonlinear processes with uncertainty in the measurements. To do that, we plan to develop a fault-detection approach based on multiscale kernel PCA or multiscale kernel PLS.

Acknowledgement

The work reported in this paper was supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No: OSR-2015-CRG4-2582.

References

References

- Abdi, H., Williams, L., 2010. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (4), 433–459.
- Aradhye, H., Bakshi, B., Strauss, R., Davis, J., 2003. Multiscale SPC using wavelets: theoretical analysis and properties. *AIChE Journal* 49 (4), 939–958.
- Bakshi, B., 1998a. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE journal* 44 (7), 1596–1610.
- Bakshi, B., 1998b. Multiscale pca with application to multivariate statistical process monitoring. *AIChE Journal* 44, 1596–1610.
- Bakshi, B., 1999. Multiscale analysis and modeling using wavelets. *Journal of Chemometrics* 13 (3), 415–434.
- Basseville, M., Nikiforov, I., 1993. Detection of abrupt changes: theory and application. Vol. 104. Prentice Hall Englewood Cliffs.
- Chiang, L., Braatz, R., Russell, E., 2001. Fault detection and diagnosis in industrial systems. Springer Science & Business Media.
- Daubechies, I., 1988. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics* 41 (7), 909–996.
- Donoho, D., Daubechies, I., Pierre, V., 1993. Wavelets on the interval and fast wavelet transforms. *Appli. Comput. Harmonic Analysis*, 1 (7), 64–81.
- Donoho, D., Johnstone, I., 1994. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81, 425–455.
- Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., Picard, D., 1995. Wavelet shrinkage: Asymptotia? *J. Roy. Stat. Soc. B* 57 (301).
- Ganesan, R., Das, T., Venkataraman, V., 2004. Wavelet-based multiscale statistical process monitoring: A literature review. *IIE transactions* 36 (9), 787–806.
- Gao, R., Yan, R., 2010. Wavelets: Theory and applications for manufacturing. Springer Science & Business Media.
- Gao, Z., Cecati, C., Ding, S., 2015. A survey of fault diagnosis and fault-tolerant techniques—part II: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Transactions on Industrial Electronics* 62 (6), 3768–3774.
- Geladi, P., Kowalski, B., 1986. Partial least-squares regression: a tutorial. *Analytica chimica acta* 185, 1–17.
- Godoy, J., Vega, J., Marchetti, J., 2013. A fault detection and diagnosis technique for multivariate processes using a PLS-decomposition of the measurement space. *Chemometrics and Intelligent Laboratory Systems* 128, 25–36.

- Harrou, F., Fillatre, L., Nikiforov, I., 2009. Bounded nuisance rejection and redundant sensor network. In: International Conference, System Identification and Control Problems, SICPRO. Vol. 9. pp. 26–30.
- Harrou, F., Fillatre, L., Nikiforov, I., 2014. Anomaly detection/detectability for a linear model with a bounded nuisance parameter. *Annual Reviews in Control* 38 (1), 32–44.
- Harrou, F., Madakyaru, M., Sun, Y., Khadraoui, S., 2016a. Improved detection of incipient anomalies via multivariate memory monitoring charts: Application to an air flow heating system. *Applied Thermal Engineering* 109, 65–74.
- Harrou, F., Nounou, M., Nounou, H., 2013a. Detecting abnormal ozone levels using PCA-based GLR hypothesis testing. In: Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on. IEEE, pp. 95–102.
- Harrou, F., Nounou, M., Nounou, H., 2013b. Enhanced monitoring using PCA-based GLR fault detection and multiscale filtering. In: IEEE Symposium on Computational Intelligence in Control and Automation (CICA). IEEE, pp. 1–8.
- Harrou, F., Nounou, M., Nounou, H., Madakyaru, M., 2013c. Statistical fault detection using PCA-based GLR hypothesis testing. *Journal of loss prevention in the process industries* 26 (1), 129–139.
- Harrou, F., Nounou, M., Nounou, H., Madakyaru, M., 2015. PLS-based EWMA fault detection strategy for process monitoring. *Journal of Loss Prevention in the Process Industries* 36, 108–119.
- Harrou, F., Sun, Y., Khadraoui, S., 2016b. Amalgamation of anomaly-detection indices for enhanced process monitoring. *Journal of Loss Prevention in the Process Industries* 40, 365–377.
- Hiroyuki, Y., Hideki, Y. B., Hiromu, F. C. E. O., Hideki, F., 2008. Canonical correlation analysis for multivariate regression and its application to metabolic fingerprinting. *Biochemical Engineering Journal* 40, 199–204.
- Höskuldsson, A., 1988. PLS regression methods. *Journal of chemometrics* 2 (3), 211–228.
- Isermann, R., 2006. Fault-diagnosis systems: an introduction from fault detection to fault tolerance. Springer Science & Business Media.
- Jia, Q., Zhang, Y., 2016. Quality-related fault detection approach based on dynamic kernel partial least squares. *Chemical Engineering Research and Design* 106, 242–252.
- Lee, G., Han, C., Yoon, E., 2004. Multiple-fault diagnosis of the tennessee eastman process based on system decomposition and dynamic PLS. *Industrial & engineering chemistry research* 43 (25), 8037–8048.
- Lehmann, E., 1996. Testing Statistical Hypotheses. Chapman and Hall, New York.
- Lenz, H.-J., Schmid, W., Wilrich, P.-T., 2012. *Frontiers in statistical quality control* 10. Vol. 10. Springer Science & Business Media.
- Li, X., Yao, X., 2005. Multi-scale statistical process monitoring in machining. *IEEE Transactions on Industrial Electronics* 52 (3), 924–927.
- Liang, W., Zhang, L., 2012. A wave change analysis (WCA) method for pipeline leak detection using gaussian mixture model. *Journal of Loss Prevention in the Process Industries* 25 (1), 60–69.
- MacGregor, J., Jaeckle, C., Kiparissides, C., Koutoudi, M., 1994. Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal* 40 (5), 826–838.
- Madakyaru, M., Nounou, M., Nounou, H., 2012. Linear inferential modeling: theoretical perspectives, extensions, and comparative analysis.
- Madakyaru, M., Nounou, M., Nounou, H., 2013a. Enhanced modeling of distillation columns using integrated multiscale latent variable regression. In: IEEE Symposium on Computational Intelligence in Control and Automation (CICA). IEEE, pp. 73–80.
- Madakyaru, M., Nounou, M. N., Nounou, H., 2013b. Integrated multiscale latent variable regression and application to distillation column. *Modelling and Simulation in Engineering* 2013, 1–7.
- Mallat, S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (7), 674–693.

- Montgomery, D. C., 2005. Introduction to statistical quality control. John Wiley& Sons, New York.
- Nomikos, P., MacGregor, J., 1995. Multi-way partial least squares in monitoring batch processes. *Chemometrics and intelligent laboratory systems* 30 (1), 97–108.
- Qin, S., 2003. Statistical process monitoring: Basics and beyond. *Journal of Chemometrics* 17 (8/9), 480–502.
- Qin, S., 2012. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control* 36 (2), 220–234.
- Shao, J., 1993. Linear model selection by cross-validation. *Journal of the American statistical Association* 88 (422), 486–494.
- Sheriff, M., Harrou, F., Nounou, M., 2014. Univariate process monitoring using multiscale Shewhart charts. In: *International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, pp. 435–440.
- Strang, G., 1989. Wavelets and dilation equations: A brief introduction. *SIAM review* 31 (4), 614–627.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S., Yin, K., 2003. A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & chemical engineering* 27 (3), 327–346.
- Wold, S., Ruhe, H., Wold, H., III, W. D., 1984. The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing* 5 (3), 735–743.
- Yin, S., Ding, S., Haghani, A., Hao, H., Zhang, P., 2012. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of Process Control* 22 (9), 1567–1581.
- Yin, S., Ding, S., Xie, X., Luo, H., 2014. A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics* 61 (11), 6418–6428.
- Yin, S., Wang, G., Gao, H., 2016a. Data-driven process monitoring based on modified orthogonal projections to latent structures. *IEEE Transactions on Control Systems Technology* 24 (4), 1480–1487.
- Yin, S., Xiangping, Z., Okayay, K., 2015. Improved PLS focused on key-performance-indicator-related fault diagnosis. *IEEE Transactions on Industrial Electronics* 62 (3), 1651–1658.
- Yin, S., Zhu, X., 2015. Intelligent particle filter and its application to fault detection of nonlinear system. *IEEE Transactions on Industrial Electronics* 62 (6), 3852–3861.
- Yin, S., Zhu, X., Qiu, J., Gao, H., 2016b. State estimation in nonlinear system using sequential evolutionary filter. *IEEE Transactions on Industrial Electronics* 63 (6), 3786–3794.
- Yoon, S., MacGregor, J., 2004. Principal-component analysis of multiscale data for process monitoring and fault diagnosis. *AIChE Journal* 50 (11), 2891–2903.
- Zhao, Y., Wang, S., Xiao, F., 2013. Pattern recognition-based chillers fault detection method using support vector data description (SVDD). *Applied Energy* 112, 1041–1048.
- Zhou, S., Sun, B., Shi, J., 2006. An SPC monitoring system for cycle-based waveform signals using Haar transform. *IEEE Transactions on Automation Science and Engineering* 3 (1), 60–72.