A novel anomaly detection scheme for high dimensional systems using Kantorovich distance statistic

Muddu Madakyaru1 · K. Ramakrishna Kini2

Received: 5 April 2022 / Accepted: 22 July 2022 / Published online: 15 August 2022
© The Author(s) 2022

Abstract The partial least squares (PLS) is a commonly applied multi-variate method in anomaly detection problems. The PLS strategy has been amalgamated with $T^2$ and squared prediction error (SPE) based statistical indicators to detect anomalies in process. These traditional indicators have few setbacks that has made them ineffective in monitoring applications. Hence, a statistical indicator based on Kantorovich distance (KD) is proposed for detecting sensor anomalies in this study. The proposed strategy integrates anomaly indicator based on KD metric with PLS based multi-variate method. The KD metric computes difference between the residuals of anomaly-free data and the data with anomaly and uses this distance as an indicator of anomaly. The proposed strategy’s critical feature is that a single anomaly indicator is sufficient to be integrated with PLS modeling framework. The Tennessee Eastman process benchmark and experimental distillation column processes data are used for assessing the performance of the proposed strategy. Further, comparisons have been provided between KD, $T^2$, SPE and Generalized Likelihood Ratio indicators. The results demonstrate the superiority of the KD statistical indicator in detecting sensor anomalies in comparison to the traditional indicators of PLS based strategy. The KD indicator integrated with PLS framework also enhances the detection of small magnitude anomalies.

Keywords Anomaly detection · Kantorovich distance · Partial least squares · Tennessee Eastman process · Experimental distillation column processes

1 Introduction

In modern world, the task of process monitoring is very much crucial for product quality as well as safety [1]. The automation-based modern revolution has brought a major improvement in petrochemical industries since it is possible to convert raw materials into final products at improved quality. However, the modern revolution also carries few risks and accidents due to the added complexity in the processes and leads to many accidents that possess a major threat to the environment. Possible manual error, poor maintenance of the process plant, and malfunctioning sensors are some of the main causes of these mishaps. The above-mentioned accidents can be controlled if the process plants are monitored continuously. The last few decades have seen improved monitoring techniques for automated processes which are built around anomaly diagnosis, anomaly detection, and isolation. The anomaly detection suggests if any anomaly is present in the system, anomaly diagnosis suggests the type of anomaly that has occurred and anomaly isolation points out the particular sensor groups that were responsible for the anomaly. With improved data acquisition methods and inexpensive data management systems available, a large volume of sensor measurement data can be recorded very easily.

The availability of sensor data has ensured frequent utilization of data-driven anomaly detection methods over their counterparts, the model-driven methods, where the duty of...
model development is exigent due to the complexity in process plants [2, 3]. The data-driven methods depend on the availability of quality process data that contains information on the process and seized intrinsic complexities within the process [4]. The uni-variate and multivariate techniques are the two sub-classes of data-driven anomaly detection methods. The multivariate strategies analyze high dimensional data to capture useful features through immeasurable variables (latent variables), thus having a clear advantage over uni-variate techniques that are restricted in analyzing single sensor measurement only [5]. Few regularly applied multivariate statistical methods include principal component analysis (PCA), partial least squares (PLS), and fisher discriminant analysis (FDA). The classic PCA technique reduces the dimension of data to retain information from the original data [6]. Though PCA has been widely applied in process monitoring problems, it represents input space without considering the features from output space.

Most chemical processes like continuous stirred tank heater and distillation column are represented using the relationship between input-output variables. Process monitoring can be enhanced if a technique can model input-output space and this is possible using partial least squares (PLS) strategy. The PLS method extracts latent factors from predictor (input) and response (output) variables to maximize the cross-correlation and generates a relationship between the two sets of variables. PLS method has also been used in soft-sensor applications to predict response variables using predictor variables [7, 8]. Recently, different variants of PLS strategy have been proposed in monitoring applications and this includes multi-block PLS [9], kernel PLS [10], dynamic PLS [11], modified PLS [12] and multi-scale PLS [13]. The anomalies occurring in the chemical process are monitored using $T^2$ and squared prediction error (SPE) anomaly indicators. However, the $T^2$ and SPE indicators are ineffective in monitoring since the decision regarding an anomaly is based only on previous observations. This makes the indicators insensitive to small magnitude faults which lead to missed detections and false alarms [14].

To overcome the shortcomings of $T^2$ and SPE indicators, the PLS method has been integrated with the generalized likelihood ratio (GLR) test to improve the task of monitoring in recent years. The GLR is a hypothesis-testing technique used to compare the goodness of fit between two data distributions. The multi-scale PLS and kernel PLS strategies have been integrated with GLRT and the performance was found to be superior than $T^2$ and SPE indicators [15, 16]. In recent years, the Kantorovitch Distance (KD) based statistical index has been applied for the task of detecting anomalies. The KD index is defined as the minimum cost needed to shift a mass of data between two distributions [17]. The KD metric has been found to have better detection of small magnitude anomalies and has provided better results when data was corrupted with noise [18, 19]. Considering the benefits offered by the KD index, an improved anomaly detection scheme is proposed that uses PLS as a modeling framework and KD as an anomaly indicator. The residuals from the PLS model capture process information and aids in providing anomaly information. Hence, the proposed PLS-KD strategy evaluates the residuals of anomaly-free data and the anomaly data to detect possible anomalies.

The organization of the paper is as follows: Sect. 2 provides an overview of PLS modeling, $T^2$, SPE, and GLR-based anomaly indicators. This is followed by the description of optimal mass transport and KD metric formulation in Sect. 3. Next, the proposed monitoring strategy that amalgamates PLS with KD metric, is discussed in Sect. 4. To assess the efficacy of the proposed PLS-KD strategy, two case studies are considered in Sect. 5. The performance of the PLS-KD scheme is contrasted against PLS-$T^2$, PLS-SPE, and PLS-GLR based strategies. In Sect. 6, the concluding remarks about the work are provided.

2 Partial least squares

PLS is a useful multi-variate tool which provides dimension reduction and prediction performance [20, 21]. It finds the relationship between input $X$ and output $Y$ by transforming both $X$ and $Y$ into their latent variables which is then followed by relating the transformed latent variables. Consider a data $X \in \mathbb{R}^{m \times p}$ and $Y \in \mathbb{R}^{m \times q}$ where $m$ and $n$ represents the number of variables and observation samples respectively. After normalization, the model is developed which involves transforming $X$ and $Y$ into lower dimensional subspace and it can be represented as [22]:

$$X = \hat{X} + F = \sum_{k=1}^{p} t_k p_k^T + F = TP^T + F$$
$$Y = \hat{Y} + E = \sum_{k=1}^{q} u_k Q_k^T + E = UQ^T + E$$

The Eq. (1) is described as the outer model of PLS. While $T \in \mathbb{R}^{m \times p}$ and $U \in \mathbb{R}^{n \times q}$ represents the matrix of transformed latent variables, $P$ and $Q$ represent the loading matrices respectively. The number of transformed latent variables (LVs) are $p$ where as $F$ and $E$ represent residual matrices. The latent variables $t_k$ and $u_k$ with $k = 1, 2, ..., p$, are extracted iteratively through non-linear iterative partial least squares (NIPALS) algorithm. For each PLS dimension, the NIPALS algorithm calculates two LVs that are called as $t$ and $u$ and they are a linear combination of $X$ matrix and $Y$ matrix. These variables are chosen such that the co-variance between $t$ and $u$ is maximized.
It is important to retain few optimum LVs in the process model to have good prediction and the cross-validation approach is utilized. Next, a relationship between latent variables of input and output data is developed via the PLS inner model that is described in Eq. (2) as follows:

$$U = TV + W$$  \hspace{1cm} (2)

where $V$ is the regression matrix and $W$ is the residual matrix. Once the PLS inner model is developed, the output $Y$ can be expressed as:

$$Y = TVQ^T + E^*$$  \hspace{1cm} (3)

where $E^*$ is a residual matrix which presents unexplained variance in Eq. (3). The NIPALS algorithm used in this work is summarized as follows: \[23\]

- **Step-1**: The $u$ is set equal to the column of $Y$.
- **Step-2**: $w = \frac{u^TW}{u^Tww}$.
- **Step-3**: The $u$ is normalized to unit length.
- **Step-4**: The scores are evaluated: $t = \frac{Xw}{w^Tw}$.
- **Step-5**: The new $u$ vector is evaluated: $u = \frac{Yt}{t^Tw}$.
- **Step-6**: The convergence on $u$ is checked: if YES, proceed to Step-7, else return to Step-1.
- **Step-7**: The $X$ loading $p = \frac{Xt}{t^Tt}$ is evaluated.
- **Step-8**: The residual matrices $E = X - tp^T$ and $F = Y - tq^T$ are evaluated.
- **Step-9**: If the additional PLS dimensions are required, $X$ and $Y$ get changed to $E$ and $F$ and then, step-1 to step-8 are repeated.

Once the reference PLS model is developed from data exhibiting normal operations, anomalies in online data can be captured using $T^2$ and SPE based indicators. The $T^2$ indicator monitors any changes in LV subspace of the model and for a new data, variations in the LVs at different time instances is represented as:

$$T^2 = \sum_{i=1}^p \frac{t_i^2}{\sigma_i^2}$$  \hspace{1cm} (4)

For the LV $t_i$, $\sigma_i^2$ is the corresponding estimated variance. In contrast, the SPE index checks for any abnormal variations in the residual subspace not defined by the LVs and it may be computed in the following way for online data:

$$SPE = e^Te$$  \hspace{1cm} (5)

where $e = x - \hat{x}$ is difference between data $x$ and predicted data $\hat{x}$. In other words, the SPE statistics defines the loss of fit from the developed PLS model. An anomaly is declared when the indicators described in Eqs. (4) and (5) exceed the threshold [24].

## 2.1 Generalized likelihood ratio test

In process monitoring problems, an anomaly is declared based on the decision taken by comparing present measurements with previous normally operating measurements. The simple and composite hypotheses have been usually employed in decision-making problems in literature. GLR is a statistical test that has been applied in solving problems related to the composite hypothesis. It solves the hypothesis problem by maximizing the likelihood ratio function over all possible anomalies. This test has been used as an anomaly indicator in many anomaly detection problems and has been found to detect faults better than conventional anomaly indicators of PLS and PCA-based strategies [16]. Consider an anomaly detection problem with $Z \in \mathbb{R}^n$ been generated by one of the distributions: $N(\mu, \sigma^2)$ or $N(\mu \neq 0, \sigma^2)$. The Eq. (6) presents the hypothesis problem, which is defined as:

$$H_A \equiv Z \sim N(\mu = 0, \sigma^2)$$

$$H_B \equiv Z \sim N(\mu \neq 0, \sigma^2)$$  \hspace{1cm} (6)

GLR test computes the ratio of anomaly-free data and the anomaly data while the unknown anomaly parameter $\Omega$ is replaced by maximum likelihood estimate and this is possible by maximization of likelihood ratio function which is defined as in Eq. (7):

$$L(a) = 2\log \frac{\sup_{\Omega \in \mathbb{R}^n} f_{\Omega}(Z)}{f_{\Omega=0}(Z)}$$

$$= 2\log \frac{\sup_{\Omega} \exp \left( - \frac{Z - \Omega^2}{2\sigma^2} \right)}{\exp \left( - \frac{Z^2}{2\sigma^2} \right)}$$  \hspace{1cm} (7)

where $f_{\Omega}(Z) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left( - \frac{Z - \Omega^2}{2\sigma^2} \right)$ is probability density function of $Z$. Re-writing the above expression and solving yields [15]:

$$L(a) = \frac{1}{\sigma^2} \left( \frac{Z^2}{2} \right)$$  \hspace{1cm} (8)

The GLR $L(a)$ in Eq. (8) adheres to chi-square $\chi^2$ distribution with $l$ degrees of freedom since $Z$ adheres to gaussianity. If $L(a)$ is less than $v(\lambda)$, the null hypothesis is accepted suggesting no anomaly in the process. If $L(a)$ exceeds $v(\lambda)$, the alternate hypothesis is accepted which suggests the presence of anomaly.
3 Kantorovich distance

The optimal mass transport has been an important topic of research in the field of economics, image processing, fluids mechanics, and process applications. Optimal mass transport deals with the relocation of data by minimizing the cost function. A convenient mode of transferring information from the first to second destination distribution concerning a cost function is referred to as optimal mass transport. The optimal mass transport theory has been applied in machine learning, medical, and image processing problems [17, 25]. It may be noted that the optimal transport problem was initially formulated by Monge who proposed a problem where a mass from source distribution can be moved to only one point in destination distribution [26]. Hence, in this formulation, there is no transport plan to rearrange one point to multiple points in the source distribution. The formulation problem was further improved by Kantorovich where it was proposed that multiple data from one distribution can be mapped to multiple data in a second and this method was named Kantorovich Distance.

Since anomaly detection problems detect anomalies by measuring the dissimilarity between data under normal operating conditions and the data with the anomaly, the KD statistic can be applied in anomaly detection problems. Moving a mass of data from the first destination to the second with respect to a cost function is termed Kantorovich distance (KD). For distributions \( G \) and \( H \), minimum shifting distance needed for the data from \( G \) to \( H \) is the Kantorovich distance. The cost is minimal when distributions are similar and large when distributions are dissimilar [27]. The cost function is defined based on the norm in Eq. (9):

\[
c(g, h) = g - h'
\]

When \( r=1 \), distance between the distributions relative to cost function is termed as 1-Wasserstein Distance and this is presented in Eq. (10):

\[
W_r(G, H) = \left( \inf_{r \in \Gamma(G, H)} \int g - h' d\gamma(g, h) \right)^{1/r}
\]  

(10)

\( \Gamma(G, H) \) represents set of joint distributions. For \( r=2 \), Kantorovich Distance for two \( n \)-dimensional random variables \( g \) and \( h \) is expressed as:

\[
W_2(G, H) = \left( \mu_g - \mu_h \right)^2 + Tr\left( \Sigma_g - \Sigma_h - 2\Sigma_g \frac{1}{\Sigma_g} \Sigma_h \frac{1}{\Sigma_h} \right) \right)^{1/2}
\]  

(11)

where \( \mu_g \) and \( \mu_h \) indicate the means while \( \Sigma_g \) and \( \Sigma_h \) indicate the covariance matrices of variables \( g \) and \( h \) in Eq. (11). The computation of the KD metric is based purely on the available data and hence, it is not restricted to any particular distribution. This makes the KD metric flexible to be applied for data adhering to any statistical distribution. The KD index involves stacking data observations into segments and evaluating them, thus providing a smooth representation of the KD metric. The KD metric implementation is dependent on the number of segments in data sets and the moving window. A proper selection of both these parameters gives a smooth transition of the KD metric. In this paper, the number of segments in each distribution is represented by \( r \) and the moving window is represented by \( j \). The computation of the KD metric between two distributions can be found in [18].

4 The PLS-KD anomaly detection strategy

In this work, a PLS model is constructed from normal operating data extracted from the process and the model is used for monitoring any abnormalities in new data through KD statistical index. The residuals from the PLS model can reveal useful process information and can be assessed to determine abnormalities in the process. When the process is operating customarily, residuals of the process are near zero. In the case of atypical process behavior, the process residuals deviate from zero thus, indicating that the process needs attention. The residuals from the PLS model are described mathematically in Eq. (12):

\[
E = Y - \hat{Y}
\]  

(12)

Next, the KD metric is calculated between the residuals of training and the testing data. The KD metric then undergoes comparison with a threshold to decide the anomaly. The PLS-KD anomaly detection strategy can be mainly divided into two stages:

1. Computation of threshold using normal operating data.
2. Monitoring new data using KD metric.

The kernel density estimation (KDE) approach is used for determining reference threshold [19]. The probability density function \( f(y) \) of a random variable is estimated by \( \hat{f}(y) \) samples \( y_i, i = 1, 2, \ldots n \):

\[
\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{y - y_i}{h} \right)
\]  

(13)

where \( K \) represents uni-variate kernel estimator and \( h \) is smoothing parameter in Eq. (13). Initially, the training data are divided into \( Tr1 \) and \( Tr2 \) respectively. This is followed by developing a PLS model and generating the residuals \( Ra1 \) and \( Ra2 \). The KD statistic \( KD \) is calculated between the two residuals. Next, the density function of the KD index
undergoes estimation through the use of a kernel estimator. The point that occupies 99% of the total area under the density function plot is considered as the reference.

Once the reference threshold is computed, the next task is to monitor anomalies in new online data. The PLS-KD strategy is presented in Fig. 1. The off-line anomaly-free data undergoes normalization and progresses for PLS model development from which the residuals $G$ are generated. When new data is available, it undergoes normalization, and residuals $H$ are generated model parameters. Next, each segment of $G$ undergoes comparison with each segment of $H$ for computation of KD metric. The statistic $KD$ is compared with the threshold $h$ to take decision regarding anomaly $\delta$ and this is presented in Eq. (14):

$$\delta = \begin{cases} 
\text{No anomaly} & \text{if } KD < h; \\
\text{Anomaly} & \text{if } KD > h; 
\end{cases} \quad (14)$$

5 Case studies

This section presents the monitoring results of PLS-KD strategy using two case studies. The PLS-KD strategy is compared with PLS-$T^2$, PLS-SPE and PLS-GLR strategies using fault detection rate (FDR) and false alarm rate (FAR) indicators [18].

5.1 Tennessee Eastman process

In the domain of process control, Tennessee Eastman (TE) benchmark is considered a standard for multi-variate control and anomaly detection tasks. Most researchers prefer validating the newly proposed anomaly detection strategy on this process [18]. The control structure of the TE process is presented in Fig. 2 and is made of a total 52 measurements that include 22 process measurements, 19 composition measurements, and 12 manipulated variables. The data has been downloaded from the following: http://depts.washington.edu/control/LARRY/TE/download.html [28]. In harsh environments having noise and disturbances, different real anomaly scenarios are proposed of which 15 are known anomalies while the remaining anomalies are defined as unknown. In this case study, the input matrix of the PLS model consists of 33 variables that include 22 process measurements and 11 manipulated variables while the composition variable 35 is treated as the output variable. The cross-validation technique is used to select 5 LVs for the PLS model while a moving window of 40 is employed for the PLS-KD formulation. The anomalies 3, 9, and 15 are not considered in this study since they give small detection rates [29].

The PLS-KD strategy is validated on different anomaly scenarios of the TE process. The monitoring performance of anomaly 14, which is a sticking anomaly in the cooling water valve, is presented. The Figs. 3 and 4 demonstrates the results of different methods in monitoring the anomaly. Though all strategies demonstrate good detection of this anomaly, the conventional anomaly detection schemes have few missed detections. However, the performance of the PLS-KD strategy is slightly better since it detects the anomaly with a higher detection rate and also has no false alarms. The FAR and FDR of the proposed PLS-KD, as well as PLS-$T^2$, PLS-SPE and PLS-GLR methods, are presented in Table 1 and Table 2. It is observed that the PLS-KD and PLS-GLR strategies detect anomaly scenarios better in comparison with PLS-$T^2$ and PLS-SPE based techniques. The PLS-KD strategy over-performs the PLS-GLR strategy as well as the conventional anomaly indicators with a better detection rate. However, it may be

![Flow chart for PLS-KD anomaly detection strategy](image-url)
observed that for a few cases all the strategies have a high FAR value, but overall, the FAR of the KD index is less as compared to other methods. It can be inferred that the PLS-KD strategy provides better detection of anomalies, thus, over-performing other methods with a smooth detection profile.

Fig. 2  TE process flowsheet [30]

Fig. 3  Monitoring of anomaly 14 by $T^2$ and SPE anomaly indicators
5.2 Experimental DC process

The distillation column (DC) is a key unit operation in the oil and gas industry. The experimental DC process used in this study is housed in the Department of Chemical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. Figure 5 depicts the bubble cap distillation process used in this work. The construction details of the bubble cap DC process can be found in [31]. The sequential perturbation of reflux, as well as feed flow, aids in the generation of process data. Once the DC process is maintained at nominal operating condition, the feed flow undergoes perturbation for a magnitude of 50 and the reflux flow is maintained constant. Next, the DC process is maintained again to nominal condition and then, the reflux flow is perturbed for a magnitude of 40 while maintaining the flow rate constant. All the above actions

---

**Table 1** FAR (%) for TE process

| Anomaly no. | PLS-T²   | PLS-SPE  | PLS-GLR | PLS-KD |
|-------------|----------|----------|----------|--------|
| 1           | 1.25     | 0.25     | 6.52     | 0.00   |
| 2           | 0.63     | 0.48     | 6.38     | 0.00   |
| 4           | 6.25     | 4.88     | 11.25    | 6.25   |
| 5           | 6.25     | 2.73     | 10.17    | 2.50   |
| 6           | 0.63     | 0.15     | 5.63     | 0.50   |
| 7           | 7.50     | 1.62     | 3.75     | 3.68   |
| 8           | 8.23     | 10.62    | 7.75     | 7.15   |
| 10          | 7.38     | 2.50     | 4.75     | 4.82   |
| 11          | 4.13     | 4.63     | 7.87     | 4.38   |
| 12          | 15.63    | 21.88    | 6.62     | 6.50   |
| 13          | 1.75     | 1.28     | 2.50     | 0.20   |
| 14          | 10.75    | 1.75     | 6.75     | 0.38   |
| 16          | 40.00    | 46.26    | 38.25    | 28.63  |
| 17          | 3.75     | 1.12     | 8.13     | 0.54   |
| 18          | 0.63     | 0.65     | 11.25    | 0.00   |
| 19          | 3.75     | 2.50     | 15.00    | 3.75   |
| 20          | 1.50     | 1.38     | 11.87    | 0.13   |
| 21          | 3.50     | 4.50     | 2.85     | 3.75   |

**Table 2** FDR (%) for TE process

| Anomaly no. | PLS-T²   | PLS-SPE  | PLS-GLR | PLS-KD |
|-------------|----------|----------|----------|--------|
| 1           | 98.12    | 99.25    | 99.05    | 99.50  |
| 2           | 97.25    | 97.88    | 97.85    | 98.50  |
| 4           | 43.25    | 47.88    | 41.88    | 63.00  |
| 5           | 35.75    | 35.63    | 45.37    | 68.50  |
| 6           | 99.25    | 99.50    | 99.50    | 99.50  |
| 7           | 78.37    | 74.62    | 90.75    | 91.50  |
| 8           | 93.23    | 94.75    | 97.25    | 97.75  |
| 10          | 61.38    | 62.00    | 60.12    | 63.12  |
| 11          | 56.13    | 58.25    | 57.87    | 70.38  |
| 12          | 96.00    | 97.88    | 96.62    | 97.50  |
| 13          | 95.75    | 95.12    | 94.50    | 96.50  |
| 14          | 90.73    | 85.62    | 96.75    | 98.15  |
| 16          | 58.37    | 55.00    | 66.25    | 80.63  |
| 17          | 78.15    | 73.12    | 86.13    | 94.54  |
| 18          | 89.62    | 91.38    | 92.25    | 92.05  |
| 19          | 15.50    | 8.13     | 18.37    | 31.25  |
| 20          | 55.87    | 52.38    | 57.75    | 70.13  |
| 21          | 35.38    | 30.88    | 28.25    | 40.52  |
result in a few modifications in the product quality (output) i.e xD which is recorded. Also, temperatures at five parts of the DC process are recorded to monitor the condition. Once the DC process data is generated, the next task is to develop a reference model. The input variables include flow rate, feed flow, and the temperatures in the DC process. The composition of light distillate (methanol) i.e., xD is considered as the output variable. Next, the input, as well as output matrices, are split equally to get training and testing data sets having 1000 observations each. The PLS model is developed on input training data while the cross-validation technique yields 2 LVs.

The proposed strategy is assessed by its ability to monitor bias, intermittent and drifts faults in the DC process. To begin with, a step signal is introduced in variable 5 between the observations 400 and the end of the testing data. The results of statistical indices in monitoring the anomaly is presented through Figs. 6 and 7. From Fig. 6, it is observed that both the conventional indicators fail in detecting this anomaly. Also, for PLS-$T^2$ strategy, false alarms are visible between sampling instants 0 and 180. Figure 7 demonstrates improved detection by the PLS-GLR method but has few missed detections in the region of anomaly and few false alarms in anomaly free region. In contrast, the KD metric demonstrates precise detection without missed detections or false alarms, thus having a clear advantage over $T^2$, SPE, and GLR-based anomaly indicators.

Even in the case of monitoring intermittent and drift anomalies, the PLS-KD strategy carries a better advantage since it detects these anomalies much better than other methods. The FAR and FDR of the proposed PLS-KD, as well as PLS-$T^2$, PLS-SPE, and PLS-GLR methods, are presented in Tables 3 and 4. In a summary, this case study demonstrates that PLS-$T^2$, PLS-SPE, and PLS-GLR methods are not effective in handling anomalies of small magnitude whereas the PLS-KD strategy detects it reasonably well. Hence, it is safe to conclude that the proposed PLS-KD strategy over-performs other strategies in handling different anomalies in an experimental distillation column process with a good detection rate.

![Fig. 5](image) A schematic of distillation column set-up

![Fig. 6](image) Monitoring of bias anomaly by $T^2$ and SPE anomaly indicators
6 Conclusion

A novel statistical anomaly indicator based on Kantorovich Distance was proposed in this study. The detection performance of $T^2$ and SPE anomaly indicators of PLS strategy was found to be abstemious. Hence, an alternative anomaly indicator based on the KD metric was integrated with the PLS modeling framework in this study. The KD metric was computed between the residuals of data under normal operating conditions and residuals of the testing data. The monitoring behavior of the integrated PLS-KD strategy was illustrated using the Tennessee Eastman process and distillation column process. The PLS-KD strategy was contrasted against PLS-$T^2$, PLS-SPE, and PLS GLR-based anomaly detection schemes. The KD statistic was able to capture sensitive process information due to segment-by-segment comparison and this aided in improved anomaly detection. The results reveal that the KD index can be effectively integrated with the PLS strategy to enhance anomaly detection. As a part of future work, the PLS scheme is planned to be integrated with wavelet functions to have a novel multi-scale PLS strategy. The multi-scale PLS scheme will be used as a modeling framework and the KD metric will be used as a fault indicator. As the process data is embedded with heavy noise, the effect of noise can be reduced using wavelets and a multi-scale PLS-KD based strategy can enhance the detection of small magnitude faults.

Acknowledgement The authors would like to thank the Manipal Institute of Technology (MIT), Manipal Academy of Higher Education (MAHE), Manipal, for supporting this work.

Funding Open access funding provided by Manipal Academy of Higher Education, Manipal.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Table 3 FAR (%) for experimental DC process

| Anomalies   | PLS-$T^2$ | PLS-SPE | PLS-GLR | PLS-KD |
|-------------|-----------|---------|---------|--------|
| Bias        | 32.50     | 0.00    | 22.00   | 0.00   |
| Intermittent| 31.25     | 1.00    | 2.75    | 0.00   |
| Drift       | 17.11     | 0.56    | 0.75    | 0.00   |

Table 4 FDR (%) for experimental DC process

| Anomalies   | PLS-$T^2$ | PLS-SPE | PLS-GLR | PLS-KD |
|-------------|-----------|---------|---------|--------|
| Bias        | 10.36     | 30.29   | 60.35   | 96.67  |
| Intermittent| 27.70     | 60.50   | 81.00   | 88.50  |
| Drift       | 80.92     | 75.56   | 81.45   | 86.50  |
References

1. Cheng H, Liu Y, Huang D, Cai B, Wang Q (2021) Rebooting kernel cca method for nonlinear quality-relevant fault detection in process industries. Process Saf Environ Prot 149:619–630
2. Li W, Peng M, Wang Q (2018) Fault identification in pca method during sensor condition monitoring in a nuclear power plant. Ann Nucl Energy 121:135–145
3. Kumar A, Bhattacharya A, Flores-Cerrillo J (2020) Data-driven process monitoring and fault analysis of reformer units in hydrogen plants: industrial application and perspectives. Comput Chem Eng 136:106756
4. Nor NM, Hasan CRC, Hussain MA (2019) A review of data-driven fault detection and diagnosis methods: applications in chemical process systems. Rev Chem Eng 36(4):513–553
5. Alauddin M, Khan F, Imtiaz S, Ahmed S (2018) A bibliometric review and analysis of data-driven fault detection and diagnosis methods for process systems. Ind Eng Chem Res 57:10719–10735
6. Zhang Q, Li P, Lang X, Miao A (2020) Improved dynamic kernel principal component analysis for fault detection. Measurement 158:107738
7. Facco P, Doplicher F, Bezzo F, Barolo M (2009) Moving average pls soft sensor for online product quality estimation in an industrial batch polymerization process. J Process Control 19:520–529
8. Wang D, Liu J, Srinivasan R (2010) Data-driven soft sensor approach for quality prediction in a refining process. IEEE Trans Ind Inf 6(1):11–17
9. MacGregor JF, Jaeckle C, Kiperissides C, Koutoudi M (1994) Process monitoring and diagnosis by multiblock pls methods. AIChE J 40(5):826–838
10. Zhang Y, Hu Z (2011) On-line batch process monitoring using hierarchical kernel partial least squares. Chem Eng Res Des 89(10):2078–2084
11. Ahn SJ, Lee CJ, Jung Y, Han C, Yoon ES, Lee G (2008) Fault diagnosis of the multi-stage flash desalination process based on signed digraph and dynamic partial least square. Desalination 228(1–3):68–83
12. Wang G, Yin S (2015) Quality-related fault detection approach based on orthogonal signal correction and modified pls. IEEE Trans Ind Inf 11(2):398–405
13. Lee HW, Lee MW, Park JM (2009) Multi-scale extension of pls algorithm for advanced on-line process monitoring. Chemom Intell Lab Syst 98:201–212
14. Harrou F, Madakaryu M, Sun Y (2017) Improved nonlinear fault detection strategy based on the Hellinger distance metric: plug flow reactor monitoring. Energy Build 143:149–161
15. Madakaryu M, Harrou F, Sun Y (2017) Improved data-based fault detection strategy and application to distillation columns. Process Saf Environ Prot 107:22–34
16. Botre C, Mansouri M, Nounou H, Nounou M, Karim MN (2016) Kernel pls-based glrt method for fault detection of chemical processes. J Loss Prev Process Ind 43:212–224
17. Li D, Martz S (2021) High-confidence attack detection via Wasserstein-metric computations. IEEE Control Syst Lett 5(2):379–384
18. Kini KR, Madakaryu M (2020) Improved process monitoring strategy using Kantorovich distance-independent component analysis: an application to tennessee eastman process. IEEE Access 8:205863–205877
19. Kini KR, Bapat M, Madakaryu M (2022) Kantorovich distance based fault detection scheme for non-linear processes. IEEE Access 10:1051–1067
20. Harrou F, Sun Y, Madakaryu M, Bouyedou B (2018) An improved multivariate chart using partial least squares with continuous ranked probability score. IEEE Sens J 18(16):6715–6726
21. Tong C, Lan T, Yu H, Peng X (2019) Distributed partial least squares based residual generation for statistical process monitoring. J Process Control 75:77–85
22. Kong X, Luo J, Xu Z, Li H (2019) Quality-relevant data-driven process monitoring based on orthogonal signal correction and recursive modified pls. IEEE Access 7:117934–117943
23. Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. Anal Chim Acta 185:151–173
24. Harrou F, Nounou MN, Nounou HN, Madakaryu M (2015) PLS-based ewma fault detection strategy for process monitoring. J Loss Prev Process Ind 36:108–119
25. Ozolek JA, Tosun AB, Wang W, Chen C, Kolouri S, Basu S, Huang H, Rohde GK (2014) Accurate diagnosis of thyroid follicular lesions from nuclear morphology using supervised learning. Med Image Anal 18(5):772–780
26. Rubin Y, Tomasi C, Guibas LJ (2000) The earth mover’s distance as a metric for image retrieval. Int J Comput Vis 40(2):99–121
27. Arifin BMS, Li Z, Shah SL (2018) Change point detection using the Kantorovich distance algorithm. IFAC Pap Online 51(18):708–713
28. Bathelt A, Ricker L, Jelali M (2015) Revision of the Tennessee Eastman process model. IFAC-Pap OnLine 48(8):309–314
29. Yin S, Ding SX, 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. J Process Control 22:1567–1581
30. Du B, Kong X, Feng X (2020) Generalized principal component analysis-based subspace decomposition of fault deviations and its application to fault reconstruction. IEEE Access 8:34177–34186
31. Madakaryu M, Harrou F, Sun Y (2019) Monitoring distillation column systems using improved nonlinear partial least squares-based strategies. IEEE Sens J 19(23):11697–11705