Nonlinear Partial Least Squares with Hellinger Distance for Nonlinear Process Monitoring

[1]Fouzi Harrou, *Member, IEEE*, [2]Muddu Madakyaru

[1]King Abdullah University of Science and Technology (KAUST)
Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division
Thuwal, 23955-6900, Saudi Arabia
fouzi.harrou@kaust.edu.sa

[1]Ying Sun

[2]Department of Chemical Engineering
Manipal Institute of Technology
Manipal University, India
muddu.m@manipal.edu

Abstract—This paper proposes an efficient data-based anomaly detection method that can be used for monitoring nonlinear processes. The proposed method merges advantages of nonlinear projection to latent structures (NLPLS) modeling and those of Hellinger distance (HD) metric to identify abnormal changes in highly correlated multivariate data. Specifically, the HD is used to quantify the dissimilarity between current NLPLS-based residual and reference probability distributions. The performances of the developed anomaly detection using NLPLS-based HD technique is illustrated using simulated plug flow reactor data.

I. INTRODUCTION

Process monitoring is employed by various process industries for improving the quality of products and enhancing process safety [1]–[3]. Among the most important applications of process safety are those related to environmental and industrial processes. To improve the reliability, safety and efficiency of advanced supervision methods, anomaly detection and diagnosis have become more and more important for numerous technical processes. Indeed, monitoring chemical and petrochemical processes is a challenging task because of the complexity and sometimes the lack of understanding about these processes. In these cases, data-based monitoring techniques are more commonly used. Moreover, data-based techniques provide efficient tools for extracting useful feature for design of monitoring schemes based on empirical models derived from the available process data [4]–[7]. Such methods require a minimal a prior knowledge about process physics, but depends on the availability of quality input data. A detailed overview of data-based monitoring techniques is presented in [5], [8]. Principal component analysis (PCA) and projection to latent structures (PLS) are two basic methods of multivariate analysis and reputed as powerful tools for monitoring multivariate processes with highly correlated process data [9], [10]. Unlike PCA, PLS finds an optimum pair of latent variables both in predictor (input) and predicted (output) variables such that these transformed variables have the largest covariance [11], [12]. Despite theirs simplicity, these linear LVR methods are based on assumption that process variables are linear and may not be appropriate for nonlinear process. However, most engineering and chemical processes are nonlinear. For highly nonlinear data the limitations of such linear methods are obvious. The use of linear LVR models (PLS and PCA) to approximate the nonlinearity may not be adequate; therefore in this situation the use of a nonlinear model is required to bypass such deficiency.

For nonlinear input-output systems, NLPLS has been extensively used to extract the linear and nonlinear relations between the process variables. Numerous variants of nonlinear PLS were developed for process monitoring, including the quadratic PLS [13]–[15], spline PLS [14], neural network PLS [16], [17], kernel PLS [11], [18] and fuzzy PLS [19], [20]. Usually, for anomaly detection purposes, NLPLS is first performed on the normal operating data (fault-free data) enabling us to obtain a reference model that will be used next for anomaly detection. Abnormal events are detected if the measurements deviate from the region of normal operation in latent space or in the residual space. The early and accurate detection of incipient anomalies represent an important step ahead that helps to design safe and reliable systems. However, conventional NLPLS-based monitoring indices such as T^2 and Q charts lose the ability to detect small changes in the mean of process data. Alternatively, one may use a NLPLS model with a distributional divergence-based monitoring metric, the Hellinger Distance (HD) to remedy the drawbacks described above and enhances the sensitivity for incipient change detection.

In this paper, we propose a statistical strategy that combines the flexibility of a NLPLS modeling approach, the greater sensitivity of HD metric to quantify dissimilarity between two distributions, and that may be efficiently enhance the sensitivity for incipient change detection. In such framework, NLPLS model is used to capture nonlinear relationships in multivariate input-output data to build predictive models. Specifically, in such NLPLS model, an optimal polynomial function in each inner relation via cross validation method is used to describe the inner relation of PLS model [21]. The HD-based detection index is used to identify deviation between actual NLPLS-based residuals and reference probability distributions. The HD metric has been used because it is very useful to quantify the similarity between two probability distributions. We apply the developed strategy to monitor a plug flow reactor. Obtained

results show the ability of the developed NLPS-HD strategy in detecting incipient anomalies, pointing out the promising application of these statistical tool in the supervision of nonlinear multivariate processes.

The rest of the paper is structured as follows. In Section II, a brief description of linear PLS and detailed discussion on NLPLS and fault detection methods are discussed. In Section III, some backgrounds of the Hellinger distance and its use for anomaly detection purpose, are briefly described. Next, in Section IV, we assess the proposed scheme and present some simulation results. Finally, some conclusions are given in Section V.

II. PRELIMINARIES

The developed method utilizes NLPLS as a modeling framework and uses HD-based index for fault detection. However, since NLPLS is utilized in this developed anomaly detection method, a brief introduction to NLPLS and how it can be used in anomaly detection, will be presented next. Consider a pair of datasets $\mathbf{X} \in \mathbb{R}^{N \times M}$ and $\mathbf{Y} \in \mathbb{R}^{N \times 1}$, where $\mathbf{X}$, $\mathbf{Y}$ are the input and output variables, respectively. After the data standardization by first subtracting the sample mean of the training data and then dividing by the sample standard deviation of the training data, PLS projects $\mathbf{X}$ and $\mathbf{Y}$ on to a lower dimension subspace defined by number of latent variable $[z_1, z_2, \ldots, z_l]$ as follows:

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

Where $\mathbf{Z} \in \mathbb{R}^{N \times l}$ (l is the number of latent variable) is the score matrix represents the projection of the variables on the subspace, $\mathbf{P} \in \mathbb{R}^{M \times l}$ represents the loading matrix for $\mathbf{X}$ and $\mathbf{Q} \in \mathbb{R}^{1 \times l}$ represents the loading matrix for $\mathbf{Y}$. $\mathbf{E}$ and $\mathbf{F}$ represents the model residue of input and output respectively. PLS calculates the input loading vectors, $\mathbf{P}_i$, so that the covariance between the estimated latent variable $\hat{\mathbf{Z}}_i$ and model output, $\mathbf{Y}$, i.e., [22]:

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

can be maximized with constraint

$$s.t. \quad \mathbf{P}_i^T \ \mathbf{P}_i = 1; \quad \mathbf{Z}_i = \mathbf{X}\mathbf{P}_i$$

where, $i = 1, ..., l$, $l \leq m$.

Note that PLS utilizes an iterative algorithm [22] to estimate the latent variables, where one latent variable or principal component is added iteratively to the model. After the inclusion of a latent variable, the input and output residuals are computed and the process is repeated using the residual data until a cross validation error criterion is minimized [23].

A. Nonlinear PLS model

The linear PLS has been extensively used as data analysis tool because of its well developed statistical foundation in the literature. Though it is used but it is limited to modeling linear relationship and most of the plant data are non linear in nature. Therefore, there is need of modeling of nonlinear relationship along with preserving the properties of linear PLS. Developing an efficient nonlinear transformation for high dimensional and collinear data set is very challenging and resulting models may suffer from overparameterization, which results in poor prediction. In the proposed nonlinear PLS modeling, outer transforamtion of linear PLS retained to get the robust properties of linear PLS. The inner relation of the linear PLS is modeled as optimal polynomial function. The advantage of doing is that dimension reduction method to eliminate collinearity among the data is achieved by the outer transformation of the linear PLS and nonlinear relationship is captured in the latent variable space. In the following subsection the proposed NLPLS is discussed.

The latent variables ($\mathbf{Z}\mathbf{c}_k$ obtained by the outer transformation of the linear PLS is used map the output variable $\mathbf{Y}$ using optimal polynomial function. which is described as,

$$\mathbf{Y}_k = f(\mathbf{Z}\mathbf{c}_k) + \mathbf{R}_k. \quad (3)$$

where $\mathbf{R}_k$ represents the model error. The parameters of the optimal polynomial function are selected by minimizing $\mathbf{R}_k$. The major advantage of the proposed NLPLS is that the outer layer of the linear PLS decomposition properties retained so that variables are projected into the directions maximizing the covariance. The proposed NLPLS algorithm can be formulated as follows:

1) Given the raw input $\mathbf{X}$ output $\mathbf{Y}$ data in to zero mean and unit variance.
2) At k=1, $\mathbf{E}_{\mathbf{X}_k}=\mathbf{X}$, $\mathbf{E}_{\mathbf{Y}_k}=\mathbf{Y}$
3) Determine the loading vector $\mathbf{P}_k$ and score vector (latent variable) $\mathbf{z}_k$).
4) Continue the step 2) till it converges to the required threshold. This step uses the nonlinear iterative least square (NIPALS) algorithm [24].
5) Determine the optimal polynomial structure for inner relation $f(.)$, which predicts the output $\mathbf{E}_{\mathbf{y}_k}$ with the input score vector (latent variable) $'Z'_k$ has the

$$E_{Y_k} = \xi_0 + \xi_1 Z c_k + \xi_2 Z c_k^2 + \xi_3 Z c_k^3 + \cdots + \mathbf{R}_k \quad (4)$$

The model is estimated by minimizing the regression vector $\mathbf{R}_k$ such that it does parsimonious in parameters. This part is computed as follows:

(i) Select the data (i.e., latent variable $\mathbf{Z}\mathbf{c}_k$ and output $\mathbf{E}_{\mathbf{Y}_k}$

(ii) Choose a order of the polynomial model structure

(iii) Split the univariate data into Training and Testing data

(iv) Estimate the parameters of the polynomial model

(v) Determine the mean square error using testing data set

(vi) Return to step (ii) with higher order model structure

(vii) Repeat the procedure (iii) to (v).

(viii) Decide the model structure based on the minimum MSE criteria.

6) Determine the input and output residuals i.e., $\mathbf{E}_{X_k} = \mathbf{X} - \hat{\mathbf{X}}$; $\mathbf{E}_{Y_k} = \mathbf{Y} - \hat{\mathbf{Y}}$

7) Let k=k+1, then return to 3) until all score vectors (latent variables) are calculated. The number of latent variables are optimized and compared using cross validation, by minimizing the output prediction mean squared error (MSE) using unseen data set.

III. HELLINGER DISTANCE-BASED MONITORING SCHEME

The proposed monitoring strategy combines modeling using NLPLS modeling with the Hellinger distance based monitoring index. HD metric, which is one of the most common measures of distance between probability distributions, is central to the problems of inference and discrimination [25]–[30]. Specifically, HD metric, which was initially developed by Ernest Hellinger, is an important statistical measure that can be used to quantify the dissimilarity or closeness between two probability density functions (PDFs) [26], [27], [30], [31]. Here, we address the problem of anomaly detection following a HD-based approach.

Definition 1 (Hellinger distance): Let us consider $P_1(x)$ and $P_2(x)$ to be two probability distributions with probability density functions $p_1(x)$ and $p_2(x)$, respectively. The HD of $p_1(x)$ relative to $p_2(x)$, which is a measure of the dissimilarity between $p_1(x)$ and $p_2(x)$, is defined by:

$$HD^2(p_1(x), p_2(x)) = \frac{1}{2} \sum (\sqrt{p_1(x)} - \sqrt{p_2(x)})^2. \quad (5)$$

The HD measure is a symmetric distance or metric in the Euclidean sense (i.e. $HD^2(p_1, p_2) = HD^2(p_2, p_1)$). It is non-negative and bounded (i.e., the HD satisfies $0 \leq HD^2(p_1, p_2) \geq 1$) and null only when the two densities are equal, $p_1 = p_2$.

The closed form expression of HD can easily be computed in the case of normal distributions. For univariate normal distributions, $p_1(x)$ and $p_2(x)$ of a random variable x, where $p_1 \sim \mathcal{N}(\mu_0, \sigma_0)$ and $p_2 \sim \mathcal{N}(\mu_1, \sigma_1)$, where μ_0 and μ_1 are the means and σ_0^2, σ_1^2 are the variances for p_1 and p_2, the HD between p_1 and p_2 is given by:

$$HD^2(p_1, p_2) = 1 - \sqrt{\frac{2\sigma_0\sigma_1}{\sigma_0^2 + \sigma_0^2}} \exp\left(-\frac{1}{4}\frac{(\mu_0 - \mu_1)^2}{\sigma_0^2 + \sigma_1^2}\right). \quad (6)$$

Note that the HD measure in the case of non-Gaussian variables requires that non-Gaussian distributions to be used instead. In some cases, the HD measure has no closed form and the function given by Equation (5) should be numerically approximated [32].

The HD metric is a useful tool for measuring the dissimilarity between two distributions, it can be a suitable statistic for anomaly detection. More specifically, the HD can be used as a anomaly indicator by comparing the statistical similarity between residual distributions before and after an anomaly. Therefore, it seems meaningful to adopt a NLPLS-based HD scheme for statistical process monitoring instead. Subsequently, the HD metric will be integrated with NLPLS to extend its anomaly detection abilities for detecting incipient anomalies.

A. NLPLS-based HD monitoring chart

In this subsection, a brief description of the NLPS-HD monitoring chart is introduced. The difference between the observed value of the output variable, y, and the predicted value, $\hat{y}$, obtained from NLPLS model represent the residual of the output variable, $\mathbf{F} = [f_1, \ldots, f_t, \ldots, f_n]$ which can be used as an indicator to detect a possible anomaly. The residual $\mathbf{F}$ obtained from PLS model is assumed to be Gaussian. It is assumed that the anomaly affect the mean parameter of residual distributions and the variance is supposed unchanged after the anomaly occurrence. In this approach, the HD has been used to measure the distance between the probability distribution of current residual $p_0(f) \sim \mathcal{N}(\mu_0, \sigma_0^2)$ against a reference one $p_1(f) \sim \mathcal{N}(\mu_1, \sigma_0^2)$, where μ_0 and μ_1 are the means and $\sigma_0^2 > 0$ is the variance for $p_0(f)$ and $p_1(f)$. Then, the HD distance based on the residuals distributions of the response variables from the NLPLS model can be computed as follows: $HD^2(p_0(f), p_1(f)) = 1 - \exp(-\frac{1}{8}\frac{(\mu_0 - \mu_1)^2}{\sigma_0^2})$. Where μ_0 and σ_0 are the mean and the standard deviation of NLPLS-based residuals obtained with fault-free data. The normal operating conditions is guaranteed by a zero HD when both PDFs are equal. Although, in real situation should deviate from zero only due to measurement errors or modeling uncertainties. The HD-based test makes decision between the null hypothesis $\mathcal{H}_0$ (absence of anomalies) and alternative hypothesis $\mathcal{H}_1$ (presence of anomalies) by comparing between the decision statistic $HD^2(p_0, p_1)$ and a given value of the threshold h.

$$HD^2(p_0, p_1) = 1 - \exp\left(-\frac{1}{8}\frac{(\mu_0 - \mu_1)^2}{\sigma_0^2}\right) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} h. \quad (7)$$

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

$$h = \mu_0^{HD} + L\sigma_0^{HD} \quad (8)$$

where μ_0^{HD} and σ_0^{HD} are the mean and the standard deviation for the nominal behavior of the anomaly indicator HD, and L is the width of the control limits which determines the confidence limits, usually specified in practice as 3 for a false alarm rate of 0.27%. If the decision function $HD^2(p_0, p_1)$ is larger than the threshold h, the HD-based test decides for $\mathcal{H}_1$, otherwise $\mathcal{H}_0$ is assumed to be true.

IV. RESULTS

In this example, the ability of the NLPLS-HD anomaly detection method is tested on simulated plug flow reactor data. The dynamic model of the PFR in which two first order reaction take place as per the equation (9). The process is modelled by a system of first order PDE's shown in

equation from 10 to 13. For the simulation purpose, the model parameters are used as listed in Table I [33].

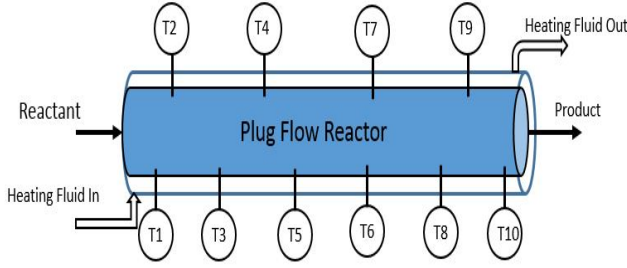$$A \rightarrow B \rightarrow C. \qquad (9)$$

Fig. 1.   Plug Flow Reactor: Schematic diagram of plug flow reactor.

The above reaction is series in nature, where A is the reactant species, which converts in to desired product B. Which is further decomposed into undesired product C. The reaction system are endothermic, therefore fluid in the jacket of the reactor is circulated to maintain the reactor temperature at desired level. The flow rate of the jacket fluid (heating fluid) is manipulated to get the desired product concentration i.e., B. There are ten temperatures measuring device used to measure the temperature at different location of the reactor as shown in Figure 1.

$$\frac{\partial C_A}{\partial t} = -v_l \frac{\partial C_A}{\partial x} - k_{10} e^{-E_1/RT_r} C_A \qquad (10)$$

$$\frac{\partial C_B}{\partial t} = -v_l \frac{\partial C_B}{\partial x} + k_{10} e^{-E_1/RT_r} C_A - k_{20} e^{-E_2/RT_r} C_B \quad (11)$$

$$\frac{\partial T_r}{\partial t} = -v_l \frac{\partial T_r}{\partial x} + \frac{\Delta H_{r_1}}{\rho_m c_{pm}} k_{10} e^{-E_1/RT_r} C_A$$
$$+ \frac{\Delta H_{r_2}}{\rho_m c_{pm}} k_{20} e^{-E_2/RT_r} C_B + \frac{U_w}{\rho_m c_{pm} V_r}(T_j - T_r) \quad (12)$$

$$\frac{\partial T_j}{\partial t} = -u \frac{\partial T_j}{\partial x} + \frac{U_{w_j}}{\rho_{mj} c_{pmj} V_j}(T_r - T_j) \qquad (13)$$

For the simulation purpose the reactor length converted to normalized length of $x \epsilon \ [0 \quad 1]$. The inlet concentration of the reactant and inlet temperature of the feed are defined as C_{A0} and T_{A0}. The flow rate of heating fluid in the jacket is u, which is used to regulate the temperature of the reacting mixture. The concentrations of reactant A and B are C_A and C_B in the reactor, the temperature inside the reactor is T and in the jacket Tj. The enthalpy of the two reactions as mentioned in the equation (9) are ΔH_{r1} and ΔH_{r2}. The density of fluid inside the reactor and heating fluid in the jacket are ρ_m and ρ_{mj} respectively. The heat capacity of the fluid in the reactor and in the jacket are c_{pm} and c_{pmj} respectively. The size of the reactor and jacket interns are of volume V_r and V_j, respectively. The overall heater transfer coefficient of reactor and jacket side are U_w and U_{wj}, respectively. The Arrhenius constants and the activation energy of the reactions (9) are k_{l0}, k_{20}, E_l, E_2.

1) Data generation: The training and testing data are generated by perturbing the input (i.e., flow rate of the reactant feed) around the steady state nominal condition mentioned in the Table I. The flow rate is perturbed using pseudo random binary signals (PRBS)signal of the frequency range [0 0.05 w_N], where $w_N = \pi/T$ represents the Nyquist frequency. In the modeling problem, Ten temperatures at different location in the reactor and feed flow considered as inputs and product concentration considered as output (i.e., C_B). The data generated are noise free, which are then added with zero mean gaussian noise. The training and testing test data are shown in Figure 2.

TABLE I
PLUG FLOW REACTOR: MODEL PARAMETERS.

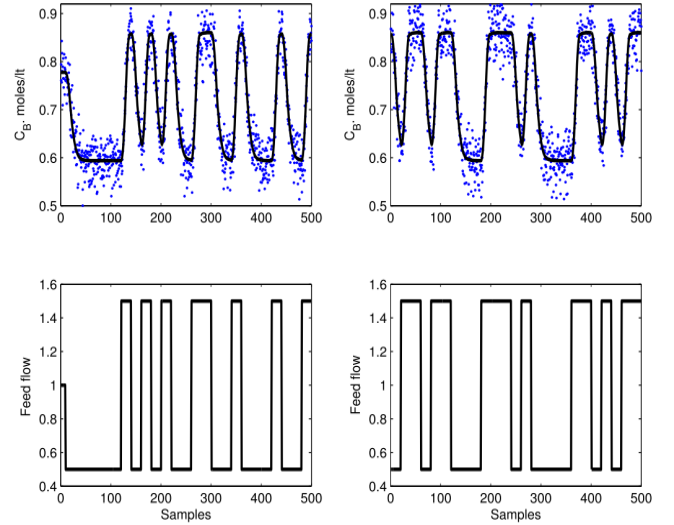| Process Variable	Value	Process Variable	Value
v_l	1 ml/min	C_{pm}	0.231 kcal/(kg K)
L	1.0 m	R	1.987 kcal/(min K)
V_r	10.0 lt	ρ_m	0.09 kg/lt
E_1	20000 kcal/kmol	U_w	0.20 kcal/(min K)
E_1	50000 kcal/kmol	C_{pmj}	0.8 kcal/(min K)
k_{10}	5.0×10^{12} min^{-1}	V_j	366 K
k_{20}	5.0×10^{2} min^{-1}	ρ_{mj}	0.10 kg/lt
H_{r_1}	0.5480 kcal/kmol	C_{A0}	4 mol/lt
H_{r_2}	0.9860 kcal/kmol	C_{B0}	0 mol/lt
T_{r0}	320 K	T_{j0}	375 K

Fig. 2.   Plug flow reactor: Dynamic input-output data used for training and testing the model with noise SNR=10 (solid black line: noise-free data; blue dots: noisy data).

A NLPLS model is fitted to the fault-free training dataset, and the goodness of fit is shown in Figure 3. It shows that 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. Therefore, the models were successful in accounting for most of the significant autocorrelations present in the data, and there

is no indication of a curvature or other anomalies. According to Figure 3, it can be seen that the scatter plots of observed and predicted data indicate a reasonable performance of the selected models.
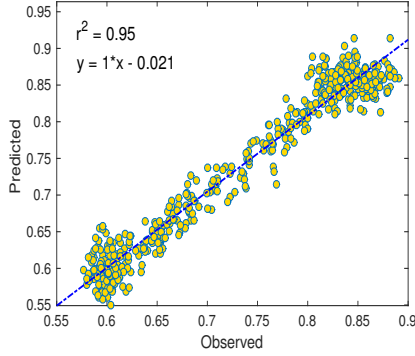

Fig. 3. Plot of observed and predicted training data via NLPLS model for the case where SNR=10.

Before applying the NLPLS-HD scheme for anomaly detection purpose, we need to check whether the residuals of response variables follows Gaussian distributions to ensure that the data are well represented using a NLPLS model. The Q-Q plot of residuals in Figure 4 evaluate the normality of the output residuals. The left panel of Figure 4 shows that the output residuals obtained from NLPLS is normal. The histogram on the right panel of Figure 4 confirms that the normality assumption appears to be a reasonable one.
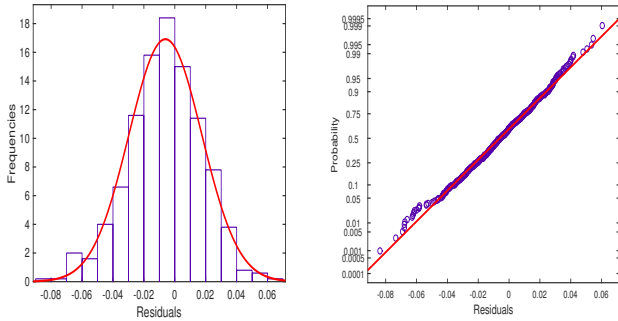

Fig. 4. Tests for Normality of Residuals (Left panel): Histograms showing the normality of the residuals. In each plot, histogram of residuals obtained from the MSPLS model have approximately normal distribution (solid curve). (Right panel): Normal Q-Q plot of residual.

A. Detection results

In this case study, the detection of an abrupt anomaly in temperature sensor is investigated. An abrupt anomaly in the temperature measurement $T5$ with a magnitude of 10% is introduced at the 300th sample and removed at the 350th sample of the testing data. The result of NLPLS-HD monitoring scheme for the considered bias anomaly, which are displayed in Figure 5, clearly indicate that the proposed strategy can successfully detect this anomaly.
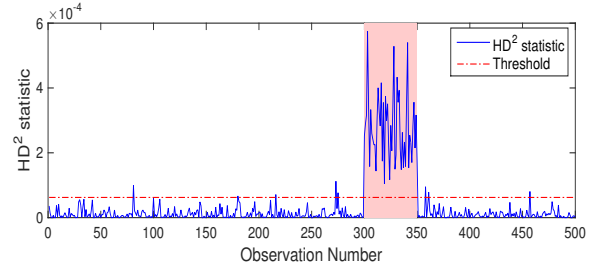

Fig. 5. The time evolution of the HD decision statistic in the presence of a bias anomaly in the temperature sensor measurements $T5$.

The aim of the second case study was to assess the potential of the proposed NLPLS-based HD method to detect intermittent anomalies. Towards this end, a small bias level, which is 5% of the total variation in $T5$, is injected between samples between intervals [200, 250] and a bias of 10% is introduced between sample interval [400,450]. The results of the NLPLS-HD scheme, which are illustrated in Figure 6, clearly show the capability of this proposed method to detect this anomaly with small number of false alarms.
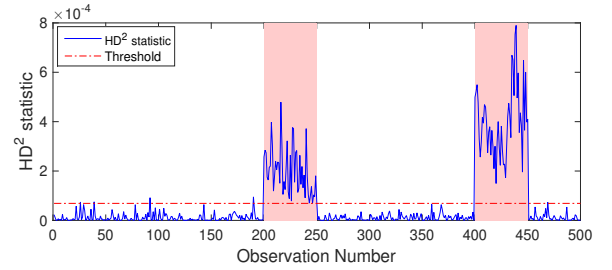

Fig. 6. The time evolution of the HD decision statistic in the presence of intermittent anomalies in $T5$.

The aim of this case study is to assess the potential of the proposed NLPLS-HD anomaly detection scheme to detect incipient or gradual anomalies. To do so, a slow increase in the input variable $T5$ with a slope of 0.01 was added to the test data starting at sample number 300 of the simulated testing data. The NLPLS-HD statistic gradually increased as the anomaly slowly developed and began to violate the threshold value when the size of the anomaly became sufficiently large that it was detected by this model. From the plot, the NLPLS-HD scheme gives a signal of anomaly at sample number 309th.
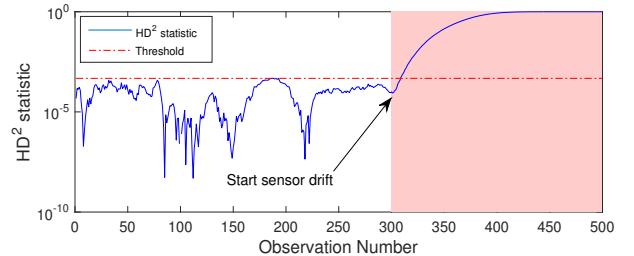

Fig. 7. The time evolution of the HD decision statistic on a semi-logarithmic scale in the presence of a drift sensor anomaly in $T5$.

V. CONCLUSION

Motivated by the sensitivity of HD-based detection to small changes in the process mean, we developed an efficient chart for detecting incipient anomalies in nonlinear multivariate processes. The HD-based anomaly detection using NLPLS is conceptually more straightforward and more sensitive for the detection of incipient anomalies. The effectiveness of the NLPLS-HD scheme was demonstrated with data from a simulated plug-flow reactor. Satisfactory detection results were obtained using the proposed method.

This work further can be extended to handle nonlinear processes with uncertainty in the measurements. To do that, we plan to develop anomaly detection approach based on multiscale NLPLS-based HD metric for nonlinear process monitoring.

ACKNOWLEDGEMENT

REFERENCES

[1] S. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools.* Springer Science & Business Media, 2008.

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

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

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

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

[6] F. Harrou, F. Kadri, S. Khadraoui, and Y. Sun, "Ozone measurements monitoring using data-based approach," *Process Safety and Environmental Protection*, vol. 100, pp. 220–231, 2016.

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

[8] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Industrial & Engineering Chemistry Research*, vol. 52, no. 10, pp. 3543–3562, 2013.

[9] T. Kourti, "Process analysis and abnormal situation detection: from theory to practice," *IEEE control systems*, vol. 22, no. 5, pp. 10–25, 2002.

[10] F. Harrou, M. Madakyaru, Y. Sun, and S. Khadraoui, "Improved detection of incipient anomalies via multivariate memory monitoring charts: Application to an air flow heating system," *Applied Thermal Engineering, DOI: 10.1016/j.applthermaleng.2016.08.047*, 2016.

[11] R. Rosipal and L. Trejo, "Kernel partial least squares regression in reproducing kernel hilbert space," *Journal of machine learning research*, vol. 2, no. Dec, pp. 97–123, 2001.

[12] F. Harrou, M. Nounou, H. Nounou, and M. Madakyaru, "PLS-based EWMA fault detection strategy for process monitoring," *Journal of Loss Prevention in the Process Industries*, vol. 36, pp. 108–119, 2015.

[13] S. Wold, N. K. Wold, and B. Skagerberg, "Nonliner pls modeling," *Chemometrics and Intelligent laboratory Systems*, vol. 7, pp. 53–65, 1989.

[14] S. Wold, "Nonlinear partial least squares modelling II. Spline inner relation," *Chemometrics and Intelligent Laboratory Systems*, vol. 14, no. 1-3, pp. 71–84, 1992.

[15] I. frank, "A nonliner pls model," *Chem. Int. Lab. Sys.*, vol. 8, pp. 109–119, 1990.

[16] Z. Tian, B. Gu, L. Yang, and Y. Lu, "Hybrid ANN–PLS approach to scroll compressor thermodynamic performance prediction," *Applied Thermal Engineering*, vol. 77, pp. 113–120, 2015.

[17] D. Lee, M. Lee, S. Woo, Y. Kim, and J. Park, "Nonlinear dynamic partial least squares modeling of a full scale biological wastewater treatment plant." *Process Biochemistry*, vol. 41, pp. 2050–2057, 1992.

[18] F. Chu, F. Wang, X. Wang, and S. Zhang, "Performance modeling of centrifugal compressor using kernel partial least squares," *Applied Thermal Engineering*, vol. 44, pp. 90–99, 2012.

[19] I. Araby, Abdel-Rahman, and G. J. Lim, "A nonliner partial least squares algorithm using quadratic fuzzy system," *J. of Chemometrics*, vol. 23, pp. 530–537, 2009.

[20] Y. H. Bang, C. K. Yoo, and I.-B. Lee, "Nonliner pls modeling with fuzzy inference system," *Chemometrics and Intelligent laboratory Systems*, vol. 64, no. 2, pp. 137–155, 2003.

[21] M. Madakyaru, M. Nounou, and H. Nounou, "Nonlinear latent variable regression," in *IEEE Symposium on Computational Intelligence in Control and Automation (CICA).* IEEE, 2013, pp. 81–88.

[22] Y. Hiroyuki, Y. B. Hideki, F. C. E. O. Hiromu, and F. Hideki, "Canonical correlation analysis for multivariate regression and its application to metabolic fingerprinting," *Biochemical Engineering Journal*, vol. 40, pp. 199–204, 2008.

[23] M. Stone and R. Brooks, "Continuum regression: Cross validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal component regression," *Journal of the Royal Statistical Society, Ser. B*, vol. 52, pp. 237–269, 1990.

[24] S. J. Qin and T. J. McAvoy, "Nonlinear pls modeling using neural networks." *Comput. Chem. Eng.*, vol. 16, no. 4, pp. 379–391, 1992.

[25] M. Basseville, "Divergence measures for statistical data processing–An annotated bibliography," *Signal Processing*, vol. 93, no. 4, pp. 621–633, 2013.

[26] G. Ditzler and R. Polikar, "Hellinger distance based drift detection for nonstationary environments," in *IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environments (CIDUE).* IEEE, 2011, pp. 41–48.

[27] C. Li, B. Huang, and F. Qian, "Hellinger distance based probability distribution approach to performance monitoring of nonlinear control systems," *Chinese Journal of Chemical Engineering*, vol. 23, no. 12, pp. 1945–1950, 2015.

[28] K. Yamanishi, J.-I. Takeuchi, G. Williams, and P. Milne, "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms," *Data Mining and Knowledge Discovery*, vol. 8, no. 3, pp. 275–300, 2004.

[29] A. Korostelev and A. Tsybakov, *Minimax theory of image reconstruction.* Springer Science & Business Media, 2012, vol. 82.

[30] J. Tajer, A. Makke, O. Salem, and A. Mehaoua, "A comparison between divergence measures for network anomaly detection," in *7th International Conference on Network and Service Management.* IEEE, 2011, pp. 1–5.

[31] I. Csiszár and P. Shields, *Information theory and statistics: A tutorial.* Now Publishers Inc, 2004.

[32] D. Romano and M. Kinnaert, "Robust fault detection and isolation based on the kullback divergence," *IFAC Proceedings Volumes*, vol. 39, no. 13, pp. 426–431, 2006.

[33] H. Shang, J. F. Fraser, and M. Guay, "Characteristics-based model prdictive control of distributed parameter system," *Americal Control Conference PRoc., Anchorage, Alaska*, pp. 4383–4388, May 8-10, 2002.