

# Virtual sensors of nonlinear industrial processes based on neighborhood preserving regression model<sup>★</sup>

Chihang Wei, Zhihuan Song

*State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: chhwei@zju.edu.cn; songzhihuan@zju.edu.cn).*

**Abstract:** The micro perspective of manifold proximity would indicate local relationships with their unique spatial geometric distribution characteristics among the data samples, which are usually neglected by traditional data-driven virtual sensors. This would not guarantee a good prediction performance. In this paper, a regression model with localized construction named neighborhood preserving regression (NPR) model is proposed. It extends the unsupervised neighborhood preserving embedding (NPE) to the supervised form. The projection vector is learned from input process variables and the output quality variable, synchronously exploring the manifold structure of the input process variables for the dimension reduction and developing the regression relationship between the projected input process variables and the output quality variable. The model is developed as a novel designed optimization whose analytical solution would be compactly and directly calculated without any iterative procedures. The effectiveness of the proposed algorithm is demonstrated by case studies carried out on a simulated penicillin production process.

**Keywords:** Monitoring of product quality and control performance, data mining and multivariate statistics, artificial intelligence, data-driven regression methods, nonlinear processes, neighborhood preserving embedding, virtual sensor, manifold learning.

## 1. INTRODUCTION

In recent years, real-time monitoring and control of industrial processes has become increasingly important. For this purpose, hardware sensors have been widely instrumented to the production facilities so that the measurements of process variables can be easily collected and recorded. However, a kind of variable closely related to product quality, known as the quality variables, is hard to be obtained online due to technological limitations and economic factors (Kadlec *et al.*, 2009; Ge *et al.*, 2017, 2018; Yao *et al.*, 2018; Kadlec *et al.*, 2011). Examples of quality variables include oxygen concentration in a furnace, melt index of polypropylene, measured height of etching process in semiconductor industry, etc (Shao *et al.*, 2018; Wei *et al.*, 2019).

The technique of virtual sensor modeling is one of the potential solutions that can estimate the measurements of the quality variables online and indicate process issues timely from easy-to-measure secondary variables (i.e. process variables, like temperature, pressure, and flow rate). With the rapid developments of computer science, data analysis, artificial intelligence, and computing hardware, data-driven virtual sensors have attracted much attention among various virtual sensor techniques (Ge *et al.*, 2017;

Qin *et al.*, 2012, 2014; Liu *et al.*, 2013). A large amount of data can be conveniently collected via distributed control systems (DCS), which reflect the true process information.

In modern industries, the collected process data is usually characterized by high dimensionality, but only a few intrinsic variables independently dominate the data variations. As the relationship among dimensions is usually nonlinear, it is thus necessary to synchronously reduce the dimensions and to handle the nonlinearity while constructing the virtual sensor models. Typical approaches are neural network and kernel-based methods (Yan *et al.*, 2016; Yu *et al.*, 2012). However, nearly all the traditional approaches are derived from the global perspective of the data structure and pursue the outer shape of the data. They cannot provide any insight into the micro perspective of manifold proximity indicating the local relationships among the data samples. This should not be effective since each local neighborhood may have its unique spatial geometric distribution characteristics. NPE, a manifold learning algorithm, was proposed by He to resolve the issue through localized construction (He *et al.*, 2005). Unlike the techniques based on global-perspective, NPE aims at exploring the intrinsic local manifold topology structure and proximity relations among data samples. Ideally, NPE reveals the real manifold structure of the nonlinear data in high dimensions (He *et al.*, 2005). More importantly, the boundary of the spatial distribution of the samples in the input space is more likely to be preserved. Besides, NPE is naturally robust to outliers without using rigorous equality

<sup>★</sup> This work was supported by the Key Project of Natural Science Foundation of China (61933013) and Ningbo 13th five-year marine economic innovation and development demonstration project NBHY-2017-Z1.

constraints. Despite the favorable properties of NPE, directly using the low dimensional embedding learned from NPE may make a worse regression model with the quality variables. This is because the quality variables are not considered impartially while exploring the intrinsic local topology and proximity relations among process variables, and both the manifold structure information among quality variables and the information between process variables and quality variables are totally discarded.

In this paper, NPR model with localized construction is proposed as a regression model. It extends the unsupervised NPE to the supervised form. The model is developed by synchronously exploring the manifold structure of the input process variables for the dimension reduction and developing the regression relationship between the projected input process variables and the output quality variable. The optimization whose analytical solution would be compactly and directly calculated is novelly designed and the projection vector is learned from input process variables and output quality variable.

All of the above techniques will be further described in more detail in the remainder of this paper. In Section 2, the background knowledge of the NPE algorithm will be briefly reviewed. Section 3 introduces the proposed NPR model with detailed motivations and derivations of the optimization. In addition, procedures about how the proposed model is utilized for virtual sensors are presented. In Section 4, case studies on a simulated fed-batch penicillin process are presented to evaluate the performance of the proposed model, while the article ends with concluding remarks in Section 5.

## 2. PRELIMINARY: NPE

The basic background knowledge of NPE is revisited as a prelude in this section. NPE is proposed to explore the local linearity of topology structure of a manifold which locally resembles a Euclidean space. More precisely, the neighborhood of each sample of a D-dimensional manifold is homeomorphic to a D-dimensional Euclidean space, which is a much weaker condition than the common assumptions, such as Gaussian distribution, linearity, and independent identically distributed. In case that high dimensional data inherently lie in or near a low dimensional manifold, expressing the data in its original space is neither efficient nor necessary since Euclidean distances are meaningful only on the local scale. In this case, it is preferred to pursue a representation so that Euclidean distances would be globally meaningful.

NPE has been proposed as an unsupervised dimensionality reduction approach (He *et al.*, 2005). It aims at exploring the intrinsic local manifold topology structure and proximity relations among data samples. Ideally, NPE reveals the real manifold structure of the nonlinear data in high dimensions. More importantly, the boundary of the spatial distribution of the samples in the input space is more likely to be preserved. Based on the geometric intuition that the global nonlinearity can be approximated by the local linearity, NPE is conducted under the assumption that the reconstructing coefficients which represent the local manifold structure data samples within a neighborhood could be represented by the linear reconstruction relationships

and should be preserved in the low dimensional embedding space.

Suppose that the matrix of process variables in the original space is expressed as  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{N \times P}$ .  $N$  and  $P$  respectively represent the number of training samples and the dimension of process variables. The algorithm is designed to find a projection matrix  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_F] \in \mathbb{R}^{P \times F}$  to project the high dimensional data  $\mathbf{X}$  into the low dimensional representation  $\mathbf{Z} = \mathbf{X}\mathbf{A} = [\mathbf{z}_1, \dots, \mathbf{z}_N] \in \mathbb{R}^{N \times F}$ , where  $F < P$ . The three procedures of NPE are described in detail as follows:

- (1) Construct the adjacency matrix representing the neighborhood relationships. The grouping technique, such as k-nearest neighbors (k-NN) or choosing the neighbor samples within a sphere with the fixed radius ( $\epsilon$ -neighborhoods), can be utilized to construct the adjacency matrix.
- (2) Compute the reconstruction weight matrix. Suppose  $\mathbf{W}$  to be the reconstruction weight matrix, where its elements  $\mathbf{W}_{ij}$  is the weight from data sample  $\mathbf{x}_i$  to data sample  $\mathbf{x}_j$ , and  $K$  to be the number of the nearest neighbors. The matrix  $\mathbf{W}$  can be optimized by the following criteria that minimizes the squared reconstruction error between all the data samples and their corresponding reconstructions,

$$L(\mathbf{W}) = \min_{\mathbf{W}} \sum_{i=1}^N \left\| \mathbf{x}_i - \sum_{j=1}^K \mathbf{W}_{ij} \mathbf{x}_j \right\|^2 \quad (1)$$

It is noted that  $\mathbf{W}$  includes the information of the intrinsic local manifold topology structure extracted from the data in the original space.

- (3) Compute the projection matrix. Based on the geometric graphical intuition that the local neighborhood manifold topology structure of the original data should be optimally preserved in the low dimensional space, NPE reconstructs each data sample to be the low dimensional representation with the same reconstruction weight matrix, that is

$$L(\mathbf{Z}) = \min_{\mathbf{Z}} \sum_{i=1}^N \left\| \mathbf{z}_i - \sum_{j=1}^K \mathbf{W}_{ij} \mathbf{z}_j \right\|^2 \quad (2)$$

Eq.(2) can be reformulated in terms of  $\mathbf{a}$  as

$$\begin{aligned} \mathbf{a}_{opt} &= \arg \min_{\mathbf{a}} \left\| \sum_{i=1}^N \mathbf{a}^T \mathbf{x}_i - \sum_{j=1}^K \mathbf{W}_{ij} \mathbf{a}^T \mathbf{x}_j \right\|^2 \\ &= \arg \min_{\mathbf{a}} \mathbf{a}^T \mathbf{X}^T \mathbf{M} \mathbf{X} \mathbf{a} \end{aligned} \quad (3)$$

where

$$\mathbf{M} = (\mathbf{I}^{N \times N} - \mathbf{W}) (\mathbf{I}^{N \times N} - \mathbf{W})^T \quad (4)$$

$$\mathbf{I}^{N \times N} = \text{diag}[1, \dots, 1] \in \mathbb{R}^{N \times N} \quad (5)$$

It is easy to check that  $\mathbf{M}$  is symmetric and positive semi-definite.

It is known that NPE can bottom-up learn the manifold topology from the process data. More importantly, the boundary of the spatial distribution of the samples in the input space is more likely to be preserved. Besides, NPE is naturally robust to outliers without using rigorous equality constraints to preserve local manifold structure.

Despite the favorable properties of NPE, directly using the low dimensional embedding learned from NPE would make a worse regression model with the quality variables. This is because the quality variables are not considered impartially while exploring the intrinsic local topology and proximity relations among process variables, and both the manifold structure information among quality variables and the information between process variables and quality variables are totally discarded. This will not guarantee a good regression performance.

### 3. NEIGHBORHOOD PRESERVING REGRESSION MODEL BASED VIRTUAL SENSORS

Modern plants have become far more complex and data-driven methods have attracted more and more attentions (Qin *et al.*, 2012, 2014). As the collected process data from modern complex systems is usually characterized by high dimensionality and nonlinearity, it is thus desirable to simultaneously reduce the dimensions and to handle the nonlinearity while constructing the virtual sensor models.

Regression is the core of the virtual sensor technology. Motivated by the concept of NPE and the supervised regression modeling, this section proposes a regression model with localized construction named NPR for virtual sensor modeling. Unlike NPE, aiming at exploring the intrinsic local manifold topology structure among process variables for nonlinear dimension-reduction, NPR synchronously explores the intrinsic local manifold topology structure among the process variables and develops the regression relationship for the estimates of the quality variable.

Specifically, NPR model is designed and developed upon an optimization including one novelly constructed objective function subject to one constraint. The objective function is compounded by two parts: (1)  $\Gamma^R$  represents regression by maximizing the regression relationship between the projected process variables in the latent space and the quality variable; (2)  $\Gamma^{DR}$  is constructed for nonlinear dimension-reduction, projecting the process variables to the latent space to explore the local manifold topology structure and proximity relations among them. The constraint is implemented to normalize the projection vector. Suppose that the vector of the quality variable in the original space is expressed as  $\mathbf{y} = [y_1, \dots, y_N] \in \mathbb{R}^{N \times 1}$ .

The latent variable of  $\mathbf{X}$  is expressed as  $\mathbf{t} = \mathbf{X}\mathbf{w}$  where  $\mathbf{w} \in \mathbb{R}^{P \times 1}$  is coefficient weight. To develop the regression relationship between the projected input process variables and the output quality variable, the objective function  $\Gamma^R$  for regression maximizes the measuring the dot product,

$$\Gamma^R = \langle \mathbf{t}, \mathbf{y} \rangle = \langle \mathbf{X}\mathbf{w}, \mathbf{y} \rangle = \mathbf{w}^T \mathbf{X}^T \mathbf{y} \quad (6)$$

To extract local topologies among process variables, the construction of  $\Gamma^{DR}$  is based on the same concept in NPE, which is, the local neighborhood manifold topology structure of the original data should be optimally preserved in the low dimensional space. Similar to (3),  $\Gamma^{DR}$  is expressed as,

$$\Gamma^{DR} = \mathbf{w}^T \mathbf{X}^T \mathbf{M}_X \mathbf{X} \mathbf{w} \quad (7)$$

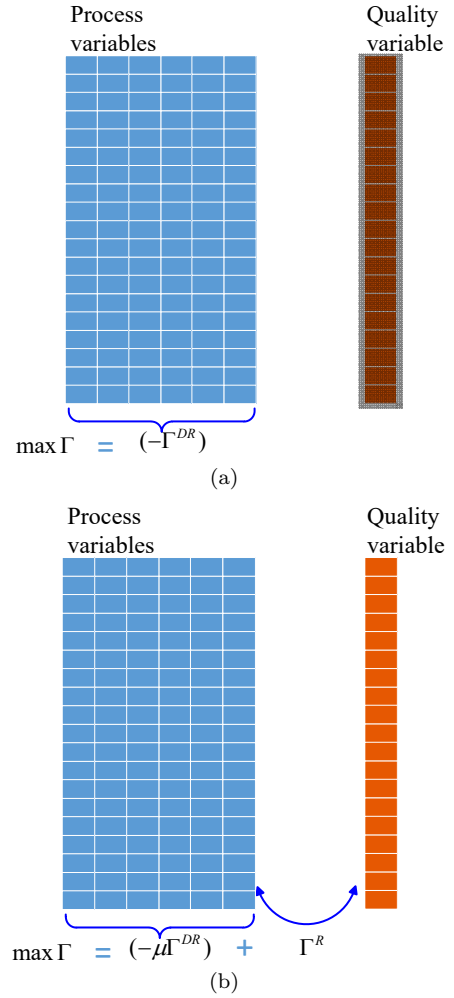


Fig. 1. Modeling sketch maps of (a) NPE, developed merely upon local topology among process variables; and (b) NPR, developed upon local topologies among process variables and regression relationship.

where  $\mathbf{M}_X$  is calculated through (1) and (4). One constraint is supplement to this optimization to normalize the projection vector of process variables,  $\mathbf{w}^T \mathbf{w} = 1$ .

The independent parts of objective functions can be easily combined together by taking the linear combination,

$$\Gamma = \Gamma^R - \mu \Gamma^{DR} = \mathbf{w}^T \mathbf{X}^T \mathbf{y} - \mu \mathbf{w}^T \mathbf{X}^T \mathbf{M}_X \mathbf{X} \mathbf{w} \quad (8)$$

where  $\mu$  is supplement to balance the order of magnitudes between  $\Gamma^R$  and  $\Gamma^{DR}$ . Eq.(8) simultaneously maximizes  $\Gamma^R$  and minimizes  $\Gamma^{DR}$ . To find the maxima of  $\Gamma$ , let  $\partial \Gamma / \partial \mathbf{w} = 0$ , it can be obtained that,

$$\mathbf{X}^T \mathbf{y} - 2\mu \mathbf{w}^T \mathbf{X}^T \mathbf{M}_X \mathbf{X} \mathbf{w} = 0 \quad (9)$$

Then the analytical solution would be compactly and directly calculated as

$$\mathbf{w} = [2\mu \mathbf{w}^T \mathbf{X}^T \mathbf{M}_X \mathbf{X} \mathbf{w}]^{-1} \mathbf{X}^T \mathbf{y} \quad (10)$$

The projection vector is then normalized to ensure the constraint  $\mathbf{w}^T \mathbf{w} = 1$ .

[Note] Parameter optimization of NPR

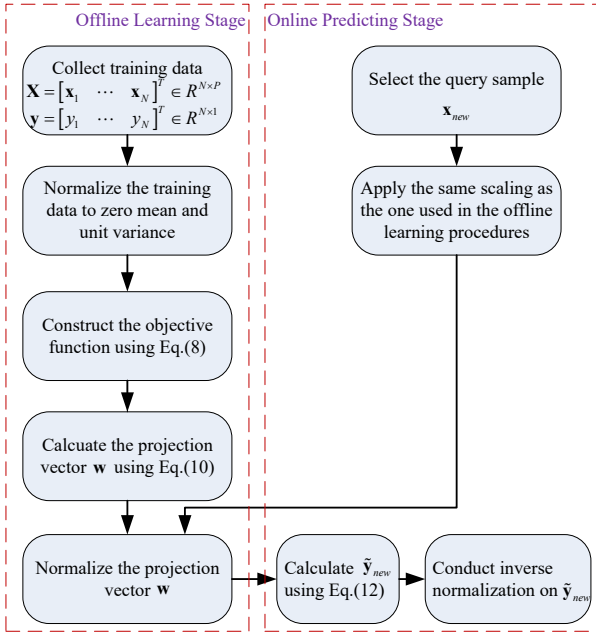


Fig. 2. Flowchart of the proposed NPR-based virtual sensors.

$K$  is the number of the nearest neighbors, which determines the size of neighborhoods and influences the linear reconstruction. If  $K$  is large, the neighborhoods would include much data information, but this may be conflicted with the local linear assumption. If  $K$  is small, there may not be sufficient information in the neighborhoods. In this paper, given  $\vartheta$  to be reconstruction errors (Kouropteva *et al.*, 2002), the value of  $K$  is determined by the following criterion,

$$\sum_{i=1}^N \left\| \mathbf{x}_i - \sum_{j=1}^K \mathbf{W}_{ij} \mathbf{x}_j \right\| \leq \vartheta \times N \times P \quad (11)$$

To illustrate the similarities and dissimilarities of modeling of NPE and NPR, Fig. 1 presents the corresponding modeling schematic diagrams. As shown in Fig. 1(a),  $\Gamma^{DR}$  in NPE ((7)) is developed merely upon process variables for extracting local topology among them, but it does not consider the quality variables (the minus sign is added for unification). In the proposed NPR shown in Fig. 1(b),  $\Gamma = \Gamma^R - \mu \Gamma^{DR}$  is modeled for discovering the intrinsic geometrical structure among both process variables and quality variables and the regression simultaneously.

To calculate the prediction  $\tilde{y}_{new}$  with a new sample of process variables, the regression system is developed and expressed as,

$$\tilde{y}_{new} = \mathbf{x}_{new} \times \mathbf{w} \quad (12)$$

The flowchart of the proposed NPR-based virtual sensors is listed in Fig. 2.

#### 4. CASE STUDIES

In this section, the effectiveness of the proposed NPR model is evaluated through case studies carried out on a simulated penicillin process. The data sets of the benchmark case are generated through the fed-batch penicillin production process. There are three quality variables. To compare

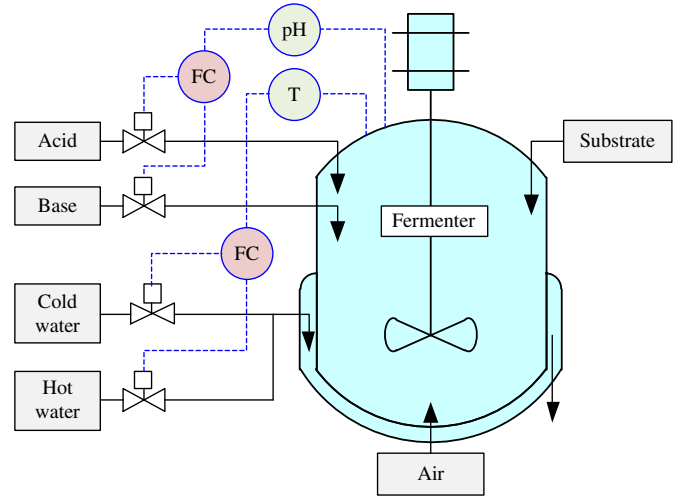


Fig. 3. Flow diagram of penicillin fermentation process.

the different regression performance, the proposed NPR in this paper is adopted. In addition, principal component regression (PCR), partial least squares (PLS), and kernel PLS (KPLS), as frequently used data-driven methods, are also adopted (Qin *et al.*, 2012, 2014). The software environment for both cases is MATLAB(R) R2016a with hardware environment Intel(R) Core(TM) i7-8700 CPU and 16-GB flash memory. The root-mean square error (RMSE) is employed to quantify the regression accuracy,

$$\text{RMSE} = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (\tilde{y}_i - y_i)^2} \quad (13)$$

where  $N_t$  represent the total number of the testing samples.  $y_i$  and  $\tilde{y}_i$  respectively represent the true value and the prediction of the quality variable. If there are multi-quality variables, one RMSE value will be calculated for each.

The penicillin production process is a fed-batch fermentation process, characterized by nonlinearity and multi-phase characteristics, which has been widely adopted for regression performance evaluation of virtual sensors (Shao *et al.*, 2018). The simulator for this penicillin production process under a variety of operating conditions, referred to as Pensim (copyright), is available at <http://simulator.iit.edu/web/pensim/simul.html>. Fig. 3 presents the flow diagram of this process. Two quality variables are considered, the penicillin concentration and biomass concentration; for virtual sensor modeling, 10 process variables are selected as the secondary variables, as listed in Table 1, referring to Shao *et al.* (2018). Two batches of data were generated with the Pensim tool under the default simulation conditions. The total simulation time was set to 400 hours and the sampling interval was set to 0.4 hour, thus, each batch contains 1000 samples. One batch is used for model construction and the other one is used for validation.

PCR, PLS, KPLS, and NPR, are respectively built for each quality variable based on the training data. The parameters of all the virtual sensor approaches are summarized as follows. For KPLS, the Gaussian kernel function issued with  $\tau = 100$  as suggested in (Lee *et al.*, 2004). For the

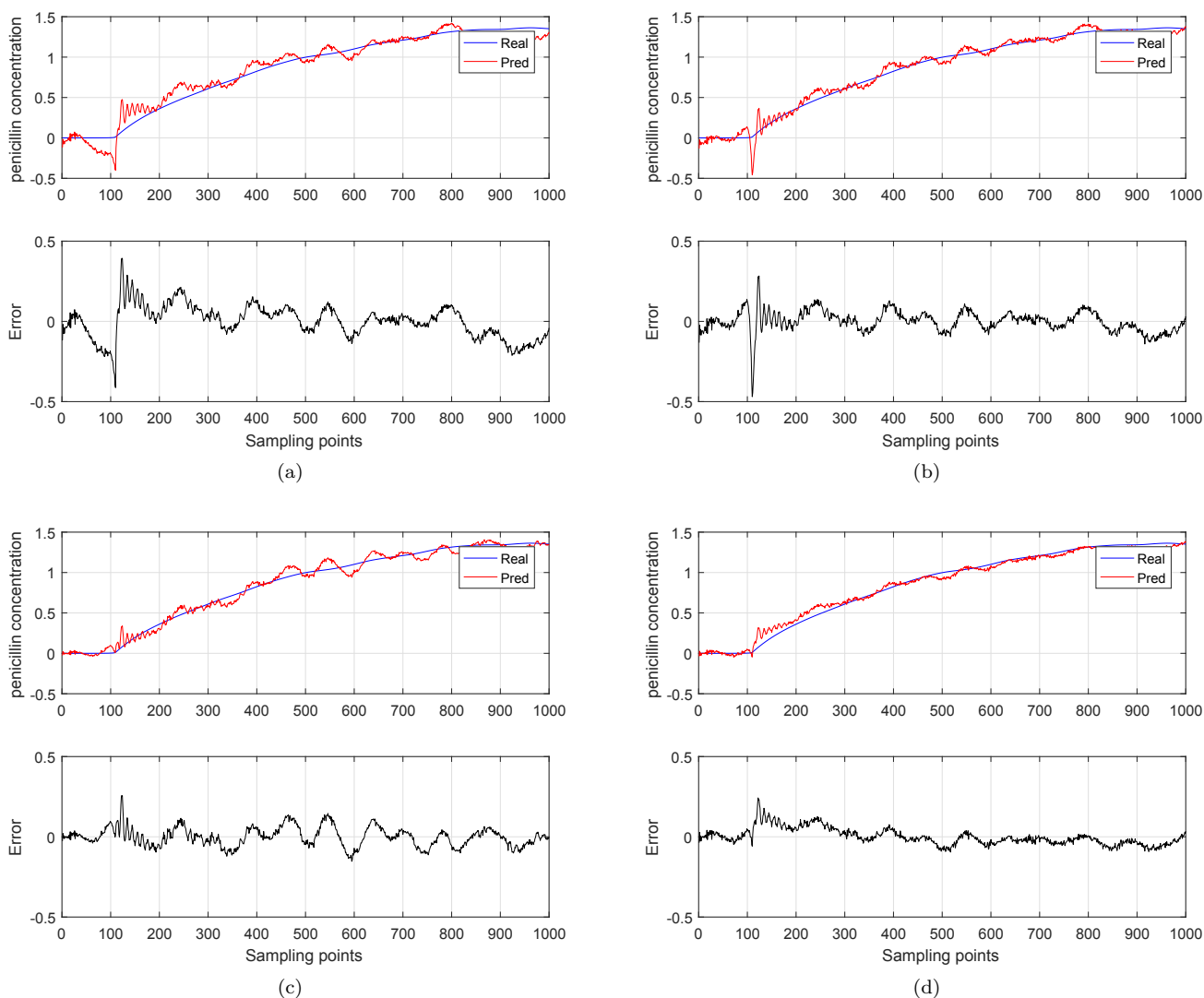


Fig. 4. Prediction result of the penicillin concentration of the penicillin fermentation process based on (a) PCR, (b) PLS, (c) KPLS, and (d) NPR.

Table 1. Secondary and quality variables of the penicillin process

No.	Secondary variables	No.	Quality variables
1	Aeration rate	1	Penicillin concentration
2	Agitator power	2	Biomass concentration
3	Substrate feed rate		
4	Substrate feed temperature		
5	Dissolved oxygen concentration		
6	Culture volume		
7	Carbon dioxide concentration		
8	pH		
9	Fermenter temperature		
10	Generated heat		

proposed NPR, the number of the nearest neighbors of process variables is set to be  $K = 5$  with  $\vartheta = 1/8$  suggested by (Kouropyteva *et al.*, 2002), while the combined objective function is conducted with  $\mu = 0.0005$  to balance the order of magnitudes between  $\Gamma^R$  and  $\Gamma^{DR}$ .

Table 2. RMSE results of penicillin process

Method	Penicillin	Biomass
PCR	0.1020	0.4068
PLS	0.0654	0.4269
KPLS	0.0600	0.4329
NPR	<b>0.0513</b>	<b>0.2612</b>

Table 2 gives the RMSE results, where the bolded numbers represent the best regression performance. The results show that the proposed NPR preforms the best with the highest accuracy. PCR and PLS have poor performances since both are limited by the linear assumption. KPLS is one traditional kernel method model. The prediction accuracy has been improved by introducing an artificially determined kernel function to deal with nonlinearity. However, the utilization of the artificially determined kernel function just focuses on how to map the training data onto the high dimensional kernel spaces, while the regression relationships between the process variables and the quality variables are totally ignored. What's more, PCR, PLS, and KPLS are not designed to explore the local manifold

topology structure and proximity relation among data samples. This would not assure a satisfying prediction performance. The proposed NPR learns the projection matrix to project the original data to low-dimension space from measured data, taking advantages of both regression behavior from the process variables to the quality variable and the dimension reduction via exploring the intrinsic local topology structures among process variables, so it performs the best with the highest accuracy.

The predictions of the penicillin concentration of the penicillin fermentation process at each sampling time are respectively depicted in detail in Fig. 4, by (a) PCR, (b) PLS, (c) KPLS, and (d) NPR. In addition, the prediction errors between the predicted penicillin concentration and the real measurements are also presented. It is easy to see that the prediction results by the proposed NPR match the real measurements best. Also, the prediction errors of NPR have the smallest undulations within the narrowest range around zero. All the results show NPR outperforms the other models.

## 5. CONCLUSIONS

The traditional data-driven virtual sensors mainly define the outer shape of the data, but they cannot provide any insight into the micro perspective of manifold proximity indicating the local relationships among the data samples; manifold learning to construct the virtual sensor models from the micro-perspective (local view) is a practical solution. In the current study, the NPR model to regression is proposed and applied in virtual sensor applications. The features of the proposed method are concluded as follows.

- NPR projects the process variables onto a new space to explore the intrinsic local manifold topology structure and proximity relations among them. Based on localized constructions, NPR is likely to provide a steady and responsible presentation to handle the nonlinear characteristics.
- In addition, NPR discovers the regression behavior between the projected process variables and quality variable to generate accurate estimates of the quality variable.
- NPR aims to synchronously discover the regression relationship as well as data representation in low dimensional manifold space through an innovatively designed optimization. The analytical solution would be compactly and directly calculated without any iterative procedures. The feasibility of the optimization is guaranteed because the two parts of objective function are trivially pursued.

The effectiveness of the proposed NPR algorithm is demonstrated through the case study carried out on a simulated penicillin production process. For space limit, this paper only includes some preliminary simulations; more simulations may be included in a future paper. Our future work will be focused on the integration of the dynamic characteristics into the NPR to catch information between the current event and the just-passed events. As there are usually much more unlabeled data without the measurements of quality variables while these data contains much useful process information, it is desired to the corresponding semi-supervised form with the state-

of-art NPR to utilize the above-mentioned information. Furthermore, the proposed NPR model ignores the information among coupled quality variables, a bilinear form of neighborhood regression model would be further desired in our future work.

## REFERENCES

- Ge Z., Song Z., Ding S.X. and Huang B., Data mining and analytics in the process industry: the role of machine learning. *IEEE Access*, 5(99):20590–20616, 2017.
- Ge, Z., Distributed predictive modeling framework for prediction and diagnosis of key performance index in plant-wide processes. *Journal of Process Control*, 65:107–117, 2018.
- He, X., Cai, D., Yan, S. and Zhang, H.J., Neighborhood preserving embedding. *Tenth IEEE International Conference on Computer Vision (ICCV'05)*, 2:1208–1213, 2005.
- Kadlec, P., Gabrys, B. and Strandt, S., Data-driven soft sensors in the process industry. *Computers and chemical engineering*, 3(44):795–814, 2009.
- Kadlec, P., Grbi, R. and Gabrys, B., Review of adaptation mechanisms for data-driven soft sensors. *Computers and chemical engineering*, 35(1):1–24, 2011.
- Kouropteva, O., Okun, O. and Pietikinen, M., Selection of the optimal parameter value for the locally linear embedding algorithm. *FSKD*, 2:359–363, 2002.
- Lee, J.M., Yoo, C., Choi, S.W., Vanrolleghem, P.A. and Lee, I.B., Nonlinear process monitoring using kernel principal component analysis. *Chemical engineering science*, 59(1):223–234, 2004.
- Liu, K., Zhang, X. and Shi, J., Adaptive sensor allocation strategy for process monitoring and diagnosis in a Bayesian network. *IEEE Transactions on Automation Science and Engineering*, 11(2):452–462, 2013.
- Qin, S.J., Survey on data-driven industrial process monitoring and diagnosis. *Annual reviews in control*, 36(2):220–234, 2012.
- Qin, S.J., Process data analytics in the era of big data. *AIChE Journal*, 60(9):3092–3100, 2014.
- Shao, W., Chen, S. and Harris, C.J., Adaptive Soft Sensor Development for Multi-Output Industrial Processes Based on Selective Ensemble Learning. *IEEE Access*, 5:55628–55642, 2018.
- Wei, C., Chen, J., Song, Z. and Chen, C.I., Adaptive virtual sensors using SNPER for the localized construction and elastic net regularization in nonlinear processes. *Control Engineering Practice*, 83:129–140, 2019.
- Yan, W., Tang, D. and Lin, Y., A data-driven soft sensor modeling method based on deep learning and its application. *IEEE Transactions on Industrial Electronics*, 64(5):4237–4245, 2016.
- Yao, L. and Ge, Z., Nonlinear Gaussian Mixture Regression for Multimode Quality Prediction with Partially Labeled Data. *IEEE Transactions on Industrial Informatics*, 2018.
- Yu, J., Multiway Gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes. *Industrial and Engineering Chemistry Research*, 51(40):13227–13237, 2012.