Knowledge-Based Systems 235 (2022) 107642

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Virtual sensing techniques for nonlinear dynamic processes using weighted probability dynamic dual-latent variable model and its industrial applications

Ze Ying [a,b], Yun Wang [c], Yuchen He [a,b,*], Jun Wang [b]

[a] Key Laboratory of Intelligent Manufacturing Quality Big Data Tracing and Analysis of Zhejiang Province, China Jiliang University, Hangzhou 310018, China
[b] College of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China
[c] Mechanical and Electrical Engineering Department, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou 311122, China

ARTICLE INFO

Article history:
Received 26 October 2020
Received in revised form 29 June 2021
Accepted 30 September 2021
Available online 29 October 2021

Keywords:
Virtual sensors
Latent variables (LVs)
Probabilistic dynamic dual-latent structure (PDDLS)
Process dynamics
Nonlinear transition process

ABSTRACT

Over the past decades, data-driven virtual sensors have been widely used to predict hard-to-measure key quality variables where process uncertainties, dynamics and nonlinearity have been considered as critical data features in modern industries. As a result, in this paper, a virtual sensing technique is developed based on a probabilistic dynamic dual-latent structure (PDDLS) in which two distinct dynamic latent variables (LVs) are introduced to take care of quality-related and quality-unrelated dynamic information within measurements respectively. By combining the local weighted (LW) strategy, the virtual sensing technique is further extended to nonlinear applications. Finally, the performance of the proposed method is verified by two industrial cases where the superiority is shown compared with previous researches.

1. Introduction

With the increasing complexity of modern industrial systems, process safety and product quality have become critical issues in industrial processes [1,2]. Meanwhile, the developments of precise instrumentation, distributed control and proliferation networks facilitate the measurement and storage of massive data, making data-driven models widely popular in industrial applications [3–6]. Among them, data-driven virtual sensing technique has been considered as a popular alternative to predict key variables which are often difficult or costly to measure in real industrial processes [7–9]. In the past decades, several machine learning methods, such as principal component regression (PCR) [10], latent factor (LF) [11,12], partial least squares (PLS) [13] and neural networks (NN) [14,15], have been widely served as virtual sensors where the LVs or layers were designed to depict the inherent characteristics of process data. However, the data collections are likely to be polluted by random noise while the above models actually lack probabilistic interpretation with data, which leads to great modeling difficulty for process uncertainty. Furthermore, the data sample collected at each moment is highly correlated with the previous ones due to a large amount of control loops in the process [16]. To this end, process uncertainty and dynamics are two significant features that should be taken seriously during soft sensor design.

To deal with process uncertainty, probabilistic methods have been developed which can provide a more natural statistical inference for virtual sensing [17]. In recent years, lots of latent variable models like probabilistic principal component regression (PPCR) [18,19] and Gaussian mixture regression (GMR) [20,21] have received great attention. Alternatively, Ge et al. [22,23] proposed a supervised factor analysis (SFA) model by introducing an extra quality-related emission equation so that quality-oriented information can be included. Compared with PPCR, disparate noise variances of measurement variables have been assigned in SFA model, hence better performance can be obtained [23]. However, SFA modeling is mainly implemented based on single latent variable where quality variables are equally considered as ordinary members of expanded process datasets. In such kind of structure, the quality information will be easily overwhelmed by large amounts of process variables which leads to insufficient retention of quality information as well as the relationship between process variables and key quality variables. Driven by the concept of PLS [24,25] models, the probabilistic partial least squares (PPLS) [26] model was then carried out where two sorts of LVs were designed to describe the statistical characteristics of quality-related and quality-unrelated information within the data. Compared with SFA models, the superior virtual sensing performance

* Corresponding author.
E-mail address: yche@cjlu.edu.cn (Y. He).

https://doi.org/10.1016/j.knosys.2021.107642
0950-7051/

of the dual-latent-variables strategy has been demonstrated in previous researches [27,28].

Unfortunately, the basic assumption of PPLS that samples are independent with each other is always invalid in real industrial processes, which indicates the existence of auto-correlations within processes. When handling process dynamics, Markov model has been considered as one of the most efficient solutions [29]. In a typical first order Markov chain, it is assumed that the latent variable of current time step only relies on the previous one, which is widely applied in linear state–space model like linear dynamic system (LDS) [30–33]. For example, Wen et al. [34] proposed a linear Gaussian state space model for dynamic process monitoring where the first order Markov chain in latent state space provided additional dynamic prediction information for the estimation of LVs, which greatly improves the efficiency and accuracy of data feature extraction. For dynamic virtual sensing modeling, Ge & Chen [32] introduced the quality-related information into LDS and developed a supervised linear dynamic model (SLDS) where quality variables are also treated as the expanded input of LDS. As the SLDS is still a single-latent-variable structure, the relationship between quality variables and process variables cannot be properly depicted. Even worse, the residuals part in this kind of latent structure seems unable to capture the dynamics in rest parts of information. As a result, the dynamic characteristics of process cannot be fully described, resulting in poor soft sensing performance.

In addition, due to fluctuation of materials, season factor etc., it is also noticed that the operating conditions in real industrial process are not always stable [35,36]. To deal with changing operating conditions, several effective local modeling methods have been proposed, among which Just-in-Time learning (JITL) [37–39] is the most popular one. In JITL methods, the similarities between training and query sample are obtained and the training samples with the greatest similarities are selected to build online models for quality variables prediction. Recently, a locally weighted linear dynamic system (WLDS) [40] was developed by introducing the locally weighting scheme into the LDS model structure. However, during the online modeling of WLDS, the selection of relevant training samples only relies on the Euclidean distance between the query sample and training samples while the quality related characteristics in latent spaces are neglected in the similarity construction. Compared with the similarity obtained from original data space, the involvement of latent information can help improve the interpretation of local weighted modeling [41].

In this paper, a novel probability dynamic dual-latent-variables structure (PDDLS) is proposed where two sets of LVs are employed which correspond to quality related and quality-unrelated information, respectively. Meanwhile, different first-order Markov chains are designed to characterize each set of LVs for process dynamic information description. Compared with relevant researches, both process dynamics and quality related information are fully concerned. To improve the virtual sensing performance in non-stable processes, the PDDLS is further extended into a weighted PDDLS (WPDDLS) model by introducing a dual-latent weighted (DLW) strategy which takes two kinds of similarities into consideration. The first one represents the similarity in the static information, which can be derived using the similarity between the LVs of training samples and query samples. While the second one indicates the similarity in dynamic information which can be expressed by the time difference increments of LVs. The two similarities are then exploited to obtain an evolving likelihood function for non-stable processes virtual sensing.

To sum up, the main contributions of this paper can be included as follows.

1. In this paper, a dual latent-variable structure is proposed which focuses on the relationship between process variables and quality variables, i.e. the extraction of quality-related information.

2. PDDLS designs different Markov chains for the two latent variables. Compared with the SLDS model, process dynamics can be well reserved in both quality related and quality-unrelated information.

3. A novel local weighted strategy is designed for the PDDLS model to improve the adaptiveness of virtual senser performance in non-stable processes where the LVs-based similarity construction can consider quality-related information as additional condition for selecting relevant training samples in online modeling.

The rest of the paper is arranged as follows: The basic ideas of traditional SLDS will be reviewed in Section 2. Then the PDDLS model will be explained in detail in Section 3 where model structure and parameter learning will be discussed. In Section 4, the novel weighted framework will be employed to extend PDDLS to non-stable occasions and virtual sensing method will then be applied. In Section 5, the performance of the proposed virtual sensor will be evaluated through two typical cases. Finally, some conclusions are given.

2. Supervised linear dynamic system model

In this section, a brief introduction of SLDS will be given. On the basis of linear Gaussian state–space structure, SLDS model is constructed after a supervised restriction is added to traditional LDS model where the quality key variables are involved. The schematic representations of LDS and SLDS model can be represented in Fig. 1(a) and (b), respectively. In these two figures, $q'_t \in \Re^{(l\times 1)}$ and $q''_t \in \Re^{(l\times 1)}$ denote the model latent variables with dimension l, $X = [x_1, x_2, \dots, x_t \dots, x_T]$; $x_t \in \Re^{(M\times 1)}$ are the observations of the process variables (model input), and $Y = [y_1, y_2, \dots, y_t \dots, y_T]$; $y_t \in \Re^{(N\times 1)}$ are the observations of the quality variables (model output), where M, N represent the number of variables contained in the observation variable respectively. As can be seen, the latent dynamic transformation and observation transformation can be found in both models. However, compared with LDS model, another quality observation transformation is established in SLDS model to make quality variable information involved. Therefore, the LVs in SLDS is also referred to as the quality-related latent variable, which try to extract the cross-correlation relationship between process variables and quality variables. The SLDS model structure can be shown as follows:

$$\begin{aligned} q''_t &= Fq''_{t-1} + e_q \\ x_t &= Aq''_t + e_x \\ y_t &= Cq''_t + e_y \end{aligned} \tag{1}$$

where $F \in \Re^{(l\times l)}$ denotes the state transition matrix that determines the latent dynamic translating relationship, and $A \in \Re^{(M\times l)}$, $C \in \Re^{(N\times l)}$ are the emission matrices corresponding to the observation transformation. e_q, e_x, e_y are the interference terms of SLDS model in which residual space $\Re^{(M+N)}$ determined by e_x, e_y can be used to describe the model fitting situation. For more details about SLDS, please refer to [31,32].

3. PDDLS

3.1. Model structure

In SLDS regression modeling, both process variables and quality variables are reflected by the same set of dynamic LVs, indicating the introduction of quality information has not changed the observation transformation between process variables and

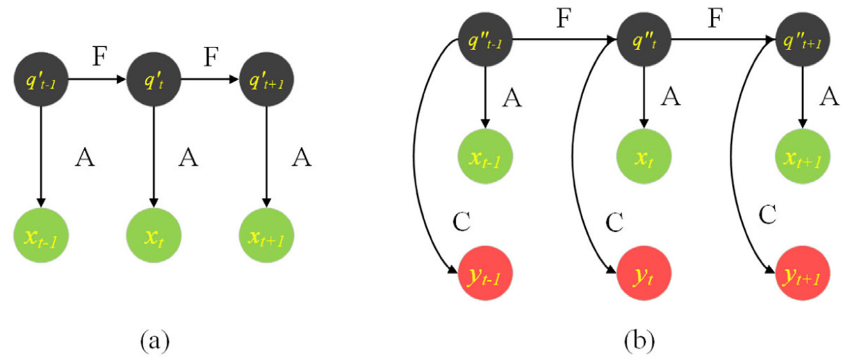


Fig. 1. Structure of LDS and SLDS model. (a) LDS (b) SLDS.

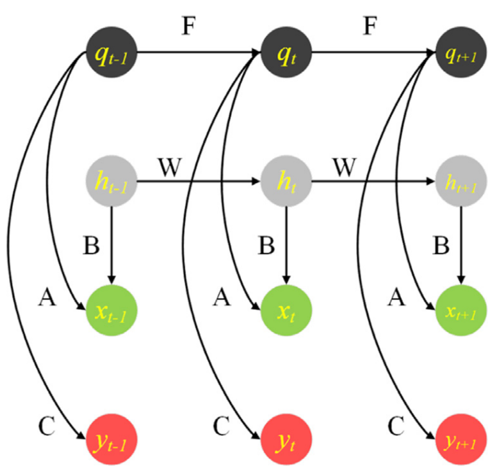


Fig. 2. Model structure of PDDLS model.

corresponding LVs. In other words, all variables are equally important to the LVs where the key quality variable may probably be neglected among massive process variables. Besides, the dynamic characteristics of process cannot be emphasized completely since a large amount of quality variable information and quality-unrelated information are unfortunately classified into the static residual space. To handle the above dilemma, an extra set of dynamic LVs is designed in PDDLS to effectively extract both the quality-related and quality-unrelated dynamic information within measurement variables, as shown in Fig. 2.

It can be clearly seen that the dynamic information of process variables x_t is pre-decomposed into two parts known as quality-related and quality-unrelated information by two dynamic LVs. The quality-related latent variable q_t provides sufficient feature space for quality variable y_t, which is used to establish mathematical relationships with process variable x_t, while the quality-unrelated latent variable h_t only stores the dynamic information of x_t itself. In this way, the quality variable is no longer simply regarded as a part of the process dataset, and the cross-correlation relationship between x_t and y_t can be obtained more effectively in latent variable $q_t \in \Re^{(l\times 1)}$. The PDDLS model is expressed as:

$$
\begin{aligned}
q_t &= Fq_{t-1} + e_q \\
h_t &= Wh_{t-1} + e_h \\
x_t &= Aq_t + Bh_t + e_x \\
y_t &= Cq_t + e_y
\end{aligned} \tag{2}
$$

in which $F \in \Re^{(l\times l)}$, $W \in \Re^{(s\times s)}$ are state transition matrices and $A \in \Re^{(M\times l)}$, $B \in \Re^{(M\times s)}$, $C \in \Re^{(N\times l)}$ are the emission matrices. In the PDDLS model, it is assumed that LVs and interference terms obey the Gaussian distribution $e_q \sim \mathcal{N}\left(0, \Sigma_q\right)$, $e_h \sim \mathcal{N}(0, \Sigma_h)$, $e_x \sim \mathcal{N}(0, \Sigma_x)$, $e_y \sim \mathcal{N}\left(0, \Sigma_y\right)$ where Σ_q, Σ_h, Σ_x, Σ_y are the corresponding covariance diagonal matrices of the interference terms and the distribution of the LVs are initialized as $q_1 \sim \mathcal{N}\left(\mu_\pi^q, \Sigma_\pi^q\right)$ and $h_1 \sim \mathcal{N}\left(\mu_\pi^h, \Sigma_\pi^h\right)$ respectively.

In addition, according to the data structure of the linear Gaussian state space model (LGSS) [29], Eq. (2) can be equally expressed as:

$$
\begin{aligned}
\begin{bmatrix} q_t \\ h_t \end{bmatrix} &= \begin{bmatrix} F & 0 \\ 0 & W \end{bmatrix}\begin{bmatrix} q_{t-1} \\ h_{t-1} \end{bmatrix} + \begin{bmatrix} e_q \\ e_h \end{bmatrix} \\
\begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} A & B \\ C & 0 \end{bmatrix}\begin{bmatrix} q_t \\ h_t \end{bmatrix} + \begin{bmatrix} e_x \\ e_y \end{bmatrix}
\end{aligned} \tag{3}
$$

For convenience, the observed samples and joint latent variables can be denoted as $G = [g_1, g_2, \ldots, g_t \ldots, g_T] \in \Re^{((M+N)\times T)}$; $g_t \triangleq \begin{bmatrix} x_t \\ y_t \end{bmatrix}$ and $Z = [z_1, z_2, \ldots, z_t \ldots, z_T] \in \Re^{((l+s)\times T)}$; $z_t \triangleq \begin{bmatrix} q_t \\ h_t \end{bmatrix}$, respectively. Meanwhile, the state transition matrix and the emission matrix of PDDLS model can be represented by joint matrices $\Lambda = \begin{bmatrix} F & 0 \\ 0 & W \end{bmatrix}$ and $\Psi = \begin{bmatrix} A & B \\ C & 0 \end{bmatrix}$, respectively. Then, the probabilistic transformation relation in Eq. (3) can be further expressed as:

$$
\begin{aligned}
p\left(z_t|z_{t-1}\right) &= \mathcal{N}\left(\Lambda z_{t-1}, \begin{bmatrix} \Sigma_q & 0 \\ 0 & \Sigma_h \end{bmatrix}\right) \\
p\left(g_t|z_t\right) &= \mathcal{N}\left(\Psi z_t, \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_y \end{bmatrix}\right)
\end{aligned} \tag{4}
$$

It should be noted that the joint form of vector/matrix is only for the representation convenience. The internal mechanism of PDDLS and its parameter learning process depend on the cost/likelihood function of the model in Eq. (5). All parameters of PDDLS model $\Theta^{PDDLS} = \left\{\mu_\pi^q, \Sigma_\pi^q, \mu_\pi^h, \Sigma_\pi^h, F, W, A, B, C, \Sigma_q, \Sigma_h, \Sigma_x, \Sigma_y\right\}$ can be solved by classical maximum likelihood estimation method.

$$
\begin{aligned}
\ln p(G) &= \ln \int_Z p(G, Z) dZ \\
&= \ln \int_{z_1}\int_{z_2}\cdots\int_{z_T}\left\{\begin{array}{c} p(z_1)\cdot\prod_{t=2}^{T} p(z_t|z_{t-1})\cdot \\ \prod_{t=1}^{T} p(g_t|z_t)\cdot dz_1 dz_2\cdots dz_T \end{array}\right\}
\end{aligned} \tag{5}
$$

which seems difficult to handle merely using the traditional maximum likelihood estimation method. To solve this problem, the expectation maximization (EM) algorithm is applied trying to maximize complete-data likelihood function including observed variables and corresponding LVs.

3.2. Parameter estimation using EM algorithm

E step: On the basis of complete-data likelihood function in Eq. (5), the EM algorithm is carried out for parameters estimation which includes iterations composed of E-step and M-step. In E-step, traditional Kalman forward filtering and backward smoothing algorithm are employed to calculate the posterior expectation of LVs, which will be given in Appendix A in details. Combined with Eqs. (26)–(28) in Appendix A, the final posterior expectation of LVs can be expressed as follows:

$$E\left\langle z_t|g_{1:T}\right\rangle = \begin{bmatrix} E\left\langle q_t\right\rangle \\ E\left\langle h_t\right\rangle \end{bmatrix} = \hat{\mu}_t$$
$$E\left\langle z_t z_t^T|g_{1:T}\right\rangle = \begin{bmatrix} E\left\langle q_t q_t^T\right\rangle & E\left\langle q_t h_t^T\right\rangle \\ E\left\langle h_t q_t^T\right\rangle & E\left\langle h_t h_t^T\right\rangle \end{bmatrix} = \hat{V}_t + \hat{\mu}_t\hat{\mu}_t^T \tag{6}$$
$$E\left\langle z_t z_{t-1}^T|g_{1:T}\right\rangle = \begin{bmatrix} E\left\langle q_t q_{t-1}^T\right\rangle & E\left\langle q_t h_{t-1}^T\right\rangle \\ E\left\langle h_t q_{t-1}^T\right\rangle & E\left\langle h_t h_{t-1}^T\right\rangle \end{bmatrix} = \hat{V}_t J_{t-1}^T + \hat{\mu}_t\hat{\mu}_t^T$$

in which $\hat{\mu}_t$ and $\hat{V}_t$ are the mean vector and covariance matrix of joint latent variables under the condition of measurements g_t collected at the sampling time t, respectively. J_t is the intermediate variable in the calculation process. After that, the expectation of the log-likelihood function, denoted as $E_{Z|G}\left\langle \ln P(G)\right\rangle$, can be given by Eq. (7).

$$\begin{aligned} \ln p\left(G\right) = & \int_{z_1} p(z_1|g_1)\cdot \ln p(z_1, \mu_\pi^q, \Sigma_\pi^q, \mu_\pi^h, \Sigma_\pi^h)dz_1 \\ & + \sum_{t=2}^{T}\int_{z_t} p(z_t|g_t)\cdot \ln p(z_t|z_{t-1}, \Lambda, \Sigma_q, \Sigma_h)dz_t \\ & + \sum_{t=1}^{T}\int_{z_t} p(z_t|g_t)\cdot \ln p(g_t|z_t, \Psi, \Sigma_x, \Sigma_y)dz_t + const \end{aligned} \tag{7}$$

here $E_{Z|G}\left\langle \ln p(G)\right\rangle \Leftrightarrow \int_Z p\left(Z|G\right)\cdot \ln p(G)dZ = \ln p(G)$.

M step: the parameter set Θ^{PDDLS} can then be updated by setting the partial derivatives of Eq. (7) to zero with respect to each parameter. More detailed explanation about the M step can be found in Appendix A. The updated parameters can be expressed as follows:

$$\begin{aligned} u_\pi^{q^{new}} &= E\left\langle q_1\right\rangle \\ u_\pi^{h^{new}} &= E\left\langle h_1\right\rangle \end{aligned} \tag{8}$$

$$\begin{aligned} \sum_{\pi}^{q^{new}} &= E\left\langle q_1 q_1^T\right\rangle - E\left\langle q_1\right\rangle E\left\langle q_1^T\right\rangle \\ \sum_{\pi}^{h^{new}} &= E\left\langle h_1 h_1^T\right\rangle - E\left\langle h_1\right\rangle E\left\langle h_1^T\right\rangle \end{aligned} \tag{9}$$

$$\begin{aligned} F^{new} &= \left(\sum_{t=2}^{T} E\left\langle q_t q_{t-1}^T\right\rangle\right)\left(\sum_{t=2}^{T} E\left\langle q_{t-1} q_{t-1}^T\right\rangle\right)^{-1} \\ W^{new} &= \left(\sum_{t=2}^{T} E\left\langle h_t h_{t-1}^T\right\rangle\right)\left(\sum_{t=2}^{T} E\left\langle h_{t-1} h_{t-1}^T\right\rangle\right)^{-1} \end{aligned} \tag{10}$$

$$\begin{aligned} \sum_{q}^{new} &= \frac{1}{T-1}\cdot\sum_{t=2}^{T}\left\{\begin{array}{c} E\left\langle q_t q_t^T\right\rangle - F^{new}\cdot E\left\langle q_{t-1} q_t^T\right\rangle \\ -E\left\langle q_t q_{t-1}^T\right\rangle\cdot\left(F^{new}\right)^T \\ +F^{new}\cdot E\left\langle q_{t-1} q_{t-1}^T\right\rangle\cdot\left(F^{new}\right)^T \end{array}\right\} \\ \sum_{h}^{new} &= \frac{1}{T-1}\cdot\sum_{t=2}^{T}\left\{\begin{array}{c} E\left\langle h_t h_t^T\right\rangle - W^{new}\cdot E\left\langle h_{t-1} h_t^T\right\rangle \\ -E\left\langle h_t h_{t-1}^T\right\rangle\cdot\left(W^{new}\right)^T \\ +W^{new}\cdot E\left\langle h_{t-1} h_{t-1}^T\right\rangle\cdot\left(W^{new}\right)^T \end{array}\right\} \end{aligned} \tag{11}$$

$$\begin{aligned} A^{new} &= \left(\sum_{t=1}^{T} x_t E\left\langle q_t^T\right\rangle\right)\left(\sum_{t=1}^{T} E\left\langle q_t q_t^T\right\rangle\right)^{-1} \\ B^{new} &= \left(\sum_{t=1}^{T} x_t E\left\langle h_t^T\right\rangle\right)\left(\sum_{t=1}^{T} E\left\langle h_t h_t^T\right\rangle\right)^{-1} \\ C^{new} &= \left(\sum_{t=1}^{T} y_t E\left\langle q_t^T\right\rangle\right)\left(\sum_{t=1}^{T} E\left\langle q_t q_t^T\right\rangle\right)^{-1} \end{aligned} \tag{12}$$

$$\begin{aligned} \sum_{x}^{new} &= \frac{1}{T}\cdot\sum_{t=1}^{T}\left\{\begin{array}{c} x_t x_t^T - \left[A^{new}B^{new}\right]\cdot E\left\langle z_t\right\rangle x_t^T \\ -x_t E\left\langle z_t^T\right\rangle\cdot\left[A^{new}B^{new}\right]^T \\ +\left[A^{new}B^{new}\right]\cdot E\left\langle z_t z_t^T\right\rangle\cdot\left[A^{new}B^{new}\right]^T \end{array}\right\} \\ \sum_{y}^{new} &= \frac{1}{T}\cdot\sum_{t=1}^{T}\left\{\begin{array}{c} y_t y_t^T - C^{new}\cdot E\left\langle q_t\right\rangle y_t^T \\ -y_t^T E\left\langle q_t^T\right\rangle\cdot\left(C^{new}\right)^T \\ +C^{new}\cdot E\left\langle q_t q_t^T\right\rangle\cdot\left(C^{new}\right)^T \end{array}\right\} \end{aligned} \tag{13}$$

4. Weighted PDDLS and virtual sensor applications

In order to realize adaptive quality prediction, the above PDDLS model is further extended to a weighted form, namely the weighted PDDLS model (WPDDLS) where parameter estimation is implemented on the basis of weighted logarithmic likelihood function. Hence, the model construction of WPDDLS is composed of weighted strategy and EM parameter estimation, which will be explained in Section 4.1. On the basis of WPDDLS method, corresponding virtual sensing applications will then be discussed in Section 4.2 to give a comprehensive explanation for key variables prediction.

4.1. WPDDLS

4.1.1. Weighted strategy of WPDDLS

Local weighted model has been considered as one of the most efficient approaches to realize local linearization for nonlinear models, which aims to build a local model for each query sample [38,40]. The contribution of each historical training sample to parameter training can be adjusted by weights which are generally determined by the similarities (spatial distance) between the training samples and the query sample corresponding to the same local model. Assume that a query sample x^{query} and the nonlinear model determined by $y_t = f\left(x_t\right) = a_4 x_t^4 + a_3 x_t^3 + a_2 x_t^2 + a_1 x_t^1 + a_0$ are employed to predict the corresponding output value, then the following local model cost function can be established:

$$\begin{aligned} &\underset{a_0 a_1 a_2 a_3 a_4}{\arg\min}\left\{\sum_{t=1}^{T}\mathbb{R}_t\left\|y_t - f(x_t)\right\|_2\right\} \\ &\mathbb{R}_t = \varphi(x_1, x_2, \ldots, x_T, x^{query}) \end{aligned} \tag{14}$$

where $\mathbb{R}_t$ describes the similarities between x^{query} and training set samples $\{x_t, y_t\}_{t\in[1,T]}$. It can be seen that $\mathbb{R}_t$ plays a regular term role in the process of cost function minimizing. The contribution of the training sample will be retained to the greatest extent if the similarity value $\mathbb{R}_{t=1,2,3\ldots T}$ is large at a certain time. Otherwise, $\mathbb{R}_{t=1,2,3\ldots T}$ will assign $\left\|y_t - f(x_t)\right\|_2$ a small value. In this way,

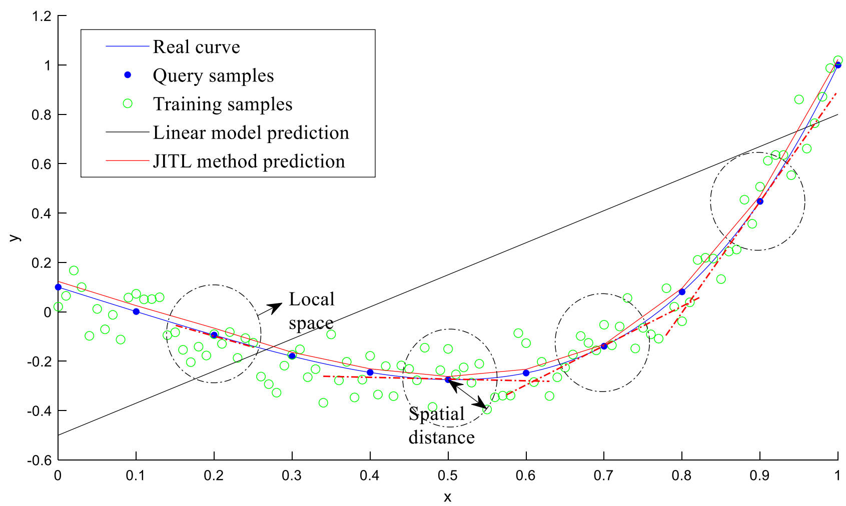


Fig. 3. Fit nonlinear curve using LW (JITL) technology.

LW technique can realize the local linearization of nonlinear processes, as shown in Fig. 3.

Inspired by the weighting strategy, the logarithmic likelihood function of PDDLS can be further modified for nonlinear modeling. It is noted that the logarithmic likelihood function of PDDLS in Eq. (7) contains two transformation terms, that is, the latent state transformation $p\left(z_t|z_{t-1}\right)$ and the state emission transformation $p\left(g_t|z_t\right)$, which cannot solve nonlinear problems as they are both linear transformation relations. In order to improve the nonlinear adaptiveness for PDDLS model, two different sorts of weights should be assigned to corresponding terms of the logarithmic likelihood function in Eq. (7) which is then maximized to obtain corresponding parameters of WPDDLS. Two different weights will be discussed in detail as follows.

Weights for state emission

Unlike traditional LW methods, the similarity of LVs in the proposed model are employed for state emission weights construction. Suppose $q_t^{train} = \left\{q_1^{train}, q_2^{train} \cdots q_T^{train} \cdots\right\}$ are the corresponding quality-related LVs for training samples and q_k^{test} is the query latent variable estimation based on the local model parameter set at the previous time. Commonly, the Euclidean distance is exploited to calculate the similarity between the query sample and training samples. On this basis, corresponding weights $\gamma_{tk} \in \Re^{T\times 1}$ are then constructed as follows:

$$\begin{aligned} d_{tk} &= \sqrt{\left(q_t^{train} - q_k^{test}\right)^T \left(q_t^{train} - q_k^{test}\right)} \\ \gamma_{tk} &= \exp\left(-d_{tk}^2/\alpha^2\right);\ k = T+1, T+2, T+3\cdots \end{aligned} \tag{15}$$

where d_{tk} represents the Euclidean distance between each pair of LVs q_t^{train} and q_k^{test}. α represents the model adjustable hyperparameters.

Weights for state transition

The latent state transformation $p\left(z_t|z_{t-1}\right)$ usually indicates state changes in the Markov chain. Therefore, the weights assigned to the state transformation term should well reflect the dynamic relationships in the time series. As a result, the first-order time difference increment of the latent variable is employed to measure the state changes and served as weights λ_{tk}.

$$\begin{aligned} \Delta q_t^{train} &= q_t^{train} - q_{t-1}^{train} \\ \Delta q_k^{test} &= q_k^{test} - q_{k-1}^{test} \end{aligned} \tag{16}$$

$$\begin{aligned} \Delta d_{tk} &= \sqrt{\left(\Delta q_t^{train} - \Delta q_k^{test}\right)^T \left(\Delta q_t^{train} - \Delta q_k^{test}\right)} \\ \lambda_{tk} &= \exp\left(-\Delta d_{tk}^2/\beta^2\right) \end{aligned} \tag{17}$$

in which β is the corresponding model adjustable hyperparameters. Finally, the weights calculated by Eqs. (15) and (17) are added to the corresponding terms of log-likelihood function to construct a weighted logarithmic likelihood function shown as follows:

$$\begin{aligned} \ln p\left(G\right) = &\int_{z_1} p(z_1|g_1)\cdot \ln p(z_1)dz_1 \\ &+ \sum_{t=2}^{T}\int_{z_t} p(z_t|g_t)\cdot \lambda_{tk}\cdot \ln p(z_t|z_{t-1})dz_t \\ &+ \sum_{t=1}^{T}\int_{z_t} p(z_t|g_t)\cdot \gamma_{tk}\cdot \ln p(g_t|z_t)dz_t + const \end{aligned} \tag{18}$$

4.1.2. EM parameter estimation for WPDDLS

It is worth noting that the posterior expectation of the LVs $q_{t=1:T}^{train}$ is calculated according to the existing training samples and is independent of the newly introduced query samples. Therefore, the E step of WPDDLS is similar with PDDLS (Refer to Eqs. (26)-(28) in Appendix A), the parameters of WPDDLS can be updated by maximizing the expectation of the weighted log-likelihood function which can be expressed as follows:

$$\begin{aligned}\widetilde{F}^{new}&=\left(\sum_{t=2}^{T}\lambda_{tk}\cdot E\left\langle q_t q_{t-1}^T\right\rangle\right)\left(\sum_{t=2}^{T}\lambda_{tk}\cdot E\left\langle q_{t-1} q_{t-1}^T\right\rangle\right)^{-1}\\ \widetilde{W}^{new}&=\left(\sum_{t=2}^{T}\lambda_{tk}\cdot E\left\langle h_t h_{t-1}^T\right\rangle\right)\left(\sum_{t=2}^{T}\lambda_{tk}\cdot E\left\langle h_{t-1} h_{t-1}^T\right\rangle\right)^{-1}\end{aligned}\tag{19}$$

$$\begin{aligned}\widetilde{\sum}_{q}^{new}&=\frac{1}{\sum_{t=2}^{T}\lambda_{tk}}\cdot\sum_{t=2}^{T}\lambda_{tk}\cdot\left\{\begin{array}{c}E\left\langle q_t q_t^T\right\rangle-\widetilde{F}^{new}\cdot E\left\langle q_{t-1}q_t^T\right\rangle\\ -E\left\langle q_t q_{t-1}^T\right\rangle\cdot\left(\widetilde{F}^{new}\right)^T\\ +\widetilde{F}^{new}\cdot E\left\langle q_{t-1}q_{t-1}^T\right\rangle\cdot\left(\widetilde{F}^{new}\right)^T\end{array}\right\}\\ \widetilde{\sum}_{h}^{new}&=\frac{1}{\sum_{t=2}^{T}\lambda_{tk}}\cdot\sum_{t=2}^{T}\lambda_{tk}\cdot\left\{\begin{array}{c}E\left\langle h_t h_t^T\right\rangle-\widetilde{W}^{new}\cdot E\left\langle h_{t-1}h_t^T\right\rangle\\ -E\left\langle h_t h_{t-1}^T\right\rangle\cdot\left(\widetilde{W}^{new}\right)^T\\ +\widetilde{W}^{new}\cdot E\left\langle h_{t-1}h_{t-1}^T\right\rangle\cdot\left(\widetilde{W}^{new}\right)^T\end{array}\right\}\end{aligned}\tag{20}$$

$$\begin{aligned}\widetilde{A}^{new}&=\left(\sum_{t=1}^{T}\gamma_{tk}\cdot x_t E\left\langle q_t^T\right\rangle\right)\left(\sum_{t=1}^{T}\gamma_{tk}\cdot E\left\langle q_t q_t^T\right\rangle\right)^{-1}\\ \widetilde{B}^{new}&=\left(\sum_{t=1}^{T}\gamma_{tk}\cdot x_t E\left\langle h_t^T\right\rangle\right)\left(\sum_{t=1}^{T}\gamma_{tk}\cdot E\left\langle h_t h_t^T\right\rangle\right)^{-1}\\ \widetilde{C}^{new}&=\left(\sum_{t=1}^{T}\gamma_{tk}\cdot y_t E\left\langle q_t^T\right\rangle\right)\left(\sum_{t=1}^{T}\gamma_{tk}\cdot E\left\langle q_t q_t^T\right\rangle\right)^{-1}\end{aligned}\tag{21}$$

$$\begin{aligned}\widetilde{\sum}_{x}^{new}&=\frac{1}{\sum_{t=1}^{T}\gamma_{tk}}\times\sum_{t=1}^{T}\gamma_{tk}\cdot\left\{\begin{array}{c}x_t x_t^T-\left[\widetilde{A}^{new}\widetilde{B}^{new}\right]\cdot E\left\langle z_t\right\rangle x_t^T\\ -x_t E\left\langle z_t^T\right\rangle\cdot\left[\widetilde{A}^{new}\widetilde{B}^{new}\right]^T\\ +\left[\widetilde{A}^{new}\widetilde{B}^{new}\right]\cdot E\left\langle z_t z_t^T\right\rangle\\ \cdot\left[\widetilde{A}^{new}\widetilde{B}^{new}\right]^T\end{array}\right\}\\ \widetilde{\sum}_{y}^{new}&=\frac{1}{\sum_{t=1}^{T}\gamma_{tk}}\cdot\sum_{t=1}^{T}\gamma_{tk}\cdot\left\{\begin{array}{c}y_t y_t^T-\widetilde{C}^{new}\cdot E\left\langle q_t\right\rangle y_t^T\\ -y_t^T E\left\langle q_t^T\right\rangle\cdot\left(\widetilde{C}^{new}\right)^T\\ +\widetilde{C}^{new}\cdot E\left\langle q_t q_t^T\right\rangle\cdot\left(\widetilde{C}^{new}\right)^T\end{array}\right\}\end{aligned}\tag{22}$$

Compared with original PDDLS model, the parameter set $\widetilde{\Theta}_k$ of WPDDLS can be modified according to the state of current query sample, which will largely improve the virtual sensing performance of nonlinear processes.

4.2. Virtual sensor applications

The virtual sensing applications are designed for nonlinear processes based on the aforementioned WPDDLS model where the LVs are utilized to reveal the relationship between process variables and quality key variables. Consequently, in this part, the final LVs $\hat{q}_k^{test}$ derived from WPDDLS modeling are employed for virtual sensor construction, which can be expressed as follows:

$$\begin{aligned}\hat{q}_k^{test}&=\widetilde{F}^{new}\hat{u}_{k-1}+K_k\left(x^{query}-\widetilde{A}^{new}\widetilde{F}^{new}\hat{u}_{k-1}\right)\\ K_k&=P_{k-1}\left(\widetilde{A}^{new}\right)^T\left[\widetilde{A}^{new}P_{k-1}\left(\widetilde{A}^{new}\right)^T+\widetilde{\sum}_{x}^{new}\right]^{-1}\\ P_{k-1}&=\widetilde{F}^{new}\hat{V}_{k-1}\left(\widetilde{F}^{new}\right)^T+\widetilde{\sum}_{q}^{new}\end{aligned}\tag{23}$$

in which parameters $\widetilde{A}^{new}$, $\widetilde{F}^{new}$, $\widetilde{\Sigma}_x^{new}$, $\widetilde{\Sigma}_q^{new}$ are the members of $\widetilde{\Theta}_k$ determined by the local model for the current sample x_k^{query}.

To sum up, the detailed steps of WPDDLS can be shown as follows:

(1) The two kinds of local weight vectors $\gamma_{tk}=\{\gamma_{1k},\gamma_{2k}\cdots\gamma_{Tk}\}$ and $\lambda_{tk}=\{\lambda_{1k},\lambda_{2k}\cdots\lambda_{Tk}\}$ can be calculated according to Eqs. (15) and (17).

(2) The above weights are then assigned to the corresponding terms of logarithmic likelihood function to construct the weighted logarithmic likelihood function. The EM algorithm is used to update the model parameter set $\widetilde{\Theta}_k$ and the latent variable estimation q_{k+1}^{test} can be obtained for next weights calculation.

When the parameter set $\widetilde{\Theta}_k$ is updated each time, the latent variable $\hat{q}_k^{test}$ of current query sample can be derived according to Eq. (23). Finally, the quality variable $y_k^{prediction}=\widetilde{C}^{new}\hat{q}_k^{test}$ is determined. The pseudocode for the proposed method can be expressed in Table 1.

Table 1
The pseudocode of WPDDLS.

Algorithm 1 Virtual Sensing Based on WPDDLS

Input: training dataset G, query samples x_k^{query}
Output: quality variable corresponding to each query sample

initialize: $\gamma_{tk}=\lambda_{tk}=\vec{0},\alpha,\beta,\varepsilon,l,s,\tilde{\Theta}_{init}^{WPDDLS}$;
data pre-processing;
function MODEL $(G,\gamma_{tk},\lambda_{tk},\alpha,\beta,\varepsilon,l,s)$
while $L\left(\tilde{\Theta}_{new}^{WPDDLS}\right)-L\left(\tilde{\Theta}_{old}^{WPDDLS}\right)>\varepsilon$ **do**
  $\langle E-step\rangle\big|_{\theta_{init}^{WPDDLS}}$ for the posteriori expectation of latent variables q_t^{train};
  % Eqs. (6) & (26)–(28)
  if $\gamma_{tk}=\lambda_{tk}=\vec{0}$ **then**
    $\langle M-step\rangle$ of PDDLS is used to update model parameters θ_{new}^{PDDLS};
  % Eqs. (8)–(13)
  else
    $\langle M-step\rangle$ of WPDDLS is used to update model parameters $\tilde{\Theta}_{new}^{WPDDLS}$;
  % Eqs. (19)–(21)
  end if
end while
end function
$k=1$;
while $k<$ *number of query samples* **do**
  $\langle E-step\rangle\big|_{\theta_{k-1}^{WPDDLS}}$ for estimation of query latent variable q_k^{train};
  % $\tilde{\Theta}_{k-1=0}^{WPDDLS}\triangleq\theta^{PDDLS}$
  calculate the similarity between q_t^{train} and q_k^{test} to update the weights $\gamma_{tk}^{new},\lambda_{tk}^{new}$;
  % Eqs. (15)–(17)
  $\tilde{\Theta}_k^{WPDDLS}=$ **MODEL** $(x_k^{query},\gamma_{tk}^{new},\lambda_{tk}^{new},\alpha,\beta,\varepsilon,l,s)$;
  $\langle E-step\rangle\big|_{\theta_k^{WPDDLS}}$ for the final calculation of the query latent variable $\hat{q}_k^{test}$;
  $y_k^{prediction}=\widetilde{C}^{new}\hat{q}_k^{test}$; % prediction
end while

5. Case study

In this section, the virtual sensing performance of the proposed model is verified by two industrial processes: the blast furnace ironmaking and Tennessee Eastman process, which are both typical industrial processes with transitions. For comparison, the performance of SFA [42], SLDS [32], WFA [38], WLDS [40] and PDDLS are also discussed in detail. Due to space limitation, the differences between the models are explained in detail in Appendix B.

Table 2
Variable selection and introduction in TE process.

No.	Variable description	No.	Variable description
1.	D feed flow (stream 2)	16.	Reactor pressure
2.	E feed flow (stream 3)	17.	Reactor level
3.	A feed flow (stream 1)	18.	Reactor temperature
4.	A & C feed flow (stream 4)	19.	Purge rate (stream 9)
5.	Separator pot liquid flow (stream 10)	20.	Product separator temperature
6.	Stripper liquid product flow (stream11)	21.	Product separator level
7.	Stripper flow valve	22.	Product separator pressure
8.	Reactor cooling water outlet temperature	23.	Product separator underflow (stream 10)
9.	Separator cooling water outlet temperature	24.	Stripper level
10.	A feed (stream 1)	25.	Stripper pressure
11.	D feed (stream 2)	26.	Stripper underflow (stream 11)
12.	E feed (stream 3)	27.	Stripper temperature
13.	A & C feed (stream 4)	28.	Stripper steam flow
14.	Recycle flow (stream 8)	29.	Reactor cooling water outlet temperature
15.	Reactor feed rate (stream 6)	30.	Reactor cooling water outlet temperature
Key quality variable		1.	A component (stream 6)

Table 3
The performance indexes and prediction results of different models in TE process.

Model	SFA	WFA	SLDS	WLDS	PDDLS	WPDDLS
RMSE	1.3916	0.6859	0.7505	0.4771	0.6249	0.3557
R^2	0.1086	0.5735	0.7221	0.8545	0.8131	0.9091

5.1. Tennessee Eastman (TE) benchmark

TE benchmark is a chemical simulation platform designed by Eastman company that is widely used to evaluate the process monitoring performance between different algorithms. A total of 41 measurement variables and 12 manipulated variables are collected in TE process. Among them, 19 quality variables measured by three component analyzers are included in the measurement variables. The remaining 22 measurement variables and 12 control variables are classified as process variables, which are easier to obtain than quality variables. In order to realize the prediction of quality key variables, 12 500 sample data are collected from TE process shown in Fig. 4. In this section, 31 process variables and one quality variable are selected as input and output of virtual sensor respectively, as shown in Table 2. Next, the collected data was pre-processed in advance, that is, each variable obeys a normal distribution after then. In addition, Fig. 5 shows the trend of each variable. It can be seen that there are five obvious transitions in 12 500 samples, which actually describes an unsteady and nonlinear industrial process.

During the virtual sensor modeling based on SFA, WFA, SLDS, WLDS, PDDLS and WPDDLS model, the first 10 000 samples are utilized as training data, and the last 2500 samples are considered as query samples. Through cross validation, it is determined that the latent variable number is set to 8. The root mean square error (RMSE) and coefficient of determination (R^2) are used in this paper to evaluate the virtual sensing performance, as shown in Table 3. Fig. 6 shows the prediction results of each model.

The prediction accuracy of virtual sensor based on SFA model is the worst since SFA does not consider the process dynamics and nonlinearity, which leads to the weak adaptiveness of model, as shown in Fig. 6(a). As a weighted form of SFA model, the WFA model uses the samples that are most similar to the query samples in the training set for model training, so as to realize the local linearization of nonlinear relationship. It can be seen from Fig. 6(b) that the prediction accuracy of SFA has been significantly improved using the LW strategy. Besides, it is worth mentioning that merely partial characteristics of process data are taken into consideration in both SLDS and WFA models, and the results they predicted are also different: ($R^2_{WFA} < R^2_{SLDS}$, $RMSE_{WFA} < RMSE_{SLDS}$). It means that the dynamic probabilistic latent variable model can improve the tracking precision while the LW strategy can improve the prediction accuracy. Therefore, the prediction performance of WSLDS can be enhanced compared with SLDS and WFA since both process dynamics and nonlinearity are considered. However, due to the equal treatment of quality key variables in the above model, the extracted quality related information is not representative. In contrast, the relationship between process data and quality data is focused in PDDLS and WPDDLS due to the dual-latent variable strategy. Both RMSE and R^2 have been greatly improved compared with other methods.

5.2. Blast furnace ironmaking

To effectively implement the quality control strategy in the blast furnace ironmaking process, the quality index of molten iron should be accurately measured. Commonly, the silicon content of molten iron is an important factor to depict the molten iron quality but is often intractable to obtain in real time [43]. We need to predict the silicon content in molten iron using the easy-to-measure process variables. In this case, 1000 samples were collected where 5 process variables and 1 quality key variable are involved, as shown in Table 4. The prediction results of each model can be found in Table 5 and Fig. 7.

Similarly, the prediction results of SFA are poor since it does not consider the dynamic characteristics of process data. The prediction performance of PDDLS is still better than SLDS since it introduces two kinds of dynamic LVs to deal with the qual-

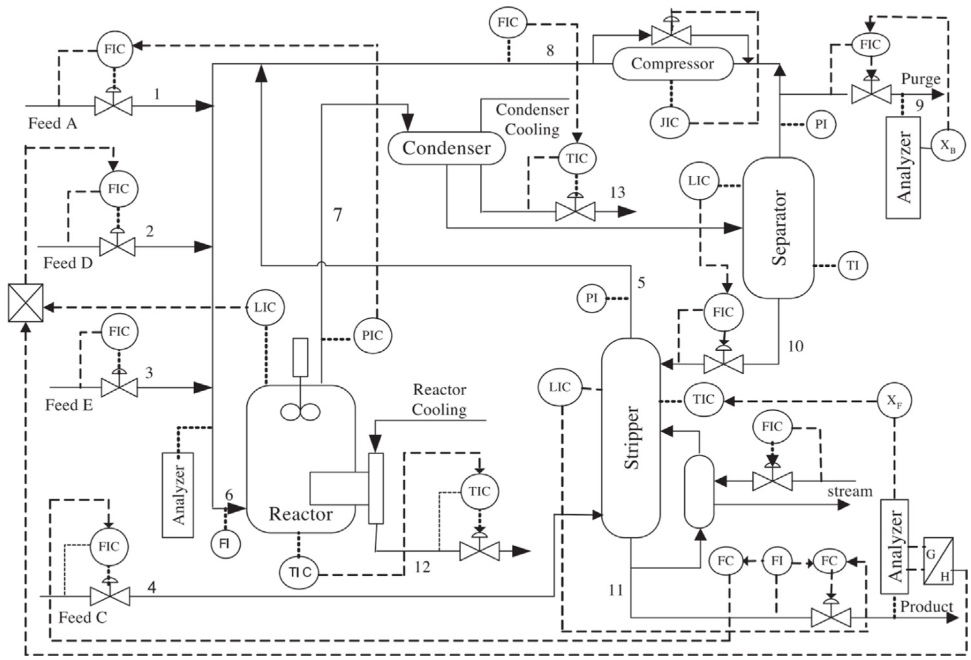


Fig. 4. Tennessee Eastman (TE) process Flow Chart.

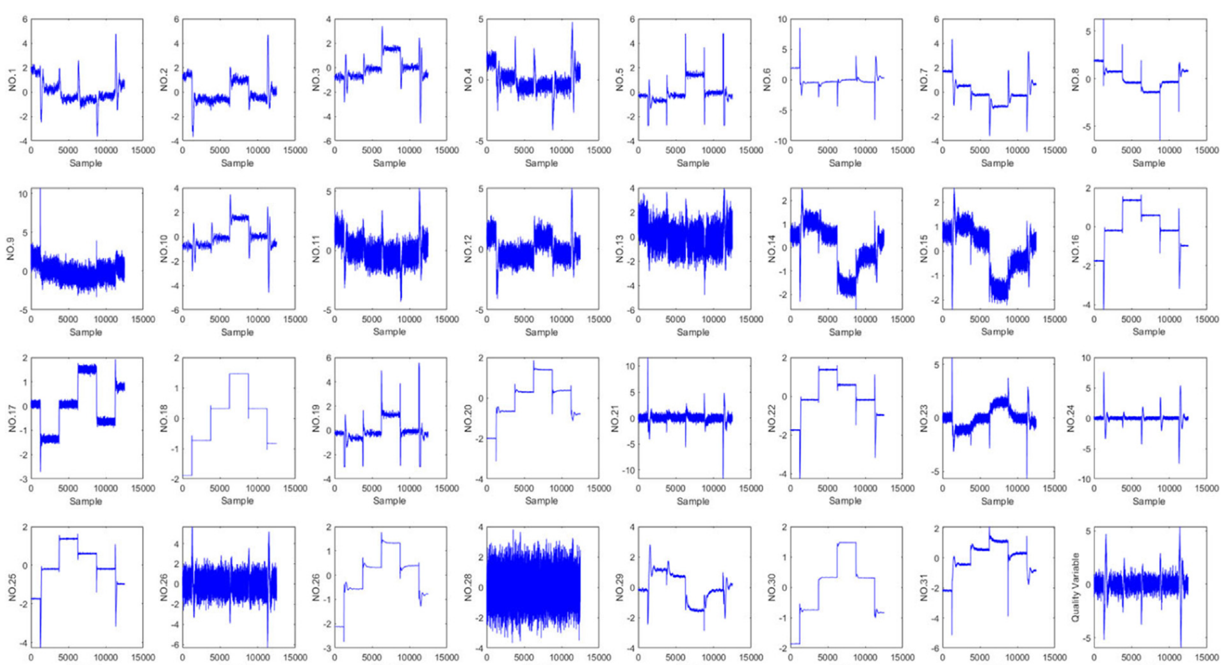


Fig. 5. Data trends for input and output variables in the TE process.

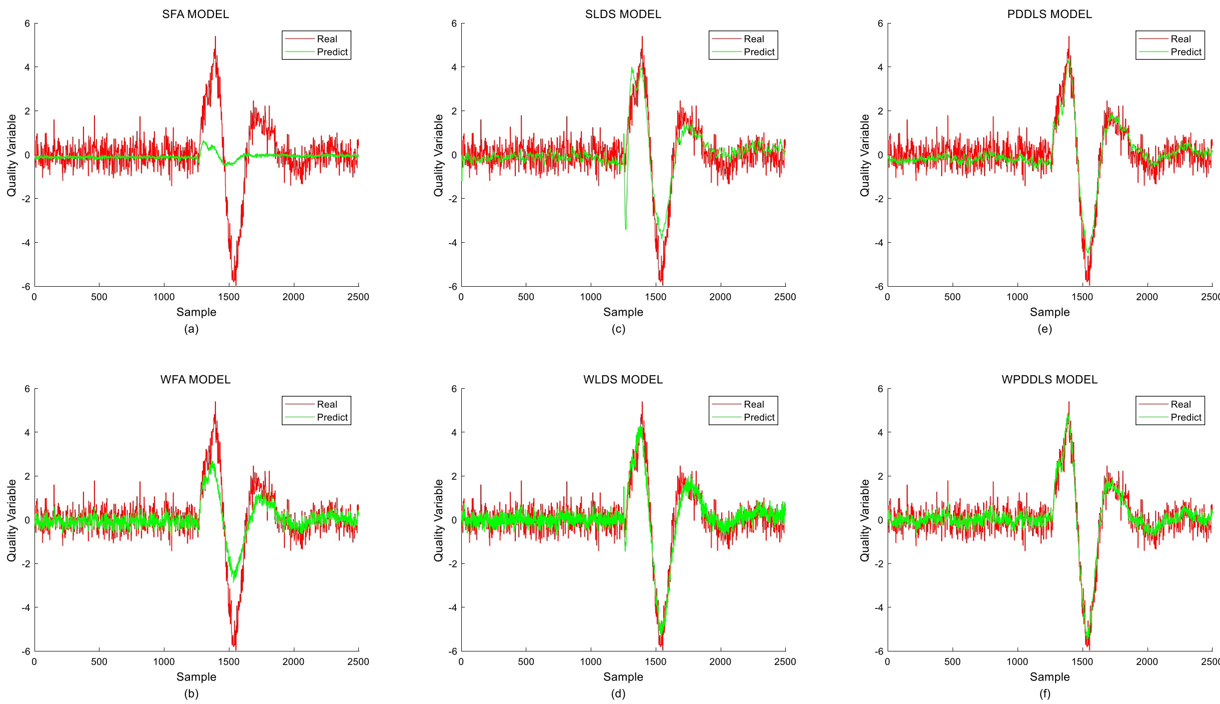


Fig. 6. Quality prediction results of different models in TE process.

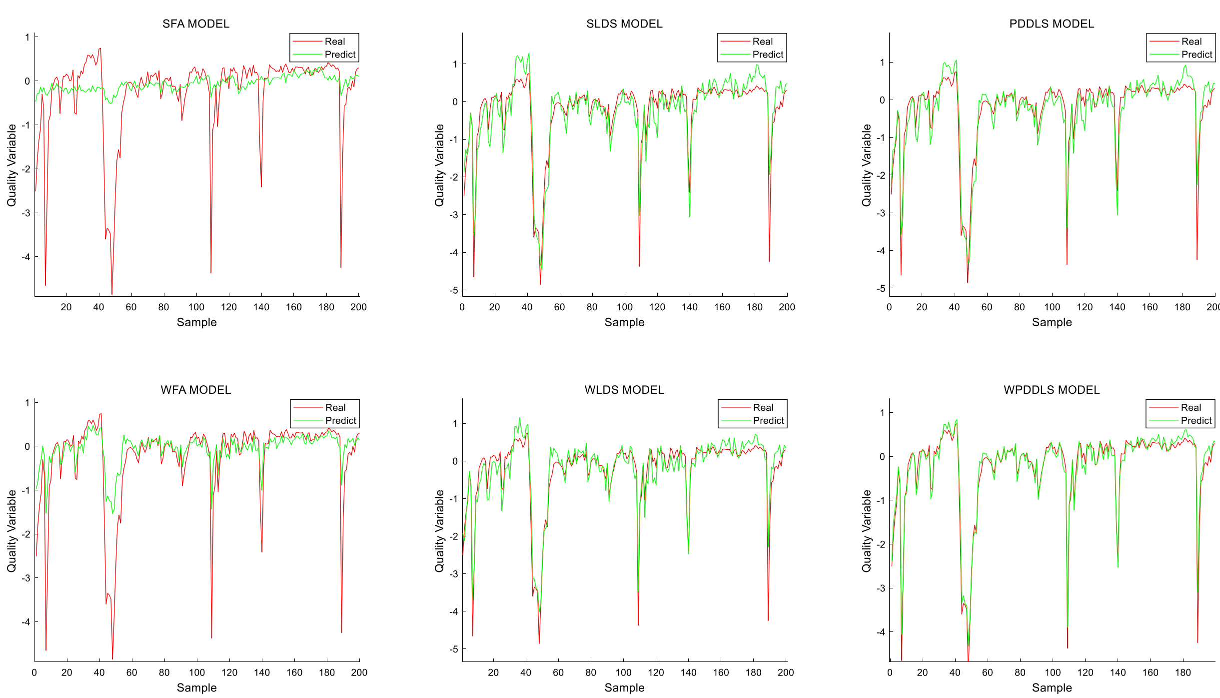


Fig. 7. The quality prediction results in different methods in the process of blast furnace ironmaking.

Table 4
Variable description in the blast furnace ironmaking process.

No.	Variable description
1.	Coal injection rate
2.	Explosion intensity
3.	Ventilation rate
4.	Material decline rate
5.	Pressure difference
6.	Silicon content of molten iron

Table 5
Quality prediction results of different methods in the blast furnace ironmaking process.

Model	SFA	SLDS	PDDLS	WFA	WLDS	WPDDLS
RMSE	0.9960	0.4249	0.3612	0.6620	0.3319	0.2398
R^2	0.0839	0.8334	0.8711	0.5452	0.8854	0.9432

ity related and quality-unrelated information in process data respectively. Therefore, PDDLS is more suitable for correlation description between process variables and quality variables in dynamic process. In addition, the prediction performance improves after applying the LW strategy. Especially, in WPDDLS modeling, different weights are designed and then assigned to different part of the log-likelihood function, which further improves the prediction accuracy. Experimental results show that WPDDLS model has a superior prediction performance over other relative methods.

6. Conclusion

In this paper, the data dynamic and nonlinear characteristics in industrial processes are mainly concerned using a weighted PDDLS model. Two different first-order Markov chains are involved to describe data dynamic characteristics. Furthermore, in comparison with the traditional supervised LDS model, the relationship between process variables and quality key variables has been emphasized in PDDLS. Therefore, the prediction accuracy of PDDLS model has been enhanced. To deal with the process nonlinearity, a weighted PDDLS is established combined with LW technique where the model parameters are estimated by the EM algorithm. Different from ordinary LW models, the quality-related information is considered as an important reference for similarity measurement. Finally, two industrial examples show that WPDDLS has great adaptive quality prediction ability and can extract nonlinear dynamic characteristics of data more effectively.

It is worth noting that the proposed model in this paper is based on the assumption of data integrity. However, due to the limitations of measurement conditions, only a small part of the sample data is marked by quality variables, which indicates the situation of data missing. Fortunately, the proposed model still has some advantages in dealing with this problem. Firstly, probabilistic frameworks can provide more natural statistical explanations for missing data which can be replaced by predictions. Secondly, the first order Markov chain in the PDDLS model can be used to learn the overall trend of the variables. WPDDLS will be extended to semi-supervised or multi-rate modeling applications, which will be discussed in our future works.

CRediT authorship contribution statement

Ze Ying: Conceptualization, Data curation, Investigation, Software, Visualization, Writing – original draft. **Yun Wang:** Methodology, Validation. **Yuchen He:** Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Jun Wang:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by National Natural Science Foundation of China (61903352), Zhejiang Province Natural Science Foundation of China (LQ19F030007), Project of department of education of Zhejiang Province under Grant (Y202044960), Fundamental Research Funds for the Provincial Universities of Zhejiang (2021YW18, 2021YW80), China Postdoctoral Science Foundation (2020M671721).

Appendix A. Detailed derivation of the EM algorithm for PDDLS

Given the training data set $\{x_t, y_t\}_{t\in[1,T]}$, the expression of complete-data logarithmic likelihood function can be given as:

$$\begin{aligned}\ln p(G,Z) &= \ln p(g_1g_2\ldots g_T, z_1z_2\ldots z_T)\\ &= \ln\left(p(z_1)\cdot\prod_{t=2}^{T}p(z_t|z_{t-1})\cdot\prod_{t=1}^{T}p(g_t|z_t)\right)\end{aligned} \tag{24}$$

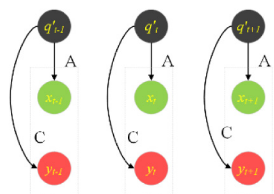
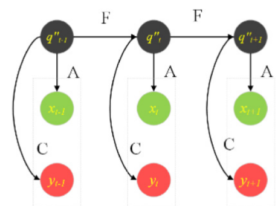
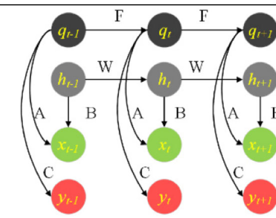
here $\ln p(G) \propto \ln p(G,Z)$

For estimating the members of Θ^{PDDLS} using EM algorithm, we take the expectation of Eq. (24) with respect to the latent variable posterior distribution $p(z_t|g_{1:T})$, corresponding to Eq. (7), which is then maximized to update parameters of PDDLS. Eq. (25) is given in Box I,

$$
\begin{aligned}
&\underset{\Theta^{PDDLS}}{\operatorname{argmax}}\left\{\begin{array}{c}
-\frac{1}{2}\ln\left|\begin{bmatrix}\sum_{\pi}^{q} & 0\\ 0 & \sum_{\pi}^{h}\end{bmatrix}\right| - E_{Z|G}\left\langle \frac{1}{2}\left(z_1-\begin{bmatrix}\mu_{\pi}^{q}\\ \mu_{\pi}^{h}\end{bmatrix}\right)^{T}\begin{bmatrix}\sum_{\pi}^{q} & 0\\ 0 & \sum_{\pi}^{h}\end{bmatrix}^{-1}\left(z_1-\begin{bmatrix}\mu_{\pi}^{q}\\ \mu_{\pi}^{h}\end{bmatrix}\right)\right\rangle\\
-\frac{T-1}{2}\ln\left|\begin{bmatrix}\sum_{q} & 0\\ 0 & \sum_{h}\end{bmatrix}\right| - E_{Z|G}\left\langle \frac{1}{2}\sum_{t=2}^{T}(z_t-\Lambda z_{t-1})^{T}\begin{bmatrix}\sum_{q} & 0\\ 0 & \sum_{h}\end{bmatrix}^{-1}(z_t-\Lambda z_{t-1})\right\rangle\\
-\frac{T}{2}\ln\left|\begin{bmatrix}\sum_{x} & 0\\ 0 & \sum_{y}\end{bmatrix}\right| - E_{Z|G}\left\langle \frac{1}{2}\sum_{t=1}^{T}(g_t-\Psi z_t)^{T}\begin{bmatrix}\sum_{x} & 0\\ 0 & \sum_{y}\end{bmatrix}^{-1}(g_t-\Psi z_t)\right\rangle\\
+const
\end{array}\right\}\\
&\Rightarrow \frac{\partial\left(\begin{array}{c}
\underbrace{\begin{array}{c}
-\frac{1}{2}\ln\left|\begin{bmatrix}\sum_{\pi}^{q} & 0\\ 0 & \sum_{\pi}^{h}\end{bmatrix}\right| - \frac{1}{2}tr\left(E\left\langle z_1z_1^{T}\right\rangle\begin{bmatrix}\sum_{\pi}^{q} & 0\\ 0 & \sum_{\pi}^{h}\end{bmatrix}^{-1}\right)+\\
E\left\langle z_1^{T}\right\rangle\begin{bmatrix}\sum_{\pi}^{q} & 0\\ 0 & \sum_{\pi}^{h}\end{bmatrix}^{-1}\begin{bmatrix}\mu_{\pi}^{q}\\ \mu_{\pi}^{h}\end{bmatrix}-\frac{1}{2}\begin{bmatrix}\mu_{\pi}^{q}\\ \mu_{\pi}^{h}\end{bmatrix}^{T}\begin{bmatrix}\sum_{\pi}^{q} & 0\\ 0 & \sum_{\pi}^{h}\end{bmatrix}^{-1}\begin{bmatrix}\mu_{\pi}^{q}\\ \mu_{\pi}^{h}\end{bmatrix}
\end{array}}_{part\ \ 1}\\
\underbrace{-\frac{T-1}{2}\ln\left|\begin{bmatrix}\sum_{q} & 0\\ 0 & \sum_{h}\end{bmatrix}\right| - \frac{1}{2}\sum_{t=2}^{T}\left\{\begin{array}{l}
tr\left(E\left\langle z_tz_t^{T}\right\rangle\begin{bmatrix}\sum_{q} & 0\\ 0 & \sum_{h}\end{bmatrix}^{-1}\right)-\\
2tr\left(E\left\langle z_tz_{t-1}^{T}\right\rangle\Lambda^{T}\begin{bmatrix}\sum_{q} & 0\\ 0 & \sum_{h}\end{bmatrix}^{-1}\right)+\\
tr\left(E\left\langle z_{t-1}z_{t-1}^{T}\right\rangle\Lambda^{T}\begin{bmatrix}\sum_{q} & 0\\ 0 & \sum_{h}\end{bmatrix}^{-1}\Lambda\right)
\end{array}\right\}}_{part\ \ 2}\\
\underbrace{-\frac{T}{2}\ln\left|\begin{bmatrix}\sum_{x} & 0\\ 0 & \sum_{y}\end{bmatrix}\right| - \frac{1}{2}\sum_{t=1}^{T}\left\{\begin{array}{l}
tr\left(g_tg_t^{T}\begin{bmatrix}\sum_{x} & 0\\ 0 & \sum_{y}\end{bmatrix}^{-1}\right)-\\
2tr\left(g_t^{*}E\left\langle z_t\right\rangle\Psi^{T}\begin{bmatrix}\sum_{x} & 0\\ 0 & \sum_{y}\end{bmatrix}^{-1}\right)+\\
tr\left(E\left\langle z_tz_t^{T}\right\rangle\Psi^{T}\begin{bmatrix}\sum_{x} & 0\\ 0 & \sum_{y}\end{bmatrix}^{-1}\Psi\right)
\end{array}\right\}}_{part\ \ 3}
\end{array}\right)}{\partial\left\{\mu_{\pi}^{q},\sum_{\pi}^{q},\mu_{\pi}^{h},\sum_{\pi}^{h},F,W,A,B,C,\sum_{q},\sum_{h},\sum_{x},\sum_{y}\right\}}\triangleq 0
\end{aligned}
\tag{25}
$$

Box I.

Table 6
Explanation of latent variables for different model structures.

MODEL	SFA	SLDS	PDDLS
Model structure			
Model formula	$x_t = Aq_t^{'} + e_x$ $y_t = Cq_t^{'} + e_y$ $here: q_t^{'} \sim \mathcal{N}(0, \mathrm{I})$	$q_t^{''} = Fq_{t-1}^{''} + e_q$ $x_t = Aq_t^{''} + e_x$ $y_t = Cq_t^{''} + e_y$	$q_t = Fq_{t-1} + e_q$ $h_t = Wh_{t-1} + e_h$ $x_t = Aq_t + Bh_t + e_x$ $y_t = Cq_t + e_y$
Likelihood function	$\ln\left(\prod_{t=1}^{T} p\left(x_t, y_t \mid q_t^{'}\right)\right)$	$\ln\left(\begin{array}{c} p\left(q_1^{''}\right) \times \prod_{t=2}^{T} p\left(q_t^{''} \mid q_{t-1}^{''}\right) \\ \times \prod_{t=1}^{T} p\left(x_t, y_t \mid q_t^{''}\right) \end{array}\right)$	$\ln\left(\begin{array}{c} p\left(q_1, h_1\right) \\ \times \prod_{t=2}^{T} p\left(q_t, h_t \mid q_{t-1}, h_{t-1}\right) \\ \times \prod_{t=1}^{T} p\left(x_t, y_t \mid q_t, h_t\right) \end{array}\right)$
Characteristics of LVs	1. Single quality-related LVs 2. LVs are independent of each other	1. Single quality-related LVs 2. Dynamic LVs	1. Double-latent-variable structure 2. Dynamic LVs

in which $p\left(z_t \mid g_{1:T}\right) \sim \mathcal{N}\left(\hat{\mu}_t, \hat{V}_t\right)$ is derived from E step:

$$\begin{aligned}
\mu_t^z &= \Lambda\mu_{t-1}^z + K_t\left(g_t - \Psi\Lambda\mu_{t-1}^z\right)\\
V_t^z &= P_{t-1} - K_t\Psi P_{t-1}\\
K_t &= P_{t-1}\Psi^T\left[\Psi P_{t-1}\Psi^T + \begin{bmatrix}\sum_x & 0\\ 0 & \sum_y\end{bmatrix}\right]^{-1}\\
P_{t-1} &= \Lambda V_{t-1}^z\Lambda^T + \begin{bmatrix}\sum_q & 0\\ 0 & \sum_h\end{bmatrix}
\end{aligned}\tag{26}$$

especially:

$$\begin{aligned}
\mu_1^z &= \begin{bmatrix}\mu_\pi^q\\ \mu_\pi^h\end{bmatrix} + \begin{bmatrix}\sum_\pi^q & 0\\ 0 & \sum_\pi^h\end{bmatrix}\Psi^T\left[\Psi\begin{bmatrix}\sum_\pi^q & 0\\ 0 & \sum_\pi^h\end{bmatrix}\Psi^T\right.\\
&\quad\left.+\begin{bmatrix}\sum_x & 0\\ 0 & \sum_y\end{bmatrix}\right]^{-1}\left(g_1 - \Psi\begin{bmatrix}\mu_\pi^q\\ \mu_\pi^h\end{bmatrix}\right)\\
V_1^z &= \begin{bmatrix}\sum_\pi^q & 0\\ 0 & \sum_\pi^h\end{bmatrix} - \begin{bmatrix}\sum_\pi^q & 0\\ 0 & \sum_\pi^h\end{bmatrix}\Psi^T\left[\Psi\begin{bmatrix}\sum_\pi^q & 0\\ 0 & \sum_\pi^h\end{bmatrix}\Psi^T\right.\\
&\quad\left.+\begin{bmatrix}\sum_x & 0\\ 0 & \sum_y\end{bmatrix}\right]^{-1}\Psi\begin{bmatrix}\sum_\pi^q & 0\\ 0 & \sum_\pi^h\end{bmatrix}
\end{aligned}\tag{27}$$

Notice μ_t^z, V_t^z are preliminary estimates of LVs posterior distribution using Kalman forward filtering. The $\hat{\mu}_t$ and $\hat{V}_t$ need to be further acquired by Kalman backward smoothing algorithm:

$$\begin{aligned}
&\hat{\mu}_T = \mu_T^z;\ \hat{V}_T = V_T^z\\
&\hat{\mu}_t = \mu_t^z + J_t\left(\hat{\mu}_{t+1} - \Lambda\mu_t^z\right)\\
&\hat{V}_t = V_t^z + J_t(\hat{V}_{t+1} - P_t)J_t^T\\
&J_t = V_t^z\Lambda^T\left(P_t\right)^{-1}
\end{aligned}\tag{28}$$

in which J_t, P_t and K_t(Kalman Gain) are intermediate variables. Then, substitute Eq. (28) into Eq. (6) to enter the next step of the EM algorithm.

In the M step, setting the partial derivative of Eq. (25) with respect to each parameter to zero to get updated value of newly Θ^{PDDLS}.

Take all parts in Eq. (25) irrelevant to each parameter into the constant term and get:

$$\begin{aligned}
&\frac{\partial\left(E_{z_1\mid g_1}\left\langle \ln p(z_1; \mu_\pi^q, \sum_\pi^q, \mu_\pi^h, \sum_\pi^h)\right\rangle\right)}{\partial\left(\left\{\mu_\pi^q, \mu_\pi^h\right\}\right)} \triangleq 0\\
&\Rightarrow \frac{\partial\left(part\ 1\right)}{\partial\left(\begin{bmatrix}\mu_\pi^q\\ \mu_\pi^h\end{bmatrix}\right)} \triangleq 0 \Rightarrow \begin{bmatrix}\mu_\pi^q\\ \mu_\pi^h\end{bmatrix} = E\left\langle z_1\right\rangle
\end{aligned}\tag{29}$$

$$\begin{aligned}
&\frac{\partial\left(E_{z_1\mid g_1}\left\langle \ln p(z_1; \mu_\pi^q, \sum_\pi^q, \mu_\pi^h, \sum_\pi^h)\right\rangle\right)}{\partial\left(\left\{\sum_\pi^q, \sum_\pi^h\right\}\right)} \triangleq 0\\
&\Rightarrow \frac{\partial\left(part\ 1\right)}{\partial\left(\begin{bmatrix}\sum_\pi^q & 0\\ 0 & \sum_\pi^h\end{bmatrix}\right)} \triangleq 0\\
&\Rightarrow \begin{bmatrix}\sum_\pi^q & \sum_\pi^{qh}\\ \sum_\pi^{hq} & \sum_\pi^h\end{bmatrix} = E\left\langle z_1 z_1^T\right\rangle - E\left\langle z_1\right\rangle E\left\langle z_1^T\right\rangle
\end{aligned}\tag{30}$$

$$\frac{\partial\left(\sum_{t=2}^{T}E_{z_t|g_t}\left\langle \ln p(z_t|z_{t-1};\Lambda,\sum_q,\sum_h)\right\rangle\right)}{\partial(\Lambda)}\triangleq 0\Rightarrow\frac{\partial\,(part\ 2)}{\partial(\Lambda)}\triangleq 0$$
$$\Rightarrow \Lambda=\left(\sum_{t=2}^{T}E\left\langle z_t z_{t-1}^T\right\rangle\right)\left(\sum_{t=2}^{T}E\left\langle z_{t-1}z_{t-1}^T\right\rangle\right)^{-1} \tag{31}$$

$$\frac{\partial\left(\sum_{t=2}^{T}E_{z_t|g_t}\left\langle \ln p(z_t|z_{t-1};\Lambda,\sum_q,\sum_h)\right\rangle\right)}{\partial\left(\left\{\sum_q,\sum_h\right\}\right)}\triangleq 0$$
$$\Rightarrow\frac{\partial\,(part\ 2)}{\partial\left(\begin{bmatrix}\sum_q & \sum_{qh}\\ \sum_{hq} & \sum_h\end{bmatrix}\right)}\triangleq 0 \tag{32}$$
$$\Rightarrow\begin{bmatrix}\sum_q & \sum_{qh}\\ \sum_{hq} & \sum_h\end{bmatrix}=\frac{1}{T-1}\cdot\sum_{t=2}^{T}\left\{\begin{array}{c}E\left\langle z_t z_t^T\right\rangle-\Lambda\cdot E\left\langle z_{t-1}z_t^T\right\rangle\\ -E\left\langle z_t z_{t-1}^T\right\rangle\cdot(\Lambda)^T\\ +\Lambda\cdot E\left\langle z_{t-1}z_{t-1}^T\right\rangle\cdot\Lambda^T\end{array}\right\}$$

$$\frac{\partial\left(\sum_{t=1}^{T}E_{z_t|g_t}\left\langle \ln p(g_t|z_t;\Psi,\sum_x,\sum_y)\right\rangle\right)}{\partial(\Psi)}\triangleq 0\Rightarrow\frac{\partial\,(part\ 3)}{\partial(\Psi)}\triangleq 0$$
$$\Rightarrow \Psi=\left(\sum_{t=1}^{T}g_t E\left\langle z_t^T\right\rangle\right)\left(\sum_{t=1}^{T}E\left\langle z_t z_t^T\right\rangle\right)^{-1} \tag{33}$$

$$\frac{\partial\left(\sum_{t=1}^{T}E_{z_t|g_t}\left\langle \ln p(g_t|z_t;\Psi,\sum_x,\sum_y)\right\rangle\right)}{\partial\left(\left\{\sum_x,\sum_y\right\}\right)}\triangleq 0$$
$$\Rightarrow\frac{\partial\,(part\ 3)}{\partial\left(\begin{bmatrix}\sum_x & \sum_{xy}\\ \sum_{yx} & \sum_y\end{bmatrix}\right)}\triangleq 0 \tag{34}$$
$$\Rightarrow\begin{bmatrix}\sum_x & \sum_{xy}\\ \sum_{yx} & \sum_y\end{bmatrix}=\left\{\begin{array}{c}g_t g_t^T-\Psi\cdot E\left\langle z_t\right\rangle g_t^T-g_t^T E\left\langle z_t^T\right\rangle\cdot\Psi^T\\ +\Psi\cdot E\left\langle z_t z_t^T\right\rangle\cdot\Psi^T\end{array}\right\}$$

where $E\langle z_t\rangle$, $E\langle z_t z_t^T\rangle$, $E\langle z_t z_{t-1}^T\rangle$ are given by Eq. (6). Furthermore, each member in Θ^{PDDLS} can be obtained according to the symmetry of matrix operations.

Appendix B. Characteristics of different LVs in different virtual sensor models

Table 6 lists the difference among SFA, SLDS and PDDLS model.

References

[1] Y.C. Jiang, S. Yin, O. Kaynak, Data-driven monitoring and safety control of industrial cyber–physical system: basics and beyond, IEEE Access 6 (2018) 47374–47384.
[2] J. Chen, B. Cui, Y.Q. Chen, B. Zhuang, An improved cooperative team spraying control of a diffusion process with a moving or static pollution source, IEEE/CAA J. Autom. Sin. 7 (2) (2020) 494–504.
[3] S. Yin, X.W. Li, H.J. Gao, O. Kaynak, Data-based techniques focused on modern industry: an overview, IEEE Trans. Ind. Electron. 62 (1) (2015) 657–667.
[4] Q.C. Jiang, X.F. Yan, B. Huang, Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes, Ind. Eng. Chem. Res. 58 (29) (2019) 12899–12912.
[5] L. Yao, Z.Q. Ge, Big data quality prediction in the process industry: A distributed parallel modeling framework, J. Process Control 68 (2018) 1–13.
[6] X. Luo, M. Zhou, S. Li, M. Shang, An inherently nonnegative latent factor model for high-dimensional and sparse matrices from industrial applications, IEEE Trans. Ind. Inf. 14 (5) (2018) 2011–2022.
[7] Z.Q. Ge, Process data analytics via probabilistic latent variable models: a tutorial review, Ind. Eng. Chem. Res. 57 (38) (2018) 12646–12661.
[8] Z.Q. Ge, Distributed predictive modeling framework for prediction and diagnosis of key performance index in plant-wide processes, J. Process Control 65 (2018) 107–117.
[9] C. Shang, B. Huang, F. Yang, D.X. Huang, Probabilistic slow feature analysis-based representation learning from massive process data for soft sensor modeling, AICHE J. 61 (12) (2015) 4126–4139.
[10] X. Yuan, Z. Ge, Z. Song, Locally weighted kernel principal component regression model for soft sensing of nonlinear time-variant processes, Ind. Eng. Chem. Res. 53 (35) (2014) 13736–13749.
[11] X. Luo, M.C. Zhou, Y. Xia, Q. Zhu, A.C. Ammari, A. Alabdulwahab, Generating highly accurate predictions for missing QoS data via aggregating nonnegative latent factor models, IEEE Trans. Neural Netw. Learn. Syst. 27 (3) (2015) 524–537.
[12] Luo Xin, M.C. Zhou, et al., Incorporation of efficient second-order solvers into latent factor models for accurate prediction of missing QoS data, IEEE Trans. Cybern. 48 (4) (2018) 1216–1228.
[13] W.M. Shao, X.M. Tian, Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models, Chem. Eng. Res. Des. 95 (2015) 113–132.
[14] L. Yao, Z.Q. Ge, Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application, IEEE Trans. Ind. Electron. 65 (2) (2018) 1490–1498.
[15] Q.Q. Sun, Z.Q. Ge, Probabilistic sequential network for deep learning of complex process data and soft sensor application, IEEE Trans. Ind. Inf. 15 (5) (2019) 2700–2709.
[16] K.D. Do, Stability in probability and inverse optimal control of evolution systems driven by Levy processes, IEEE/CAA J. Autom. Sin. 7 (2) (2020) 405–419.
[17] Khatibisepehr Shima, B. Huang, Dealing with irregular data in soft sensors: bayesian method and comparative study, Ind. Eng. Chem. Res. 47 (22) (2008) 8713–8723.
[18] Z.Q. Ge, Mixture bayesian regularization of PCR model and soft sensing application, IEEE Trans. Ind. Electron. 62 (7) (2015) 4336–4343.
[19] P.B. Zhu, X. Liu, Y.B. Wang, X.Q. Yang, Mixture semisupervised bayesian principal component regression for soft sensor modeling, IEEE Access 6 (2018) 40909–40919.
[20] W.M. Shao, Z.Q. Ge, L. Yao, Z.H. Song, BayesIan nonlinear gaussian mixture regression and its application to virtual sensing for multimode industrial processes, IEEE Trans. Autom. Sci. Eng. 17 (2) (2020) 871–885.
[21] W.M. Shao, Z.Q. Ge, Z.H. Song, Soft-sensor development for processes with multiple operating modes based on semisupervised gaussian mixture regression, IEEE Trans. Control Syst. Technol. 27 (5) (2019) 2169–2181.
[22] Z.Y. Yang, L. Yao, Z.Q. Ge, A novel streaming variational bayesian supervised factor analysis for industrial adaptive soft sensor modeling, in: Asia Control Conference, 2019, pp. 850–855.
[23] Z. Ge, Supervised latent factor analysis for process data regression modeling and soft sensor application, IEEE Trans. Control Syst. Technol. 24 (3) (2016) 1004–1011.
[24] J.F. MacGregor, C. Jaeckle, C. Kiparissides, M. Koutoudi, Process monitoring and diagnosis by multiblock PLS methods, AIChE J. 40 (5) (1994) 826–838.
[25] G. Li, S.J. Qin, D. Zhou, Brief paper: geometric properties of partial least squares for process monitoring, vol. 46, Pergamon Press, Inc., 2010, pp. 204–210, no. 1.
[26] J.H. Zheng, Z.H. Song, Z.Q. Ge, Probabilistic learning of partial least squares regression model: theory and industrial applications, Chemometr. Intell. Lab. Syst. 158 (2016) 80–90.
[27] J.H. Zheng, Z.H. Song, Semisupervised learning for probabilistic partial least squares regression model and soft sensor application, J. Process Control 64 (2018) 123–131.
[28] S. Li, J. Gao, J.O. Nyagilo, D.P. Dave, Probabilistic partial least square regression: a robust model for quantitative analysis of raman spectroscopy data, in: IEEE International Conference on Bioinformatics and Biomedicine, 2011, pp. 526–531.
[29] C.M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag New York, 2006, p. xx, 738.
[30] Y. Cong, L. Zhou, Z.H. Song, Z.Q. Ge, Multirate dynamic process monitoring based on multirate linear gaussian state-space model, IEEE Trans. Autom. Sci. Eng. 16 (4) (2019) 1708–1719.

[31] Z.Q. Ge, X.R. Chen, Supervised linear dynamic system model for quality related fault detection in dynamic processes, J. Process Control 44 (2016) 224–235.
[32] Z.Q. Ge, X.R. Chen, Dynamic probabilistic latent variable model for process data modeling and regression application, IEEE Trans. Control Syst. Technol. 27 (1) (2019) 323–331.
[33] Y. Wu, X.L. Luo, A novel calibration approach of soft sensor based on multirate data fusion technology, J. Process Control 20 (10) (2010) 1252–1260.
[34] Q. Wen, Z. Ge, Z. Song, Data-based linear Gaussian state-space model for dynamic process monitoring, AICHE J. 58 (12) (2012) 3763–3776.
[35] C.H. Wei, J.H. Chen, Z.H. Song, C.I. Chen, Development of self-learning kernel regression models for virtual sensors on nonlinear processes, IEEE Trans. Autom. Sci. Eng. 16 (1) (2019) 286–297.
[36] W. Shao, Z. Ge, Z. Song, L. Zhou, Data-driven predictive model based on locally weighted bayesian gaussian regression, in: IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS), 2019.
[37] W.M. Shao, Z.Q. Ge, Z.H. Song, BayesIan just-in-time learning and its application to industrial soft sensing, IEEE Trans. Ind. Inf. 16 (4) (2020) 2787–2798.
[38] L. Yao, Z.Q. Ge, Locally weighted prediction methods for latent factor analysis with supervised and semisupervised process data, IEEE Trans. Autom. Sci. Eng. 14 (1) (2017) 126–138.
[39] X.F. Yuan, Z.Q. Ge, Z.H. Song, Y.L. Wang, C.H. Yang, H.W. Zhang, Soft sensor modeling of nonlinear industrial processes based on weighted probabilistic projection regression, IEEE Trans. Instrum. Meas. 66 (4) (2017) 837–845.
[40] X.F. Yuan, Y.L. Wang, C.H. Yang, Z.Q. Ge, Z.H. Song, W.H. Gui, Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes, IEEE Trans. Ind. Electron. 65 (2) (2018) 1508–1517.
[41] X. Yuan, Z. Ge, B. Huang, Z. Song, A probabilistic just-in-time learning framework for soft sensor development with missing data, IEEE Trans. Control Syst. Technol. 25 (3) (2017) 1124–1132.
[42] Z.Q. Ge, Supervised latent factor analysis for process data regression modeling and soft sensor application, IEEE Trans. Control Syst. Technol. 24 (3) (2016) 1004–1011.
[43] J.S. Zeng, C.H. Gao, H.Y. Su, Data-driven predictive control for blast furnace ironmaking process, Comput. Chem. Eng. 34 (11) (2010) 1854–1862.