Received May 3, 2021, accepted May 16, 2021, date of publication May 20, 2021, date of current version June 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3082284

A Dynamic Latent Structure With Time-Varying Parameters for Virtual Sensing of Industrial Process With Irregular Missing Data

ZE YING[1], YUN WANG[2], JUN WANG[1], AND YUCHEN HE[1]

[1]College of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China

[2]Mechanical and Electrical Engineering Department, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou 311123, China

Corresponding author: Yun Wang (wy@zjtongji.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61903352, in part by the Zhejiang Province Natural Science Foundation of China under Grant LQ19F030007, in part by the Department of Education of Zhejiang Province Project under Grant Y202044960, in part by the Project of Zhejiang Tongji Vocational College of Science and Technology under Grant TRC1904, and in part by the China Postdoctoral Science Foundation under Grant 2020M671721.

ABSTRACT This paper proposes a novel dynamic latent structure with time-varying parameters for virtual sensing of industrial process with irregular missing data. The proposed latent structure is based on the linear dynamic system (LDS) model. In order to capture the time-varying process characteristics, a Karman filter based parameter updating method is developed and a virtual sensor is constructed to predict hard-to-measure quality variables. The latent variable structure of the improved model enables the virtual sensor to capture the variable cross-correlation and autocorrelation in the missing data by considering both spatial and temporal information, so that the information in the missing data can be learned from those not missing in the samples as well as the Markov process in the temporal stream. Incorporation of both spatial and temporal information renders more flexibility, resulting in a virtual sensor with higher prediction accuracy even with irregular missing data. The better performance of the proposed method is verified by two industrial applications with different ratios of irregular missing data.

INDEX TERMS Data-driven process monitoring, virtual sensors, irregularly missing data, linear dynamic systems (LDS), linear time-varying parameters, Kalman filtering, EM algorithm.

I. INTRODUCTION

With the development of distributed control and proliferation networks in modern industries, a large amount of data is collected and stored, making data-driven modeling, monitoring and control an attractive research direction [1]–[3]. In many industrial processes, quality variables are usually key indicators that intuitively reflect the quality and efficiency of industrial production. However, in some cases, quality variables are difficult to measure online due to high cost or hostile environment. In order to overcome this difficulty, data-driven virtual sensors have been widely used as a popular alternative to estimate hard-to-measure key quality variables from easy-to-measure process variables [4]–[6]. Such virtual sensors are often based on methods like principal components regression [7], partial least squares [8], artificial neural network [9], or more recently, deep learning techniques [10]. As these methods are data driven, the quality of data becomes a critical problem, which may subject to noise, outliers or missing data etc. [11], [12].

Among the data quality issues, handling of missing data is perhaps the most important. Plenty of methods have been proposed to deal with this issue, from missing data removal to imputation using regression or probabilistic methods [13]–[15]. Whilst data removal may cause loss of meaningful information, data imputation becomes very important, especially for tasks like online learning or monitoring. For instance, Masuda *et al.* [16] proposed a fault detection method with a pre-designed virtual sensing model to predict the missing values. Deng *et al.* [17] proposed a missing data imputation method based on the space-time nearest neighbor value to ensure the performance of subsequent data processing and available data entries. Alternatively, Luo *et al.* [18] adopted least squares regression to train the

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Xiao.

required data features from the existing observations in the original data matrix, so as to achieve an accurate prediction for missing serve-quality data. Lin *et al.* [19] reconstructed missing data by introducing deep learning framework to deal with random and large-scale missing problem. The above mentioned work showed mixed success, as inappropriate imputation brings additional information and may deteriorate the performance of modeling and monitoring. In addition, these deterministic methods do not perform well when faced with uncertainty and disturbances, which is common in industrial processes.

Compared to deterministic methods, probabilistic methods have attracted significant attentions in dealing with missing data problem [20], [21]. One popular probabilistic method to deal with missing data is the EM algorithm, which replaces the missing values by the posterior expectation given the existing observations [22], [23]. Another effective method is based on probabilistic latent variable model, by approximating the real likelihood using incomplete likelihood, the data latent feature can reflect the relationship among variables in the missing data set. For example, Zhou *et al.* [24] constructed a multi-rate probabilistic principal component regression model by integrating the cross correlation between multi-rate variable, resulting in the capable of quality-prediction with the non-randomly missing data. Similarly, Ge & Zhou [25], [26] proposed a semi-supervised probability latent variable regression model to deal with insufficient quality labels. Zheng *et al.* [27] proposed a semi-supervised probabilistic partial least squares regression model for virtual sensing to accommodate for missing quality data. Despite the research progress, most of aforementioned methods are built under the assumption that the data samples are i.i.d. Such assumption, however, is not valid in many industrial processes exhibiting dynamic characteristics. For dynamic processes, it is necessary to consider autocorrelation among data samples. A popular dynamic model structure is the Linear Dynamic Systems (LDS) model [28]–[30]. The LDS can be modified to accommodate missing values as well as multi-sampling rate data using Kalman filtering [31], [32]. In addition, Cong *et al.* [33] proposed a dynamic multi-sampling rate linear Gaussian state space model to handle data samples with three sampling rates, whose parameters are estimated using the EM algorithm. Also, Ref. [34] adopted a similar idea to effectively improve the accuracy of virtual sensor when there is missing data in the output observations.

The problem of the LDS methods in handling missing data is that they require prior knowledge about data missing patterns. However, since the time and frequency of missing data cannot be predicted, i.e., the missing pattern can be completely at random, the LDS based methods may not be useful in practice. In addition, most of the work in literature assumes the parameters of LDS model are time-invariant when faced with missing data, which may not be practical in real applications. In order to overcome these difficulties, this paper develops a virtual sensor for time-varying dynamic system based on the LDS model, which can adapt to data with unknown missing patterns. The latent variable structure of the improved LDS model enables the virtual sensor to capture the information in the missing data by considering both spatial and temporal information. That is to say, the information in the missing data can be learned from those not missing in the samples as well as the Markov process in the temporal stream. Considering of both spatial and temporal information renders more flexibility to the developed virtual sensor, so that it can easily cope with missing data.

The rest of this article is arranged as follows. Preliminaries are given in Section 2 where the missing data characteristics and LDS are briefly introduced. In Section 3, the proposed model will be mainly discussed. The performance of relevant methods is validated through two cases in Section 4. Finally, some conclusions are made in Section 5.

II. PRELIMINARIES

In this section, the characteristics of the irregular missing data are introduced in subsection A, and a brief review of linear dynamic systems is then given in subsection B.

A. THE PATTERNS AND CHARACTERISTICS OF MISSING DATA

Missing values usually occur in the data which is caused by the abnormalities in measuring equipment, data transmission, and storage. Among them, the irregular missing data in the form of Fig.1 is the most common missing type in actual industry. The main characteristics can be summarized as follows.

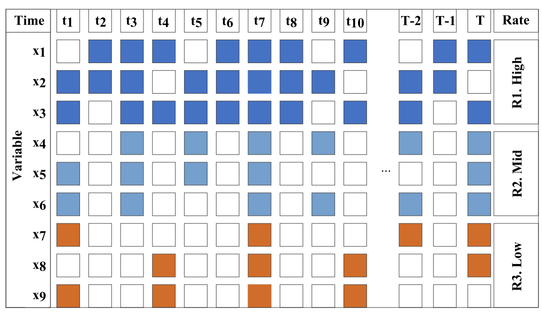


FIGURE 1. **The schematic of irregular data missing patterns (Different colored squares represent variables with different sampling rates where the white squares represent missing values).**

1) *Irregular Missing* – Random missing value can be found in almost every single variable (row) and sample (column), which makes it difficult to describe the behavior of the process using traditional methods.
2) *Data Objective Existence* – The missing data caused by collection errors truly exists, which means the variable-wise and sample-wise relationship has not actually been interrupted by the data missing problem.
3) *Information Imbalance* – As important indications for process safety, most quality variables can only be

obtained through laboratory analysis and complex calculation, which increase the level of information imbalance between process variables and quality variables.

According to the characteristics of irregular missing data, the maximum possible data inherent characteristics should be extracted via the existing measurements while reducing information loss during data modeling. For convenience, the data missing situation can be expressed by a state indicator $\varpi_{r,c}$ which can indicate the blank in row r, column c by simply setting the corresponding element $\varpi_{r,c}$ to 0, otherwise 1.

B. LINEAR DYNAMIC SYSTEM (LDS)

LDS is a typical probabilistic dynamic model, which can deal with the dynamic and uncertain characteristics of process simultaneously [35]. The model structure of LDS is shown in Eq 1. The feature latent variables z_t are used to reflect the main information of the measurement data, which are also linked by a first-order Markov chain to characterize the process dynamics. For virtual sensing modeling, the measurement sample g_t can contain both process variables (input variables x_t) and quality variables (output variables y_t) to obtain the relationship between the both. More about virtual sensing applications with LDS can be found in [30].

$$\begin{aligned} z_t &= \Lambda z_{t-1} + \Gamma \\ g_t &= W z_t + \Omega \end{aligned} \tag{1}$$

Furthermore, the main mathematical symbols appearing in this paper are defined as follows.

- z_t represent the i-dimensional latent feature vector.
- g_t represent the sample vector $\Re^{((N+M)\times 1)}$.
- G represents the original data matrix $\Re^{((N+M)\times T)}$, where the row represents the variable dimension, the column represents the sample dimension.
- N, M represent the number of process variables and quality variables respectively.
- μ_π, Σ_π represent the initial probability distribution of the latent variable, $z_{t=1} \sim \mathcal{N}(\mu_\pi, \Sigma_\pi)$.
- μ_t, V_t represent the forward estimate of the latent variable probability distribution, $z_t \sim \mathcal{N}(\mu_t, V_t)$.
- $\hat{\mu}_t, \hat{V}_t$ represent the backward estimate of the latent variable probability distribution, $z_t \sim \mathcal{N}\left(\hat{\mu}_t, \hat{V}_t\right)$.
- Λ represents the state transition matrix $\Re^{(i\times i)}$, which reflects the dynamic transfer relationship between the latent variables.
- W represents the observation matrix $\Re^{((M+N)\times i)}$, reflecting the linear conversion relationship between measurement and feature.
- Γ, Ω represent the model measurement noise.
- Q, R represent the covariance matrix of the noise term, denote as $\Gamma \sim \mathcal{N}(0, Q)$ and $\Omega \sim \mathcal{N}(0, R)$.
- $*, \varpi_{r,c}$ represent time-varying tag and missing status indicator respectively.

III. THE PROPOSED MODEL CONSTRUCTION

In this section, the construction process of the proposed model is introduced in detail, which is followed by the model parameter solution and some further discussions. Finally, a virtual sensing technique based on the model is developed.

A. LINEAR DYNAMIC TIME-VARYING PARAMETER STRUCTURE (LDVPS)

During the LDS modeling, all variables and their cross correlation can be obtained throughout the process, which is not available within incomplete data set. The invariant parameter structure of LDS seems impossible to transfer remaining variable information in different sampling time to the common latent variable simultaneously. To address this problem, the linear transformation relationship between each measurement variable and the common latent variable should be established separately, so that the irregular missing data and its remaining variable information are fully reflected in a common feature space. Furthermore, in such a structure, a first-order Markov chain is designed between the adjacent latent variables to describe the process dynamic. The proposed model can be expressed as:

$$\begin{aligned} z_t &= \Lambda z_{t-1} + \Gamma \\ x_t^{(n)} &= \alpha_n z_t + \varepsilon_n \\ y_t^{(m)} &= \beta_m z_t + \xi_m \end{aligned} \tag{2}$$

where $n = 1, 2, \ldots, N$ and $m = 1, 2, \ldots, M$ enumerate all measured entries that may be involved at each sampling moment. $\alpha_n \in \Re^{(1\times i)}$ and $\beta_m \in \Re^{(1\times i)}$ denote the loadings of process variables $x_t^{(n)}$ and quality variables $y_t^{(m)}$ respectively. $\varepsilon_n \sim \mathcal{N}(0, \delta_n)$ and $\xi_m \sim \mathcal{N}(0, \sigma_m)$ stand for measurement noise of single variable where δ_n and σ_m are the variance. When dealing with irregular missing data, the measured variables $x_t^{(n)}$ or $y_t^{(m)}$ may be null in each sampling sample. Therefore, the parameter $\{\alpha_n, \delta_n, \beta_m, \sigma_m\}$ and noise item corresponding to the missing entries can be removed from the structure to establish the relationship between remaining variables and the common latent variable. For distinguish, the incomplete sample can be denoted as g_t^*, and the remaining parameters in the sampling time t are incorporated as the observation matrix W_t^* and the interference covariance matrix R_t^*, respectively. Actually, the missing status of variable change with sampling time, which means that the settings of W_t^* and R_t^* are related to time. With the assumption that all variables obey Gauss, the probability evolution formula of the model can be abbreviated as:

$$\begin{cases} p\left(z_t \mid z_{t-1}\right) = \mathcal{N}\left(\Lambda z_{t-1}, Q\right) \\ p\left(g_t^* \mid z_t\right) = \mathcal{N}\left(W_t^* z_t, R_t^*\right) \end{cases} \tag{3}$$

The joint logarithmic likelihood function of the proposed model can be expressed as:

$$\begin{aligned}
&\ln p\left(g_1^* g_2^* \cdots g_t^*, z_1 z_2 \cdots z_t\right) \\
&\quad = \ln\left\{p\left(z_1\right)\prod_{t=2}^{T} p\left(g_t^* \middle| z_t\right) p\left(z_t \middle| z_{t-1}\right) p\left(z_{t-1}\right)\right\} \\
&\quad = \ln p\left(z_1\right) + \sum_{t=2}^{T} \ln p\left(z_t | z_{t-1}\right) + \sum_{t=1}^{T} \ln p\left(g_t^* | z_t\right) \qquad (4)
\end{aligned}$$

Contrast to complete likelihood term $\sum_{t=1}^{T} \ln p\left(g_t | z_t\right)$, $\sum_{t=1}^{T} \ln p\left(g_t^* | z_t\right)$ needs to ensure consistency of information in both the common feature space and the sample space with different measurement conditions simultaneously, which requires the learning of model parameters can reflect the constraint relationship among variables even if some values are missing. In other words, there is common information among different measurement samples, i.e., the cross correlation of remaining variables, which can be fully described by the latent feature variable after obtaining the optimal parameter solution. In this paper, the EM algorithm is utilized to learn the proposed model parameters.

B. MODEL PARAMETER SOLUTION USING EM

The EM algorithm is an efficient optimization strategy for latent variable estimation and parameter learning, which operates E-step and M-step iteratively [33].

E-Step: Deriving the expectation of the log-likelihood function with respect to the latent variable. In the general LDS modeling process, the posterior distribution of latent variable $p\left(z_t \middle| g_{t=1:T}^*\right) = \mathcal{N}\left(\mu_t, V_t\right)$ is usually estimated by the Kalman filtering algorithm combined with current observations and parameters. As for missing data, the proposed model tries to modify the traditional Kalman forward filtering algorithm to ensure that the filtering parameters are synchronized with the time-varying model parameters.

$$\mu_t = \Lambda \mu_{t-1} + K_t\left(g_t^* - W_t^* \Lambda \mu_{t-1}\right) \qquad (5)$$

$$V_t = \left(I - K_t W_t^*\right)\left(\Lambda V_{t-1} \Lambda^T + Q\right) \qquad (6)$$

$$\begin{aligned}
K_t &= \left(\Lambda V_{t-1} \Lambda^T + Q\right)\left(W_t^*\right)^T \\
&\quad \times \left[W_t^*\left(\Lambda V_{t-1} \Lambda^T + Q\right)\left(W_t^*\right)^T + R_t^*\right]^{-1} \qquad (7)
\end{aligned}$$

Especially, the initial distribution is calculated as:

$$\begin{aligned}
\mu_1 &= \mu_\pi + K_1\left(g_1^* - W_1^* \mu_\pi\right) \\
V_1 &= \Sigma_\pi - K_1 W_1^* \Sigma_\pi \\
K_1 &= \Sigma_\pi\left(W_1^*\right)^T\left[W_1^* \Sigma_\pi\left(W_1^*\right)^T + R_1^*\right]^{-1} \qquad (8)
\end{aligned}$$

It should be mentioned that the distribution of latent variables need to be optimized by the following Kalman backward smoothing algorithm which already proves itself in missing data treatment [35].

$$\hat{\mu}_T = \mu_T; \hat{V}_T = V_T \qquad (9)$$

$$\hat{\mu}_t = \mu_t + V_t \Lambda^T\left(\Lambda V_t \Lambda^T + Q\right)^{-1}\left(\hat{\mu}_{t+1} - \Lambda \mu_t\right) \qquad (10)$$

$$\begin{aligned}
\hat{V}_t &= V_t + V_t \Lambda^T\left(\Lambda V_t \Lambda^T + Q\right)^{-1} \\
&\quad \times \left(\hat{V}_{t+1} - \left(\Lambda V_t \Lambda^T + Q\right)\right) \\
&\quad \times \left(V_t \Lambda^T\left(\Lambda V_t \Lambda^T + Q\right)^{-1}\right)^T \qquad (11)
\end{aligned}$$

in which $\hat{\mu}_t$ and $\hat{V}_t$ are the ultimate posterior distributions of latent variables, and now the irregularly missing data is successfully mapped into a latent sequence space with the same dimension. Consequently, the expectation representation of latent variables can be derived as.

$$E\left\langle z_t | g_{1:T}^*\right\rangle = \hat{\mu}_t \qquad (12)$$

$$E\left\langle z_t z_t^T | g_{1:T}^*\right\rangle = \hat{V}_t + \hat{\mu}_t \hat{\mu}_t^T \qquad (13)$$

$$\begin{aligned}
E\left\langle z_t z_{t-1}^T | g_{1:T}^*\right\rangle &= \hat{V}_t\left(\Lambda V_{t-1} \Lambda^T + Q\right)^{-1} \Lambda V_{t-1} \\
&\quad + \hat{\mu}_t \hat{\mu}_{t-1}^T \qquad (14)
\end{aligned}$$

Finally, the expectation of the logarithmic likelihood function with respect to the latent variable is expressed as.

$$\begin{aligned}
&E_{z_t|g_t^*}\left\langle \ln p\left(g_1^* g_2^* \cdots g_t^*, z_1 z_2 \cdots z_t\right)\right\rangle \\
&\quad = E_{z_1|g_1^*}\left\langle \ln p\left(z_1\right)\right\rangle + \sum_{t=2}^{T} E_{z_t|g_t^*}\left\langle \ln p\left(z_t | z_{t-1}\right)\right\rangle \\
&\qquad + \sum_{t=1}^{T} E_{z_t|g_t^*}\left\langle \ln p\left(g_t^* | z_t\right)\right\rangle \qquad (15)
\end{aligned}$$

M-Step: Setting the partial derivative of the likelihood expectation with respect to each parameter to zero to get updated value of new parameters. To this end, substitute Eq 3 into Eq 15, and then expand.

$$\begin{aligned}
&\partial E_{z_1|g_1^*}\left\langle \ln p\left(z_1\right)\right\rangle \Big|_{\mu_\pi^{new}, \Sigma_\pi^{new}} \\
&\quad = -\frac{1}{2} \ln\left|\Sigma_\pi\right| \\
&\qquad - \frac{1}{2} tr\left(E\left\langle z_1 z_1^T\right\rangle \Sigma_\pi^{-1}\right) + E\left\langle z_1^T\right\rangle \Sigma_\pi^{-1} \mu_\pi - \mu_\pi^T \Sigma_\pi^{-1} \mu_\pi \qquad (16)
\end{aligned}$$

$$\begin{aligned}
&\partial E_{z_t|g_t^*}\left\langle \ln p\left(z_t | z_{t-1}\right)\right\rangle \Big|_{\Lambda^{new}, Q^{new}} \\
&\quad = -\frac{T-1}{2} \ln|Q| \\
&\qquad - \frac{1}{2} \sum_{t=2}^{T}\left\{\begin{array}{l} tr\left(E\left\langle z_t z_t^T\right\rangle Q^{-1}\right) - 2tr\left(E\left\langle z_t z_{t-1}^T\right\rangle \Lambda^T Q^{-1}\right) \\ + tr\left(E\left\langle z_{t-1} z_{t-1}^T\right\rangle \Lambda^T Q^{-1} \Lambda\right) \end{array}\right\} \qquad (17)
\end{aligned}$$

$$\begin{aligned}
&\partial E_{z_t|g_t^*}\left\langle \ln p\left(g_t^*|z_t\right)\right\rangle \Big|_{\alpha_{n=1:N}^{new},\beta_{m=1:M}^{new},\delta_{n=1:N}^{new},\sigma_{m=1:M}^{new}} \\
&\quad = -\frac{T}{2}\ln\left|R_t^*\right| \\
&\qquad -\frac{1}{2}\sum_{t=1}^{T}\left\{\begin{array}{l} tr\left(g_t^*\left(g_t^*\right)^T\left(R_t^*\right)^{-1}\right) \\ -2tr\left(g_t^* E\left\langle z_t\right\rangle\left(W_t^*\right)^T\left(R_t^*\right)^{-1}\right) \\ +tr\left(E\left\langle z_t z_t^T\right\rangle\left(W_t^*\right)^T\left(R_t^*\right)^{-1}W_t^*\right)\end{array}\right\}
\end{aligned} \tag{18}$$

Set the Eq 16 - Eq 18 to 0 to obtain the model parameter update expression.

$$\begin{aligned}
u_\pi^{new} &= E\left\langle z_1\right\rangle \\
\Sigma_\pi^{new} &= E\left\langle z_1 z_1^T\right\rangle - E\left\langle z_1\right\rangle E\left\langle z_1^T\right\rangle
\end{aligned} \tag{19}$$

$$\begin{aligned}
\Lambda^{new} &= \left(\sum_{t=2}^{T}E\left\langle z_t z_{t-1}^T\right\rangle\right)\left(\sum_{t=2}^{T}E\left\langle z_{t-1}z_{t-1}^T\right\rangle\right)^{-1} \\
Q^{new} &= \frac{1}{T-1}\sum_{t=2}^{T}\left\{\begin{array}{c} E\left\langle z_t z_t^T\right\rangle - \Lambda^{new}E\left\langle z_{t-1}z_t^T\right\rangle \\ -E\left\langle z_t z_{t-1}^T\right\rangle\left(\Lambda^{new}\right)^T \\ +\Lambda^{new}E\left\langle z_{t-1}z_{t-1}^T\right\rangle\left(\Lambda^{new}\right)^T\end{array}\right\}
\end{aligned} \tag{20}$$

$$\begin{aligned}
\alpha_n^{new} &= \left(\sum_{t=1}^{T}\varpi_{n,t}\cdot x_t^{(n)}\cdot E\left\langle z_t^T\right\rangle\right) \\
&\quad\times\left(\sum_{t=1}^{T}\varpi_{n,t}\cdot E\left\langle z_t z_t^T\right\rangle\right)^{-1} \\
\beta_m^{new} &= \left(\sum_{t=1}^{T}\varpi_{(N+m),t}\cdot y_t^{(m)}\cdot E\left\langle z_t^T\right\rangle\right) \\
&\quad\times\left(\sum_{t=1}^{T}\varpi_{(N+m),t}\cdot E\left\langle z_t z_t^T\right\rangle\right)^{-1}
\end{aligned} \tag{21}$$

$$\begin{aligned}
\delta_n^{new} &= \frac{1}{\sum_{t=1}^{T}\varpi_{n,t}} \\
&\quad\times\sum_{t=1}^{T}\varpi_{n,t}\left\{\begin{array}{l} x_t^{(n)^2} - \alpha_n^{new}\cdot E\left\langle z_t\right\rangle\cdot x_t^{(n)} \\ -x_t^{(n)}\cdot E\left\langle z_t^T\right\rangle\cdot\left(\alpha_n^{new}\right)^T \\ +\alpha_n^{new}\cdot E\left\langle z_t z_t^T\right\rangle\cdot\left(\alpha_n^{new}\right)^T\end{array}\right\} \\
\sigma_m^{new} &= \frac{1}{\sum_{t=1}^{T}\varpi_{(N+m),t}} \\
&\quad\times\sum_{t=1}^{T}\varpi_{(N+m),t}\left\{\begin{array}{l} y_t^{(m)^2} - \beta_m^{new}\cdot E\left\langle z_t\right\rangle\cdot y_t^{(m)} \\ -y_t^{(m)}\cdot E\left\langle z_t^T\right\rangle\cdot\left(\beta_m^{new}\right)^T \\ +\beta_m^{new}\cdot E\left\langle z_t z_t^T\right\rangle\cdot\left(\beta_m^{new}\right)^T\end{array}\right\}
\end{aligned} \tag{22}$$

It is worth noting that although $\alpha_n, \beta_m, \delta_n, \sigma_m$ are calculated separately, the learning of each model parameter is influenced by other variables. Taking α_n as an example, the terms $E\left\langle z_t^T\right\rangle$ and $E\left\langle z_t z_t^T\right\rangle$ correspond to the posterior expectations of the latent variables in E step. As mentioned in E step, the common latent variable is shared by non-missing variables at each sampling moment, which means that the relationship among the remaining variables can be retained and accumulated over the whole sampling time. Therefore, during repeated iterations of E step and M step, i.e., the model parameters learning process, the information of missing data can be extracted from the existing variable to the greatest extent.

C. SOME DISCUSSION OF THE PROPOSED MODEL

In this sub-section, several characteristics of the proposed model are further discussed, followed by the graphical explanation of latent information extraction within missing data from two aspects of spatial and temporal.

Remark 1: When all matrix elements of $\varpi_{r,c}$ are set to "1", the proposed model can be simplified as a traditional dynamic model built on the complete data set. While the quality-related elements in $\varpi_{r,c}$ are partially set to "0", it is equivalent to a semi-supervised method. If all elements in $\varpi_{r,c}$ are orderly set to "0", it can be served as a multi-sampling rate model. As a consequence of the above, it can be seen that the proposed model has strong adapt abilities to different patterns of missing data.

Remark 2: On the basis of the Markov chain in the proposed model, the dynamic trend of process can be extracted to help predict the latent information of missing data set even though measurements are seriously uncollected.

Remark 3: From the existing variables and dynamic prediction, the latent information of the original data matrix can be restored to the greatest extent in feature space.

To better explain the mechanism of the remark 3, the visualization of E step of the proposed model is given in Fig.2. By using time-varying Kalman filtering algorithm, the existing collections will be transmitted into corresponding latent feature information in a posteriori form, which can be further combined with the dynamic prediction information of latent variables to achieve the maximum utilization of the incomplete data. As shown in Fig.2(a), each sample is mapped into the low-dimensional space to assist the estimation of latent variable distribution $\mathcal{N}\left(z_t\left|g_t^*, W_t^*, R_t^*\right.\right)$. Simultaneously, in the low dimensional space, assuming that the distribution of latent variable at the previous time $t-1$ is known, that is:

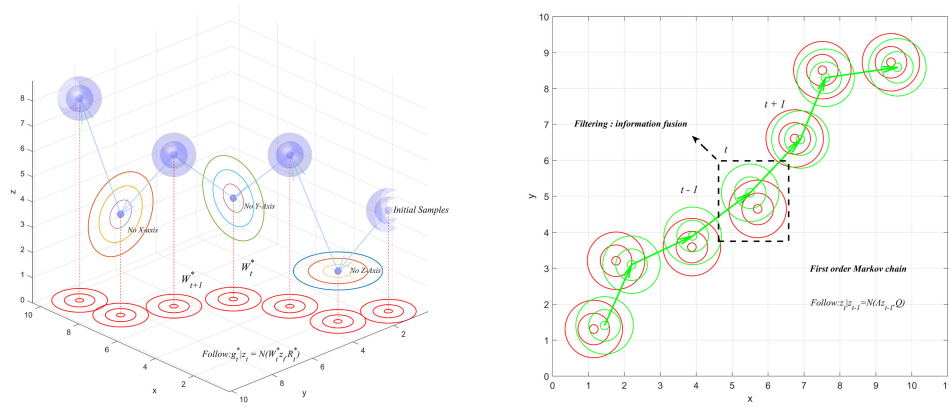
$$\begin{aligned}
&z_{t-1}|\, g_1^*, g_2^*, \cdots, g_{t-1}^* \\
&\quad = \underbrace{E\left\langle z_{t-1}\right\rangle}_{\mu_{t-1}} + \Delta z_{t-1};\ \Delta z_{t-1}\sim\mathcal{N}\left(0, V_{t-1}\right)
\end{aligned} \tag{23}$$

According to the model definition in Eq 2, the following results can be obtained:

$$\underbrace{z_t|\, g_1^*, g_2^*, \cdots, g_{t-1}^*}_{prediction_t} = \Lambda z_{t-1} + \Gamma \tag{24}$$

Substitute Eq 23 into Eq 24:

$$\begin{aligned}
\underbrace{z_t|\, g_1^*, g_2^*, \cdots, g_{t-1}^*}_{prediction_t} &= \Lambda\left[\mu_{t-1} + \Delta z_{t-1}\right] + \Gamma \\
&= \Lambda\mu_{t-1} + \Lambda\Delta z_{t-1} + \Gamma
\end{aligned} \tag{25}$$



(a) (b)

FIGURE 2. Visualization of the E step of the proposed model. In Fig.2(a), the high dimensional space "x-y-z" represents observation space where the blue spheres denote the distribution diagrams of the complete measurement samples and the planetary orbit diagrams denote the measured samples with missing variables. The low dimensional space "x-y" is the latent feature space, in which the red circles denote the projection from the measured samples, as shown in Fig.2(b). The green circles denote the dynamic prediction of the latent variables and the green arrow lines represent the Markov chain which reflects the process dynamic trend.

On the basis of the above equations, the predicted distribution of dynamic latent variable at sampling time t can be derived:

$$\begin{aligned}
&E\left\langle z_t|\, g_1^*, g_2^*, \cdots, g_{t-1}^*\right\rangle \\
&\quad = E\left\langle \Lambda\mu_{t-1}\right\rangle + E\left\langle \Lambda\Delta z_{t-1}\right\rangle + E\left\langle \Gamma\right\rangle \\
&\quad = \Lambda\mu_{t-1} \\
&Cov\left\langle z_t|\, g_1^*, g_2^*, \cdots, g_{t-1}^*\right\rangle \\
&\quad = E\left\langle \begin{array}{l} \left[\left(z_t|\, g_1^*, g_2^*, \cdots, g_{t-1}^*\right) - \Lambda\mu_{t-1}\right] \\ \times\left[\left(z_t|\, g_1^*, g_2^*, \cdots, g_{t-1}^*\right) - \Lambda\mu_{t-1}\right]^T \end{array}\right\rangle \\
&\quad = E\left\langle \left[\Lambda\Delta z_{t-1} + \Gamma\right]\cdot\left[\Lambda\Delta z_{t-1} + \Gamma\right]^T\right\rangle \\
&\quad = \Lambda V_{t-1}\Lambda^T + Q
\end{aligned} \tag{26}$$

Consequently, the latent feature information at sampling time t can be described using two different aspects, including the projection from observation space in Fig.2(a) and the prediction from latent sequence space in Fig.2(b). Then, the fusion results of the two types of information can be regarded as a beneficial complement for latent information construction, as shown in the dotted box in Fig.2(b). The final distribution expressions of the latent variable after the combination can be expressed in Eq 5 - Eq 8, and the specific derivation is shown in the Appendix.

D. VIRTUAL SENSOR APPLICATIONS

As mentioned in the previous sub-sections, the inherent characteristics of the proposed model make it an efficient approach for missing data treatment, which can be naturally employed for virtual sensing in irregular data missing situations. The detailed virtual sensor process is depicted in Fig.3. When a query sample $X_{k=1,2,\ldots}^{query}$ enters, its latent variable can be calculated as follows:

$$\mu_k = \widehat{\Lambda}\mu_{k-1} + K_k\left(X_k^{query} - \widehat{W}_\alpha^*\widehat{\Lambda}\mu_{k-1}\right)$$

$$here: \widehat{W}_\alpha^* = \left[\widehat{\alpha}_1^T, \cdots, \widehat{\alpha}_n^T, \cdots, \widehat{\alpha}_N^T\right]^T \tag{27}$$

$$V_k = \left(I - K_k\widehat{W}_\alpha^*\right)\left(\widehat{\Lambda}V_{k-1}\widehat{\Lambda}^T + \widehat{Q}\right) \tag{28}$$

$$\begin{aligned}
K_k &= \left(\widehat{\Lambda}V_{k-1}\widehat{\Lambda}^T + \widehat{Q}\right)\times\left(\widehat{W}_\alpha^*\right)^T \\
&\quad \times\left[\begin{array}{l} diag\left\{\widehat{\delta}_1, \cdots, \widehat{\delta}_n, \cdots, \widehat{\delta}_N\right\} + \widehat{W}_\alpha^* \\ \quad\times\left(\widehat{\Lambda}V_{k-1}\widehat{\Lambda}^T + \widehat{Q}\right)\left(\widehat{W}_\alpha^*\right)^T \end{array}\right]^{-1}
\end{aligned} \tag{29}$$

in which $\left\{\widehat{\Lambda}, \widehat{Q}, \widehat{\alpha}_{n=1:N}, \widehat{\beta}_{m=1:M}, \widehat{\delta}_{n=1:N}\right\}$ represents the update result in the ultimate M step. Then, the prediction can be obtained.

$$Y_k^{pre} = \left[\widehat{\beta}_1^T, \cdots, \widehat{\beta}_m^T, \cdots, \widehat{\beta}_M^T\right]^T \times u_k \tag{30}$$

IV. CASE STUDY

In this section, the proposed model LDVPS is utilized for quality prediction in two industrial data missing cases, including a Debutanizer column and Sulfur Recovery Unit. The superiority of LDVPS in virtual sensing applications is verified and two missing-data treatment models are also introduced for comparison, namely semi-supervised probabilistic latent variable regression (SSPLVR) and dynamic probabilistic latent variable model (LDVPS) [30].

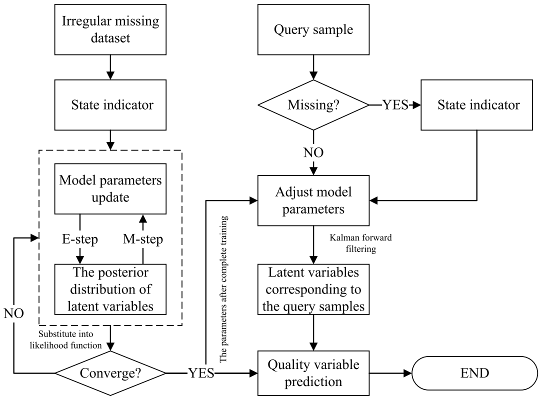


FIGURE 3. Flow chart for virtual sensing modeling.

A. DEBUTANIZER COLUMN

The debutanizer column is an important part of the desulfurization and naphtha splitter plant that is widely used to evaluate the process monitoring performance among different virtual sensing techniques [25]. The main purpose of this process is to remove butane from the product and effective monitoring of butane content is fully essential. For the purpose of virtual sensors modeling, 7 process variables have been collected from the process shown in Fig.4. These process variables are closely related to butane content and listed in Table 1. 1000 samples in the process have been collected as the training data set, and the missing mechanism of dataset is set using state indicator. Meanwhile, another 1000 samples are collected as test data under the same working condition. As a feasible reference scheme, it is assumed that the test samples are complete, which can compare the models' tolerance for training data missing. In addition, the cumulative variance contribution rate has been employed to select the dimension of latent variables before implementing virtual sensors applications of SSPLVR, DPLVM and LDVPS, as shown in Fig.5.

TABLE 1. The variables of the virtual sensor in the debutanizer column.

Input variable	Variable description
x1 (U1)	Top temperature
x2 (U2)	Top pressure
x3 (U3)	Reflux flow
x4 (U4)	Flow to next process
x5 (U5)	6th tray temperature
x6 (U6)	Bottom temperature
x7 (U7)	Bottom pressure

In the aspect of model performance evaluation, mean square error (MSE) and goodness of fit (R^2) are usually adopted as regression evaluation indexes. Table 2 shows the index value trends of the three virtual sensors under different variable missing rates where the smaller MSE or the larger the R^2 denote the higher prediction accuracy of virtual-sensor model. To guarantee reliability, every index values in Table 2 are the average results of 100 experiments, and the missing positions of variables in each experiment are randomly reset. It can be seen that the prediction performance of SSPLVR and DPLVM models decreases with the missing rate rising from 1% to 15%, while LDVPS can maintain a good prediction performance steadily. This means that the LDVPS model is less affected by missing data and can provide more robust virtual-sensor modeling technology for missing data treatment. Taking 5% and 15% missing rate as examples, the predictive effect of the three models on butane content is shown in Fig.6 and Fig.7. In these two figures, both LDVPS and DPLVM models have better prediction results than SSPLVR models due to their consideration of process

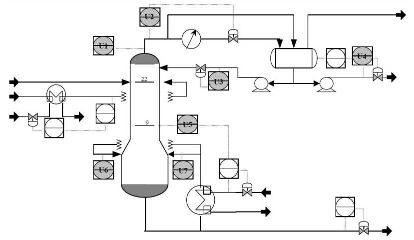


FIGURE 4. Flow chart for virtual sensor modeling.

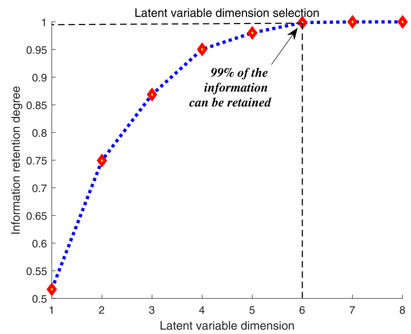


FIGURE 5. The cumulative variance contribution plot.

TABLE 2. The index values of three virtual sensor methods under different variable missing rate.

Miss rate	SSPLVR MSE	SSPLVR R^2	DPLVM MSE	DPLVM R^2	LDVPS MSE	LDVPS R^2
1%	0.829	0.161	0.225	0.775	**0.204**	**0.796**
2%	0.827	0.172	0.253	0.747	**0.200**	**0.800**
3%	0.915	0.084	0.300	0.700	**0.271**	**0.729**
4%	0.924	0.076	0.362	0.697	**0.255**	**0.755**
5%	0.950	0.049	0.357	0.643	**0.241**	**0.755**
6%	0.956	0.043	0.376	0.623	**0.285**	**0.715**
7%	0.982	0.017	0.378	0.621	**0.322**	**0.681**
8%	0.959	0.040	0.415	0.585	**0.313**	**0.687**
9%	0.979	0.020	0.450	0.550	**0.298**	**0.702**
10%	0.982	0.003	0.469	0.531	**0.308**	**0.692**
11%	0.997	0.002	0.508	0.491	**0.264**	**0.735**
12%	1.154	-0.155	0.637	0.363	**0.283**	**0.717**
13%	1.146	-0.147	0.677	0.322	**0.295**	**0.705**
14%	1.152	-0.153	0.809	0.190	**0.313**	**0.684**
15%	1.257	-0.166	0.834	0.165	**0.285**	**0.715**
16%	1.217	-0.218	0.870	0.129	**0.316**	**0.684**
17%	1.411	-0.350	0.881	0.119	**0.313**	**0.687**
18%	1.352	-0.353	0.948	0.051	**0.316**	**0.684**
19%	1.396	-0.398	0.978	0.021	**0.305**	**0.695**
20%	1.314	-0.316	1.235	-0.060	**0.319**	**0.682**

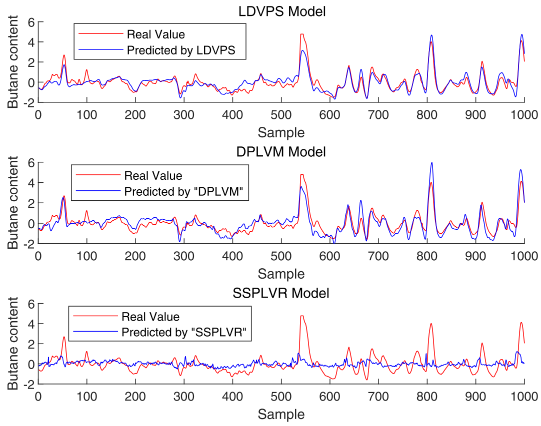


FIGURE 6. The prediction effect of three models on butane content when 5% variable missing rate.

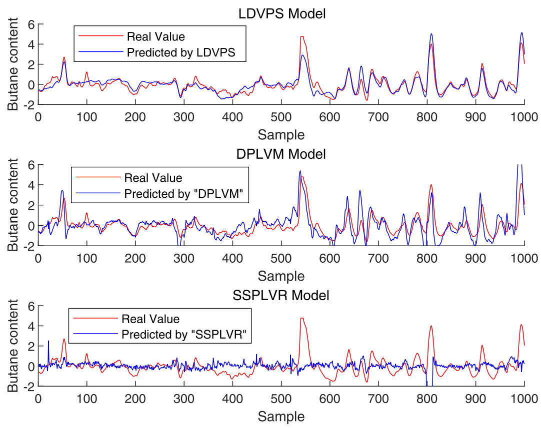


FIGURE 7. The prediction effect of three models on butane content when 15% variable missing rate.

dynamics. However, with the increase of missing rate, the performance gap between DPLVM and SSPLVR model gradually narrowed, and the advantages of LDVPS model became more and more prominent. There are two main reasons for this phenomenon. Firstly, the error caused by the removal of incomplete samples by DPLVM is acceptable when dealing with small-scale irregular missing data, as shown in Fig.6. But when the scale of irregular missing data increases, DPLVM will lose the entire sample even if the sample is missing only one variable, which is much higher cost than dealing with multi-sampling rate data or small-scale missing data. Worse still, DPLVM will not be sufficient to accurately extract the process dynamics along with the loss of mass data information and the destruction of data structures. It can be clearly seen from Fig.7 that the prediction curve of the DPLVM model is seriously deviated after 600s. Secondly, the negative effects of quality-related variables missing can be "immune" by SSPLVR semi-supervised modeling method, and the samples that missing process variables can be directly removed without considering the auto-correlation between them. As a result, although SSPLVR does not consider process dynamic, its utilization rate of data is much higher than that of DPLVM, and the predicted results are relatively stable.

On the contrary, LDVPS can establish a probabilistic dynamic model that is consistent with the characteristics of missing data without discarding any samples or part of variables, which can more thoroughly extract the cross-correlation between variables and the auto-correlation between samples. From the experimental results, LDVPS occupies an absolute advantage in the three models. Moreover, when the missing rate is as high as 50%, LDVPS can still maintain a high prediction accuracy, which is a capability that SSPLVR and DPLVM models do not have, as shown in Fig.8.

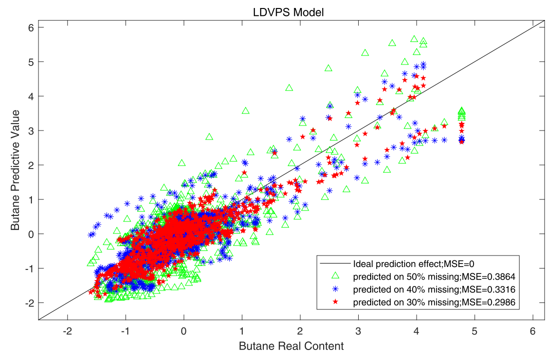


FIGURE 8. The prediction effect of LDVPS models on butane content when the variable missing rate is as high as 50%.

B. SULFUR RECOVERY UNIT (SRU)

SRU is an important green cleaning device in the oil refining process and is widely used in the treatment and recovery of sulfur-containing waste gases [36]. The simplified schematic diagram of SUR process is shown in Fig.9. Firstly, two waste gases, MEA (rich in H_2S) and SWS (rich in H_2S and NH_3), are sent into the waste heat boiler for combustion, and the air flow (oxygen content) is controlled by adjusting valve $S1$ to ensure the appropriate stoichiometric ratio. Next, the combustion products from the furnace are cooled by $C1 - C3$ to produce liquid sulfur which is increased by a subsequent catalytic reaction of $R1 - R3$. The specific chemical reaction equation of the process is shown in Eq 31. To monitor the performance of sulfur conversion process and control the air feed ratio, the tail gas analysis of residual H_2S and SO_2 gases is required. However, hardware sensors are often corroded by sulphur-containing gases and require regular cleaning and maintenance. Therefore, virtual sensing techniques for SRU processes need to be developed to achieve on-line prediction

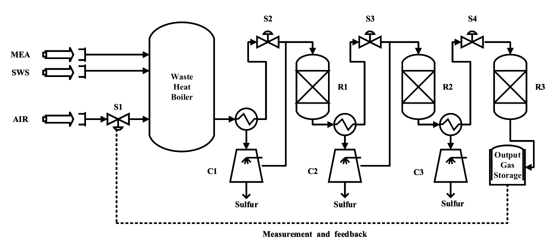


FIGURE 9. The flowchart of the Sulfur Recovery Unit.

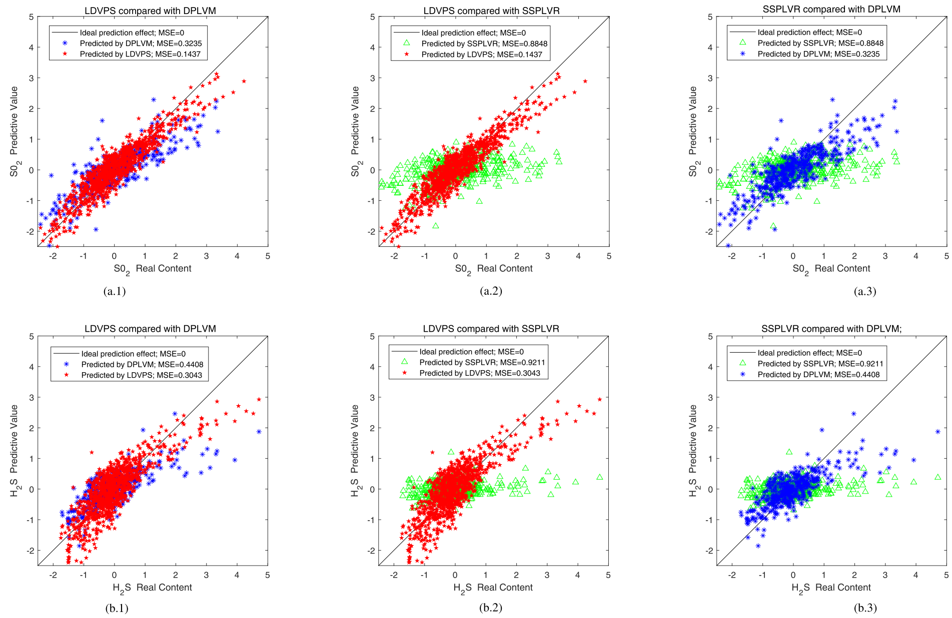


FIGURE 10. The prediction effect of three virtual sensor methods when the variable missing rate is 15%, (a.1-a.3) SO_2 content prediction; (b.1-b.3) H_2S content prediction.

of H_2S) and SO_2 gases. In this example, 5 process variables are selected as the inputs of virtual sensor modeling, and H_2S and SO_2 contents are output variables, as shown in Table 3.

$$\begin{aligned} &3H_2S + \frac{1}{2}O_2 \rightarrow SO_2 + 2H_2S + H_2O \\ &4NH_3 + 3O_2 \rightarrow 2N_2 + 6H_2O \\ &SO_2 + 2H_2S \rightarrow S_3 + 2H_2O \end{aligned} \tag{31}$$

TABLE 3. The variables of the virtual sensor in the debutanizer column.

Variable	Variable description
x1	The gas flow in MEA zone (MEA GAS)
x2	The air flow in MEA zone 1 (AIR MEA 1)
x3	The air flow in MEA zone (AIR MEA 2)
x4	The air flow in SWS zone (AIR SWS)
x5	The gas flow in SWS zone (SWS GAS)
y1	H_2S gas content in exhaust gas
y2	SO_2 gas content in exhaust gas

A total of 2000 sets of data samples in SRU process have been collected, and then the state indicator is used to label the variable missing cases. To further verify the virtual sensor performance of the proposed model, *MSE* and R^2 were employed again for the evaluation indexes of SSPLVR, DPLVM and LDVPS models. Table 4 and Fig.10 respectively show the prediction indexes and results of the three models when the variable missing rate is 15%. Because the LDVPS model has the highest utilization rate of data and can consider the dynamic process, its prediction results are better than the other two models. It is worth noting that the number of red dots representing the predicted result in Fig.10 is always more than the number of other colors. The main reason for this phenomenon is that SSPLVR and DPLVM models remove the test samples with missing variables, which makes it impossible to give the prediction results at the corresponding time. In contrast, LDVPS can give the predicted value of the test sample at each moment, and the predicted points are closer to the diagonal (highest prediction accuracy). In summary, the proposed model LDVPS is more suitable for handling irregular missing data and has higher prediction accuracy under the same situation.

TABLE 4. The prediction index values of three virtual sensor methods when the variable missing rate is 15%.

Index	SSPLVR	DPLVM	LDVPS
$MSE(SO_2)$	0.885	0.324	**0.144**
$R^2(SO_2)$	0.141	0.686	**0.856**
$MSE(H_2S)$	0.921	0.441	**0.304**
$R^2(H_2S)$	0.122	0.580	**0.694**

V. CONCLUSION

In this paper, a novel linear dynamic time-varying parameters structure for irregular missing data treatment is proposed and applied to virtual sensing modeling. Unlike traditional data deletion and filling methods, LDVPS can construct a probabilistic dynamic model that meets the characteristics of missing data without discarding any valuable data information or introducing additional noise. By setting time-varying parameters, the linear transformation relationship between the common latent variable and the remaining variables is established, so that the information of missing data can be captured from those not missing in the samples. In addition, the process dynamics can be reflected by the Markov chain between adjacent latent variables, which can provide a reasonable supplement for latent information prediction.

The EM algorithm is introduced to solve the model parameters where the E step is improved by the time-varying Karman filtering algorithm to estimate the information of latent variables of missing data set. Experiments show that the LDVPS model is more suitable for missing data treatment, which can maintain good stable prediction accuracy under large scale irregular missing data.

APPENDIX

E-step: Improved Karman forward filtering inference

The accurate estimation of the latent variable distribution is required.

$$\underbrace{p\left(z_t \middle| g_1^*, g_2^*, \cdots, g_t^*\right)}_{update_t} = \frac{p\left(g_t^* \middle| z_t, g_1^*, g_2^*, \cdots, g_{t-1}^*\right) \cdot p\left(z_t \middle| g_1^*, g_2^*, \cdots, g_{t-1}^*\right)}{p\left(g_1^*, g_2^*, \cdots, g_t^*\right)} \times \frac{p\left(g_1^*, g_2^*, \cdots, g_{t-1}^*\right)}{p\left(g_1^*, g_2^*, \cdots, g_t^*\right)} \tag{32}$$

in which $\frac{p(g_1^*, g_2^*, \cdots, g_{t-1}^*)}{p(g_1^*, g_2^*, \cdots, g_t^*)}$ tends to 1 in EM iteration process.

$$\underbrace{p\left(z_t \middle| g_1^*, g_2^*, \cdots, g_t^*\right)}_{update_t} = p\left(g_t^* \middle| z_t\right) \cdot \underbrace{p\left(z_t \middle| g_1^*, g_2^*, \cdots, g_{t-1}^*\right)}_{prediction_t} \tag{33}$$

where $\underbrace{p\left(z_t \middle| g_1^*, g_2^*, \cdots, g_{t-1}^*\right)}_{prediction_t}$ represents the current latent variable distribution prediction using all observations at the previous moment, that is, latent dynamic prediction:

$$\begin{aligned}\underbrace{p\left(z_t \middle| g_1^*, g_2^*, \cdots, g_{t-1}^*\right)}_{prediction_t} &= \int p\left(z_t, z_{t-1} \middle| g_1^*, g_2^*, \cdots, g_{t-1}^*\right) dz_{t-1} \\ &= \int p\left(z_t \middle| z_{t-1}, g_1^*, g_2^*, \cdots, g_{t-1}^*\right) \\ &\quad \times p\left(z_{t-1} \middle| g_1^*, g_2^*, \cdots, g_{t-1}^*\right) dz_{t-1} \\ &= \int p\left(z_t \middle| z_{t-1}\right) \underbrace{p\left(z_{t-1} \middle| g_1^*, g_2^*, \cdots, g_{t-1}^*\right)}_{update_{t-1}} dz_{t-1}\end{aligned} \tag{34}$$

Hence, we can get the probability distribution recursion formula of latent variables with respect to time:

$$\begin{aligned}\underbrace{p\left(z_t \middle| g_1^*, g_2^*, \cdots, g_t^*\right)}_{updata_t} &= p\left(g_t^* \middle| z_t\right) \cdot \int p\left(z_t \middle| z_{t-1}\right) \\ &\quad \times \underbrace{p\left(z_{t-1} \middle| g_1^*, g_2^*, \cdots, g_{t-1}^*\right)}_{update_{t-1}} dz_{t-1}\end{aligned} \tag{35}$$

equivalent to:

$$\begin{aligned}&\mathcal{N}\left(z_t \middle| \mu_t, V_t\right) \\ &= \mathcal{N}\left(g_t^* \middle| B_t^* z_t, R_t^*\right) \\ &\quad \times \int \mathcal{N}\left(z_t \middle| \Lambda z_{t-1}, Q\right) \mathcal{N}\left(z_{t-1} \middle| \mu_{t-1}, V_{t-1}\right) dz_{t-1}\end{aligned} \tag{36}$$

Finally, the Eq 5 - Eq7 are obtained via Eq 36.

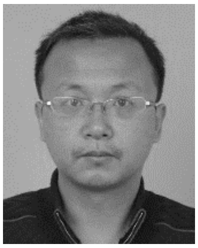
REFERENCES

[1] Q. Jiang, X. Yan, and B. Huang, "Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes," *Ind. Eng. Chem. Res.*, vol. 58, no. 29, pp. 12899–12912, Jul. 2019.
[2] Z. Ge, "Process data analytics via probabilistic latent variable models: A tutorial review," *Ind. Eng. Chem. Res.*, vol. 57, no. 38, pp. 12646–12661, Sep. 2018.
[3] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 10, pp. 3543–3562, Mar. 2013.
[4] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795–814, Apr. 2009.
[5] F. A. A. Souza, R. Araújo, and J. Mendes, "Review of soft sensor methods for regression applications," *Chemometric Intell. Lab. Syst.*, vol. 152, pp. 69–79, Mar. 2016.
[6] Y. Jiang, S. Yin, J. Dong, and O. Kaynak, "A review on soft sensors for monitoring, control and optimization of industrial processes," *IEEE Sensors J.*, early access, Oct. 22, 2020, doi: 10.1109/JSEN.2020.3033153.
[7] X. Yuan, Z. Ge, and Z. Song, "Locally weighted kernel principal component regression model for soft sensing of nonlinear time-variant processes," *Ind. Eng. Chem. Res.*, vol. 53, no. 35, pp. 13736–13749, Sep. 2014.
[8] W. Shao and X. Tian, "Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models," *Chem. Eng. Res. Des.*, vol. 95, pp. 113–132, Mar. 2015.
[9] J. C. B. Gonzaga, L. A. C. Meleiro, C. Kiang, and R. Maciel Filho, "ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process," *Comput. Chem. Eng.*, vol. 33, no. 1, pp. 43–49, Jan. 2009.
[10] Q. Sun and Z. Ge, "A survey on deep learning for data-driven soft sensors," *IEEE Trans. Ind. Informat.*, early access, Jan. 20, 2021, doi: 10.1109/TII.2021.3053128.
[11] J. L. Schafer and J. W. Graham, "Missing data: Our view of the state of the art," *Psychol. Methods*, vol. 7, no. 2, p. 147, Jun. 2002.
[12] J. Zhu, Z. Ge, Z. Song, and F. Gao, "Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data," *Annu. Rev. Control*, vol. 46, pp. 107–133, 2018.
[13] N. A. Lazar, "Statistical analysis with missing data," *Technometrics*, vol. 45, no. 4, pp. 364–365, Jan. 2002.
[14] A. Ilin and T. Raiko, "Practical approaches to principal component analysis in the presence of missing values," *J. Mach. Learn. Res.*, vol. 11, pp. 1957–2000, Aug. 2010.
[15] J. M. Jerez, I. Molina, P. J. García-Laencina, E. Alba, and L. Franco, "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," *Artif. Intell. Med.*, vol. 50, no. 2, pp. 105–115, 2010.
[16] Y. Masuda, H. Kaneko, and K. Funatsu, "Multivariate statistical process control method including soft sensors for both early and accurate fault detection," *Ind. Eng. Chem. Res.*, vol. 53, no. 20, pp. 8553–8564, 2014.
[17] Y. Deng, C. Han, J. Guo, and L. Sun, "Temporal and spatial nearest neighbor values based missing data imputation in wireless sensor networks," *Sensors*, vol. 21, no. 5, p. 1782, Mar. 2021.
[18] X. Luo, M. Zhou, Y. Xia, Q. Zhu, A. C. Ammari, and A. Alabdulwahab, "Generating highly accurate predictions for missing QoS data via aggregating nonnegative latent factor models," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 524–537, Mar. 2016.
[19] J. Lin, N. Li, M. A. Alam, and Y. Ma, "Data-driven missing data imputation in cluster monitoring system based on deep neural network," *Int. J. Speech Technol.*, vol. 50, no. 3, pp. 860–877, Mar. 2020.
[20] S. Khatibisepehr and B. Huang, "Dealing with irregular data in soft sensors: Bayesian method and comparative study," *Ind. Eng. Chem. Res.*, vol. 47, no. 22, pp. 8713–8723, 2008.

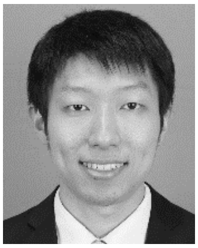
[21] S. Khatibisepehr, B. Huang, and S. Khare, "Design of inferential sensors in the process industry: A review of Bayesian methods," *J. Process Control*, vol. 23, no. 10, pp. 1575–1596, Nov. 2013.
[22] A. P. Dempster, "Maximum likelihood from incomplete data Via the *EM* algorithm," *J. Roy. Stat. Soc.: B (Methodol.)*, vol. 39, no. 1, pp. 1–22, 1977.
[23] R. Razavi-Far, B. Cheng, M. Saif, and M. Ahmadi, "Similarity-learning information-fusion schemes for missing data imputation," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104805.
[24] L. Zhou, Y. Wang, and Z. Ge, "Multi-rate principal component regression model for soft sensor application in industrial processes," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 230–232, Apr. 2020.
[25] Z. Ge and Z. Song, "Semisupervised Bayesian method for soft sensor modeling with unlabeled data samples," *AIChE J.*, vol. 57, no. 8, pp. 2109–2119, Aug. 2011.
[26] L. Zhou, J. Chen, Z. Song, and Z. Ge, "Semi-supervised PLVR models for process monitoring with unequal sample sizes of process variables and quality variables," *J. Process Control*, vol. 26, pp. 1–16, Feb. 2015.
[27] J. Zheng and Z. Song, "Semisupervised learning for probabilistic partial least squares regression model and soft sensor application," *J. Process Control*, vol. 64, pp. 123–131, Apr. 2018.
[28] Q. Wen, Z. Ge, and Z. Song, "Data-based linear Gaussian state-space model for dynamic process monitoring," *AIChE J.*, vol. 58, no. 12, pp. 3763–3776, Dec. 2012.
[29] Z. Ge and X. Chen, "Supervised linear dynamic system model for quality related fault detection in dynamic processes," *J. Process Control*, vol. 44, pp. 224–235, Aug. 2016.
[30] Z. Ge and X. Chen, "Dynamic probabilistic latent variable model for process data modeling and regression application," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 1, pp. 323–331, Jan. 2019.
[31] Y. Wu and X. Luo, "A novel calibration approach of soft sensor based on multirate data fusion technology," *J. Process Control*, vol. 20, no. 10, pp. 1252–1260, Dec. 2010.
[32] A. Fatehi and B. Huang, "Kalman filtering approach to multi-rate information fusion in the presence of irregular sampling rate and variable measurement delay," *J. Process Control*, vol. 53, pp. 15–25, May 2017.
[33] Y. Cong, L. Zhou, Z. Song, and Z. Ge, "Multirate dynamic process monitoring based on multirate linear Gaussian state-space model," *IEEE Trans. Automat. Sci. Eng.*, vol. 16, no. 4, pp. 1708–1719, Oct. 2019.
[34] H. Raghavan, A. K. Tangirala, R. Bhushan Gopaluni, and S. L. Shah, "Identification of chemical processes with irregular output sampling," *Control Eng. Pract.*, vol. 14, no. 5, pp. 467–480, May 2006.
[35] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York, NY, USA: Springer-Verlag, 2006.
[36] L. Fortuna, A. Rizzo, M. Sinatra, and M. G. Xibilia, "Soft analyzers for a sulfur recovery unit," *Control Eng. Pract.*, vol. 11, no. 12, pp. 1491–1500, Dec. 2003.

ZE YING is currently pursuing the master's degree with the School of Mechanical and Electrical Engineering, China Jiliang University. His research interests include probability modeling, fault detection, soft sensor, and their applications in industrial processes.

YUN WANG received the B.S. and M.S. degrees from the Zhejiang University of Technology, in 2010 and 2012, respectively. She is a Faculty Member with the Zhejiang Tongji Vocational College of Science and Technology, where she is currently a Lecturer with the Mechanical and Electrical Engineering Department. Her research interests include fault diagnosis and data analysis and their applications in industrial processes.



JUN WANG received the B.S. and M.S. degrees from Wuhan University, in 1991 and 2001, respectively, and the Ph.D. degree from the Huazhong University of Science and Technology, China, in 2013. He is a Faculty Member with China Jiliang University, where he is currently a Lecturer with the College of Mechanical and Electrical Engineering. His research interests include fuel flow, combustion, and solar power systems.



YUCHEN HE received the B.S. degree from the Zhejiang University of Technology, in 2010, and the Ph.D. degree from Zhejiang University, China, in 2017. He is a Faculty Member of China Jiliang University, where he is currently an Associate Professor with the College of Mechanical and Electrical Engineering. His research interests include multivariate statistical process monitoring, fault diagnosis, and data analysis and their applications in industrial processes.