IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 8 April 2023, accepted 9 May 2023, date of publication 12 May 2023, date of current version 18 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3275762

RESEARCH ARTICLE

A Quality-Related Concurrent Dual-Latent Variable Model and Its Semi-Supervised Soft Sensor Applications

YUN WANG[1], ZE YING[2], AND YUCHEN HE[3], (Member, IEEE)
[1]Department of Mechanical and Electrical Engineering, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou 311123, China
[2]College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
[3]Key Laboratory of Intelligent Manufacturing Quality Big Data Tracing and Analysis of Zhejiang Province, China Jiliang University, Hangzhou 310018, China

Corresponding author: Yuchen He (yche@cjlu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB3304100, 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 LY23F030004, in part by the Project of Department of Education of Zhejiang Province under Grant Y202249499, in part by the Youth Fund of the Zhejiang Tongji Vocational College of Science and Technology under Grant FRF22QN002, and in part by the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant 2021YW18 and Grant 2021YW80.

ABSTRACT To provide accurate key variable prediction, data-driven soft sensing techniques have extracted much attention in recent years. Due to different control strategies in industrial processes, it is noticed that the variables in the control loops can be autocorrelated while the others may be static, which needs to be considered simultaneously. In this paper, a quality-related concurrent dual-latent variable (CDLV) model is proposed for soft sensing construction. Two different kinds of latent variables are adopted to learn quality-related dynamic information and quality-related static information respectively. Both quality-related variables are then applied for quality prediction purposes. On this basis, the CDLV model is extended to a semi-supervised form to provide a comprehensive description for the soft sensor design with insufficient quality information. The proposed models are demonstrated by two industrial cases which show superiority over other relative methods in the accuracy of key quality variables prediction.

INDEX TERMS Concurrent dual-latent variable (CDLV), soft sensor, probabilistic latent variable model, quality-related information, semi-supervised learning.

I. INTRODUCTION

To realize precise process control in modern industrial systems, it is of great significance to obtain a reliable value of key variables that are closely related to product quality [1], [2], [3]. However, due to the increasing complexity of industrial environments, limitations of measurement equipment, and other economic considerations, some key quality variables cannot be measured properly or promptly. In the past decades, a large number of soft sensing techniques have been developed to analyze the relationship between hard-to-measure quality variables and easy-to-measure process variables [4], [5], [6], which mainly includes three types, namely mechanism-model-based, knowledge-based, and data-driven soft sensors [7], [8]. The accuracy of the first two types can be guaranteed only when adequate process mechanisms or prior expert knowledge is available, which is often time-consuming in practical situations. Meanwhile, due to the popularity of distributed control systems (DCS) and the internet of things (IoT), a large amount of process data has been collected, transmitted, and stored, which makes the data-driven soft sensing techniques quite reasonable solutions for quality prediction tasks [9], [10].

To date, various multivariate statistical process control models have been developed for industrial soft sensors, such as principal component regression (PCR) [11] and partial least square (PLS) [12], which tried to project the original data into latent space to extract corresponding data characteristics. These models perform well when the process is hardly affected by stochastic noise, but do not handle the

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

process uncertainty problem well. Alternatively, a series of probabilistic frameworks [10], [13] have been developed to cope with the above situations, including the probabilistic principal component regression (PPCR) model [14] and the supervised factor analysis (SFA) model [15]. Specifically, the probabilistic latent variable models are a class of methods that find variable relationships using maximum likelihood estimation, which not only preserves the average value operations in measurements, but also contains a covariance term that represents the process uncertainty. Despite the progress, the above models are always constructed based on an assumption that the samples are independent of each other, which lacks a dynamic description for the autocorrelations of process data [16], [17].

In most industrial processes, the process data could exhibit dynamical characteristics due to the existence of the systems control loops, resulting in the deficiency of traditional static latent variable models [18], [19]. In fact, the autocorrelation characteristics among the process data always play important roles in quality prediction tasks. To this end, several pioneering works have been carried out, such as the time-series models [20], and the autoregressive models [21]. Among them, Ge et al. recently proposed a dynamic probabilistic latent variable (DPLV) model [22] and a supervised linear dynamic system (SLDS) model [23] by introducing the linear Gaussian state space method into the latent variable modeling framework. In this way, both cross-correlation and autocorrelation within the data can be employed for the regression purpose. On this basis, literatures [24], [25] improved the above dynamic latent variable structure with the concept of autoregression modeling where an autoregression dynamic latent variable (ARDLV) model and its supervised version were then carried out for high-order dynamical information description. Nevertheless, the above latent information derived from the dynamic structure cannot be directly applied for quality prediction since the quality-related information has not been emphasized.

Recently, supervised deep learning modeling has made great advances in soft sensing techniques. For instance, Shen proposed a supervised weighted nonlinear dynamic system to extract the deep dynamic and nonlinear features within the process [26]. Guo introduced a semi-supervised dynamic soft sensor to capture dynamic characteristics while eliminating noise and redundancy within the raw data [27]. Besides, to capture hierarchical local nonlinear dynamic features, Yuan designed a dynamic convolutional neural network and applied it in soft sensor modeling [28]. These methods mainly focused dynamic or static feature within the process. Unfortunately, due to process complexity, different control strategies can be found in one system, which will lead to the coexistence of dynamic and static data. As a result, both quality-related dynamic and quality-related static information should be considered in a uniformed model.

On the other hand, due to the limitations in measurement technology, it is difficult and time-consuming to collect some quality variables, resulting in severe incomplete data problems during the modeling [29]. Particularly, the down and up sampling strategies are often exploited by simply deleting incomplete samples or estimating missing values in advance, so that traditional models can be applied on the modified dataset [30], [31]. However, this kind of data form regularization not only introduces extra noise into the original data but abandons a large amount of valuable quality data information. Thus, it is of particular interest if the existing data can be fully incorporated into those models without any modification. In recent years, to tackle the above problem, the semi-supervised modeling methods have attracted much attention [32], [33]. For example, literature [34] proposed a semi-supervised soft sensing method based on a hierarchical extreme learning machine to combine both labeled and unlabeled data through sample similarity where the final latent layer serves as a integration of all information, so that accurate regression model can be established with the existing data [35]. In addition, semi-supervised learning can also be applied to static probabilistic generative models, such as the semi-supervised PPCA model [36] and the semi-supervised PPLS model [37], so as to capture the variable cross-correlation for soft sensing modeling. Alternatively, to prevent dynamic drift due to data incompleteness, [38] and [39] proposed dynamical soft sensor models for quality variable missing situations where the process global information can be learned from those incomplete samples. Unfortunately, the above semi-supervised methods have not well balanced the relationship between dynamic information and static information when extracting quality-related information.

In this paper, a concurrent dual-latent variable (CDLV) model is proposed for soft sensing purpose in which two kinds of latent variables are designed to deal with the quality-related dynamic information and quality-related static information, respectively. Specifically, the dynamical latent variable defined by the first-order Markov process is designed to describe the quality-related dynamic information among the original data, while the static latent variable mainly focuses on remaining quality-related information extraction. In addition, inspired by the previous research, the CDLV model is further extended to a semi-supervised form, namely the semi-supervised dual latent variable structure (Ss-CDLV) to handle the insufficient quality data problem. In summary, the main contribution can be summarized as follows.

- A novel CDLV structure is proposed where dynamical and statical latent variables are designed to give a full explanation for soft sensor construction.
- The Ss-CDLV is further developed to deal with insufficient quality variables issue.

The remainder of this article is organized as follows. Preliminaries are given in Section II where two classical probabilistic latent variable regression models are briefly reviewed. In Section III, the proposed CDLV model will be discussed in detail, followed by the Ss-CDLV model construction in Section IV. The performance of proposed methods is validated

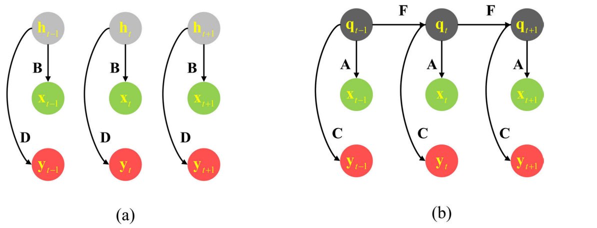


FIGURE 1. Models structure; (a) supervised factor analysis; (b) supervised linear dynamic system.

through two industrial cases in Section V. Finally, some conclusions are made.

II. PRELIMINARIES

A. PREVIOUS RESEARCHES

The SFA model is a linear Gaussian latent variable model that takes quality variable information into consideration for regression modeling. Compared with the traditional factor analysis (FA) model that only uses variables $\mathbf{x}_t$, SFA treats additional quality variables $\mathbf{y}_t$ as a constraint where the follow common latent variable $\mathbf{h}_t$ is created to capture the quality-related information. The structure of the SFA model is shown in Figure 1(a).

$$\begin{aligned} \mathbf{x}_t &= \mathbf{B}\mathbf{h}_t + \mathbf{e}^{\mathbf{x}} \\ \mathbf{y}_t &= \mathbf{D}\mathbf{h}_t + \mathbf{e}^{\mathbf{y}} \end{aligned} \tag{1}$$

However, the $\mathbf{h}_{t=1:T} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ in (1) means that only part of the static information related to the quality variables can be obtained using the SFA model.

In contrast, the SLDS builds a new latent variable $\mathbf{q}_t$ to learn quality-related dynamic information, as shown in (2).

$$\begin{aligned} \mathbf{q}_t &= \mathbf{F}\mathbf{q}_{t-1} + \mathbf{e}^{\mathbf{q}} \\ \mathbf{x}_t &= \mathbf{A}\mathbf{q}_t + \mathbf{e}^{\mathbf{x}} \\ \mathbf{y}_t &= \mathbf{C}\mathbf{q}_t + \mathbf{e}^{\mathbf{y}} \end{aligned} \tag{2}$$

where the dynamical latent variable $\mathbf{q}_t$ tries to describe variable cross-correlations from the current observed variables $\mathbf{x}_t$ and $\mathbf{y}_t$, which adds additional relationship between $\mathbf{q}_t$ and $\mathbf{q}_{t-1}$. $\mathbf{e}^{\mathbf{q}}$, $\mathbf{e}^{\mathbf{x}}$, $\mathbf{e}^{\mathbf{y}}$ represent model noises. The structure of the SLDS model is shown in Figure 1(b).

B. PROBLEM STATEMENT

Although both SFA and SLDS try to involve quality variables, the two structures have not focused on the relationship between process variables and quality variables which is of great importance for soft sensor construction. That means these two methods are naturally incapable of handling quality-related information. Besides, the latent variables introduced in these methods are assumed to be static or dynamic, which indicates processes characterized by both features cannot be described by these methods. Therefore, it seems more reasonable to introduce two different latent variables to care about static and dynamic features respectively.

TABLE 1. Descriptions of all mathematical terms.

Mathematical term	Description
x_t, y_t	represent observed data.
h_t	represents the quality-related static latent variable.
q_t	indicates the quality-related dynamic latent variable depicted by the first-order Markov chain.
s,l,m,n	represent the vector dimensions in $\mathbf{q}_t \in \Re^{(l\times1)}$, $\mathbf{h}_t \in \Re^{(s\times1)}$, $\mathbf{x}_t \in \Re^{(m\times1)}$, $\mathbf{y}_t \in \Re^{(n\times1)}$, respectively.
A,B,C,D	represent the observation transition matrices.
e^q,e^x,e^y	represent the noise terms with zero means and diagonal covariances.
θ_{CDLV}	represents the model parameter set.
$\sum_x$,$\sum_y$,$\sum_q$	represent thediagonal covariances.
T	number of the sample set.
u_t^q,V_t^q	represent the corrected posterior disributions of q_t at time t.
$\sum^{qh}$ $\sum^{hq}$	represent zero matrices.
K_t	represents the intermediate variable.
ϖ_t	represents sampling indicator.
$X_{k=1,2,3,\ldots}^{query}$	represents a query sample.

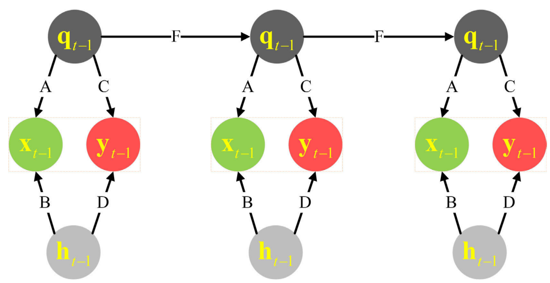


FIGURE 2. CDLV model structure.

Furthermore, the main mathematical symbols appearing in this paper are defined in Table 1.

III. QUALITY-RELATED CONCURRENT DUAL-LATENT VARIABLE MODELING

A. MODEL DESCRIPTION

In this paper, the CDLV model is designed to extract complete quality-related information using two types of latent variables, as shown in Figure 2, where $\mathbf{q}_t \in \Re^{(l\times1)}$ indicates the quality-related dynamic latent variable depicted by the first-order Markov chain while $\mathbf{h}_t \in \Re^{(s\times1)}$ is the quality-related static latent variable. The relationship between observation and latent variables can be expressed as (3), where s, l, m, n represent the vector dimensions in $\mathbf{q}_t \in \Re^{(l\times1)}$, $\mathbf{h}_t \in \Re^{(s\times1)}$, $\mathbf{x}_t \in \Re^{(m\times1)}$, $\mathbf{y}_t \in \Re^{(n\times1)}$, respectively. Compared to previous methods, quality-related information reflected by the latent variables $\mathbf{q}_t$ and $\mathbf{h}_t$ can take both quality-related dynamic information and static information simultaneously.

$$\begin{aligned} \mathbf{q}_t &= \mathbf{F}\mathbf{q}_{t-1} + \mathbf{e}^{\mathbf{q}} \\ \mathbf{x}_t &= \mathbf{A}\mathbf{q}_t + \mathbf{B}\mathbf{h}_t + \mathbf{e}^{\mathbf{x}} \\ \mathbf{y}_t &= \mathbf{C}\mathbf{q}_t + \mathbf{D}\mathbf{h}_t + \mathbf{e}^{\mathbf{y}} \end{aligned} \tag{3}$$

in which $\mathbf{h}_{t=1:T} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is assumed to be independent of $\mathbf{q}_{t=1:T}$. $\mathbf{A} \in \Re^{(m\times l)}, \mathbf{B} \in \Re^{(m\times s)}, \mathbf{C} \in \Re^{(n\times l)}$ and $\mathbf{D} \in \Re^{(n\times s)}$ are the observation transition matrices. $\mathbf{e^q} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma_q}\right), \mathbf{e^x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_x}), \mathbf{e^y} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma_y}\right)$ are the noise terms with zero means and diagonal covariances. The model parameter set $\Theta_{CDLV} = \left\{\mathbf{u}_0^\mathbf{q}, \mathbf{V}_0^\mathbf{q}, \mathbf{F}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{\Sigma_q}, \mathbf{\Sigma_x}, \mathbf{\Sigma_y}\right\}$ can be calculated by maximizing the joint log-likelihood function in (4).

$$\begin{aligned}
&\ln p\left(\mathbf{x}_{t=1:T}, \mathbf{y}_{t=1:T}, \mathbf{q}_{t=1:T}, \mathbf{h}_{t=1:T}\right) \\
&= \ln p\left(\mathbf{q}_1\right) + \sum_{t=2}^{T} \ln p\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}\right) + \sum_{t=1}^{T} \ln p\left(\mathbf{x}_t \mid \mathbf{q}_t, \mathbf{h}_t\right) \\
&\quad + \sum_{t=1}^{T} \ln p\left(\mathbf{y}_t \mid \mathbf{q}_t, \mathbf{h}_t\right)
\end{aligned} \tag{4}$$

where T is the total number of the sample set and the log-likelihood function in (4) can be derived by the application of the product rule of probability.

B. MODEL PARAMETER ESTIMATION

The maximization of log-likelihood function is always difficult. To solve this problem, the above parameters can be obtained by the expectation maximization (EM) algorithm including E-step and M-step whose convergence has been demonstrated in [40], [41], and [42].

E-step: The posterior distributions of the latent variables $\mathbf{q}_t \sim \mathcal{N}\left(\mathbf{u}_t^\mathbf{q}, \mathbf{V}_t^\mathbf{q}\right)$ and $\mathbf{h}_t \sim \mathcal{N}\left(\mathbf{u}_t^\mathbf{h}, \mathbf{V}_t^\mathbf{h}\right)$ should be calculated as well as the expectation form of the model log-likelihood function with respect to these latent variables. Given the observed data $\mathbf{x}_{t=1:T}$ and $\mathbf{y}_{t=1:T}$, the posterior distribution of the latent variables can be calculated by the improved Kalman filter algorithm after state prediction and observation correction. In the prediction stage, the state of dynamic latent variables $\mathbf{q}_t$ is obtained from its previous moment.

$$p\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}\right) \sim \mathcal{N}\left(\mathbf{F}\mathbf{u}_{t-1}^\mathbf{q}, \mathbf{F}\mathbf{V}_{t-1}^\mathbf{q}\mathbf{F}^T + \mathbf{\Sigma_q}\right) \tag{5}$$

where the statistics $\mathbf{F}\mathbf{u}_{t-1}^\mathbf{q}$ and $\mathbf{F}\mathbf{V}_{t-1}^\mathbf{q}\mathbf{F}^T + \mathbf{\Sigma_q}$ are the predictive distributions of $\mathbf{q}_t$ at time t. $\mathbf{u}_{t-1}^\mathbf{q}, \mathbf{V}_{t-1}^\mathbf{q}$ are the corrected posterior distributions of $\mathbf{q}_{t-1}$ at time $t-1$. In particular, $p\left(\mathbf{q}_{t=2} \mid \mathbf{q}_{t=1}\right) \sim \mathcal{N}\left(\mathbf{F}\mathbf{u}_0^\mathbf{q}, \mathbf{F}\mathbf{V}_0^\mathbf{q}\mathbf{F}^T + \mathbf{\Sigma_q}\right)$. On the other hand, the static latent variables $\mathbf{h}_{t=1:T}$ obey a prior distribution throughout the whole process:

$$p\left(\mathbf{h}_t \mid \mathbf{h}_{t-1}\right) = p\left(\mathbf{h}_t\right) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{6}$$

According to the independence assumption between $\mathbf{q}_t$ and $\mathbf{h}_t$, the joint distribution of current time t can be defined as:

$$\begin{aligned}
&p\left(\begin{bmatrix}\mathbf{q}_t \\ \mathbf{h}_t\end{bmatrix} \begin{bmatrix}\mathbf{q}_{t-1} \\ \mathbf{h}_{t-1}\end{bmatrix}\right) \\
&\quad \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{F}\mathbf{u}_{t-1}^\mathbf{q} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{F}\mathbf{V}_{t-1}^\mathbf{q}\mathbf{F}^T + \mathbf{\Sigma_q} & \mathbf{\Sigma^{qh}} \\ \mathbf{\Sigma^{hq}} & \mathbf{I}\end{bmatrix}\right)
\end{aligned} \tag{7}$$

Combined with the diagonal property of $\mathbf{F}$, $p\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}\right)$ and $p\left(\mathbf{h}_t\right)$ are also considered to be independent of each other. That means the two sub-matrices $\mathbf{\Sigma^{qh}}$ and $\mathbf{\Sigma^{hq}}$ in (7) are zero matrices. Otherwise, $\mathbf{q}_t$ and $\mathbf{h}_t$ will be collinear, which results in the incomplete extraction of either dynamic information or static feature information.

In the correction stage, the observation evolution equation in (3) can be rewritten as:

$$\begin{aligned}
&\begin{bmatrix}\mathbf{x}_t \\ \mathbf{y}_t\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{bmatrix} \begin{bmatrix}\mathbf{q}_t \\ \mathbf{h}_t\end{bmatrix} + \begin{bmatrix}\mathbf{e^x} \\ \mathbf{e^y}\end{bmatrix} \\
&\begin{bmatrix}\mathbf{e^x} \\ \mathbf{e^y}\end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\mathbf{\Sigma_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_y}\end{bmatrix}\right)
\end{aligned} \tag{8}$$

Therefore, the conditional distribution of the joint latent variables $\begin{bmatrix}\mathbf{q}_t \\ \mathbf{h}_t\end{bmatrix}$ will be modified as:

$$\begin{aligned}
\begin{bmatrix}\mathbf{u}_t^\mathbf{q} \\ \mathbf{u}_t^\mathbf{h}\end{bmatrix} &= \begin{bmatrix}\mathbf{F}\mathbf{u}_{t-1}^\mathbf{q} \\ \mathbf{0}\end{bmatrix} + \mathbf{K}_t \left[\begin{bmatrix}\mathbf{x}_t \\ \mathbf{y}_t\end{bmatrix} \right. \\
&\quad \left. - \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{bmatrix} \begin{bmatrix}\mathbf{F}\mathbf{u}_{t-1}^\mathbf{q} \\ \mathbf{0}\end{bmatrix}\right]
\end{aligned} \tag{9}$$

$$\begin{aligned}
\begin{bmatrix}\mathbf{V}_t^\mathbf{q} & \hat{\mathbf{\Sigma}}_t^\mathbf{qh} \\ \hat{\mathbf{\Sigma}}_t^\mathbf{hq} & \mathbf{V}_t^\mathbf{h}\end{bmatrix} &= \left[\mathbf{I} - \mathbf{K}_t \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{bmatrix}\right] \\
&\quad \times \begin{bmatrix}\mathbf{F}\mathbf{V}_{t-1}^\mathbf{q}\mathbf{F}^T + \mathbf{\Sigma_q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\end{bmatrix}
\end{aligned} \tag{10}$$

$$\begin{aligned}
\mathbf{K}_t &= \begin{bmatrix}\mathbf{F}\mathbf{V}_{t-1}^\mathbf{q}\mathbf{F}^T + \mathbf{\Sigma_q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\end{bmatrix} \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{bmatrix}^T \\
&\quad \left[\begin{bmatrix}\mathbf{\Sigma_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_y}\end{bmatrix} + \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{bmatrix} \right. \\
&\quad \left. \times \begin{bmatrix}\mathbf{F}\mathbf{V}_{t-1}^\mathbf{q}\mathbf{F}^T + \mathbf{\Sigma_q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\end{bmatrix} \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{bmatrix}^T\right]
\end{aligned} \tag{11}$$

in which $\mathbf{K}_t$ is the intermediate variable. On this basis, the marginal distributions of the two latent variables can be updated according to the condition of observation, which can be used as the input of the next prediction stage.

$$\begin{aligned}
p\left(\mathbf{q}_t \mid \mathbf{x}_t, \mathbf{y}_t\right) &\sim \mathcal{N}\left(\mathbf{u}_t^\mathbf{q}, \mathbf{V}_t^\mathbf{q}\right) \\
p\left(\mathbf{h}_t \mid \mathbf{x}_t, \mathbf{y}_t\right) &\sim \mathcal{N}\left(\mathbf{u}_t^\mathbf{h}, \mathbf{V}_t^\mathbf{h}\right)
\end{aligned} \tag{12}$$

Finally, the posterior expectation of the latent variables can be obtained as follows:

$$\begin{aligned}
E\left\langle \mathbf{q}_t \mid \mathbf{x}_{1:T}, \mathbf{y}_{1:T}\right\rangle &= \mathbf{u}_t^\mathbf{q} \\
E\left\langle \mathbf{h}_t \mid \mathbf{x}_{1:T}, \mathbf{y}_{1:T}\right\rangle &= \mathbf{u}_t^\mathbf{h}
\end{aligned}$$

$$
\begin{aligned}
E\left\langle \mathbf{q}_t\mathbf{q}_t^T \mid \mathbf{x}_{1:T}, \mathbf{y}_{1:T}\right\rangle &= \mathbf{V}_t^{\mathbf{q}} + \mathbf{u}_t^{\mathbf{q}}\left(\mathbf{u}_t^{\mathbf{q}}\right)^T \\
E\left\langle \mathbf{h}_t\mathbf{h}_t^T \mid \mathbf{x}_{1:T}, \mathbf{y}_{1:T}\right\rangle &= \mathbf{V}_t^{\mathbf{h}} + \mathbf{u}_t^{\mathbf{h}}\left(\mathbf{u}_t^{\mathbf{h}}\right)^T \\
E\left\langle \mathbf{q}_t\mathbf{q}_{t-1}^T \mid \mathbf{x}_{1:T}, \mathbf{y}_{1:T}\right\rangle &= \mathbf{V}_{t-1}^{\mathbf{q}}\mathbf{F}^T\left(\mathbf{F}\mathbf{V}_{t-1}^{\mathbf{q}}\mathbf{F}^T + \mathbf{\Sigma}_{\mathbf{q}}\right)^{-1}\mathbf{V}_t^{\mathbf{q}} \\
&\quad + \mathbf{u}_t^{\mathbf{q}}\left(\mathbf{u}_{t-1}^{\mathbf{q}}\right)^T
\end{aligned} \tag{13}
$$

The expectation of the log-likelihood function can be expressed as:

$$
\begin{aligned}
&E_{\mathbf{q}_t,\mathbf{h}_t}\left\langle \ln p\left(\mathbf{x}_{t=1:T}, \mathbf{y}_{t=1:T}, \mathbf{q}_{t=1:T}, \mathbf{h}_{t=1:T}\right)\right\rangle \\
&\quad = E_{\mathbf{q}_1}\left\langle \ln p\left(\mathbf{q}_1\right)\right\rangle + \sum_{t=2}^{T} E_{\mathbf{q}_t}\left\langle \ln p\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}\right)\right\rangle \\
&\qquad + \sum_{t=1}^{T} E_{\mathbf{q}_t,\mathbf{h}_t}\left\langle \ln p\left(\mathbf{x}_t \mid \mathbf{q}_t, \mathbf{h}_t\right)\right\rangle \\
&\qquad + \sum_{t=1}^{T} E_{\mathbf{q}_t,\mathbf{h}_t} \ln p\left(\mathbf{y}_t \mid \mathbf{q}_t, \mathbf{h}_t\right)
\end{aligned} \tag{14}
$$

where each part can be expanded as:

$$
\begin{aligned}
&E_{\mathbf{q}_1}\left\langle \ln p\left(\mathbf{q}_1\right)\right\rangle \\
&\Leftrightarrow \int_{\mathbf{q}_1} \ln \mathcal{N}\left(\mathbf{u}_0^{\mathbf{q}}, \mathbf{V}_0^{\mathbf{q}}\right) d\mathbf{q}_1 \\
&= -\frac{1}{2}\left\{\begin{array}{c}
\ln\left|\mathbf{V}_0^{\mathbf{q}}\right| + E\left\langle \mathbf{q}_1^T\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}\mathbf{q}_1\right\rangle - \left(\mathbf{u}_0^{\mathbf{q}}\right)^T\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}E\left\langle \mathbf{q}_1\right\rangle \\
-E\left\langle \mathbf{q}_1^T\right\rangle\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}\mathbf{u}_0^{\mathbf{q}} + \left(\mathbf{u}_0^{\mathbf{q}}\right)^T\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}\mathbf{u}_0^{\mathbf{q}}
\end{array}\right\}
\end{aligned} \tag{15}
$$

$$
\begin{aligned}
&E_{\mathbf{q}_t}\left\langle \ln p\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}\right)\right\rangle \\
&\Leftrightarrow \int_{\mathbf{q}_t} \ln \mathcal{N}\left(\mathbf{F}\mathbf{q}_{t-1}, \mathbf{\Sigma}_{\mathbf{q}}\right) d\mathbf{q}_t \\
&= -\frac{1}{2}\left\{\begin{array}{c}
\ln\left|\mathbf{\Sigma}_{\mathbf{q}}\right| + E\left\langle \mathbf{q}_t^T\mathbf{\Sigma}_{\mathbf{q}}^{-1}\mathbf{q}_t\right\rangle - E\left\langle \mathbf{q}_{t-1}^T\mathbf{F}^T\mathbf{\Sigma}_{\mathbf{q}}^{-1}\mathbf{q}_t\right\rangle \\
-E\left\langle \mathbf{q}_t^T\mathbf{\Sigma}_{\mathbf{q}}^{-1}\mathbf{F}\mathbf{q}_{t-1}\right\rangle + E\left\langle \mathbf{q}_{t-1}^T\mathbf{F}^T\mathbf{\Sigma}_{\mathbf{q}}^{-1}\mathbf{F}\mathbf{q}_{t-1}\right\rangle
\end{array}\right\}
\end{aligned} \tag{16}
$$

$$
\begin{aligned}
&E_{\mathbf{q}_t,\mathbf{h}_t}\left\langle \ln p\left(\mathbf{x}_t \mid \mathbf{q}_t, \mathbf{h}_t\right)\right\rangle \\
&\Leftrightarrow \int_{\mathbf{q}_t}\int_{\mathbf{h}_t} \ln \mathcal{N}\left(\mathbf{A}\mathbf{q}_t + \mathbf{B}\mathbf{h}_t, \mathbf{\Sigma}_{\mathbf{x}}\right) d\mathbf{h}_t d\mathbf{q}_t \\
&= -\frac{1}{2}\left\{\begin{array}{c}
\ln\left|\mathbf{\Sigma}_{\mathbf{x}}\right| + \mathbf{x}_t^T\left(\mathbf{\Sigma}_{\mathbf{x}}\right)^{-1}\mathbf{x}_t \\
-\left[E\left\langle \mathbf{q}_t^T\right\rangle E\left\langle \mathbf{h}_t^T\right\rangle\right]\left[\mathbf{A}\ \mathbf{B}\right]^T\mathbf{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_t \\
-\mathbf{x}_t^T\mathbf{\Sigma}_{\mathbf{x}}^{-1}\left[\mathbf{A}\ \mathbf{B}\right]\left[E\left\langle \mathbf{q}_t^T\right\rangle E\left\langle \mathbf{h}_t^T\right\rangle\right]^T \\
+E\left\langle\left[\mathbf{q}_t^T\ \mathbf{h}_t^T\right]\left[\mathbf{A}\ \mathbf{B}\right]^T\mathbf{\Sigma}_{\mathbf{x}}^{-1}\left[\mathbf{A}\ \mathbf{B}\right]\left[\mathbf{q}_t^T\ \mathbf{h}_t^T\right]^T\right\rangle
\end{array}\right\}
\end{aligned} \tag{17}
$$

$$
\begin{aligned}
&E_{\mathbf{q}_t,\mathbf{h}_t}\left\langle \ln p\left(\mathbf{y}_t \mid \mathbf{q}_t, \mathbf{h}_t\right)\right\rangle \\
&\Leftrightarrow \int_{\mathbf{q}_t}\int_{\mathbf{h}_t} \ln \mathcal{N}\left(\mathbf{C}\mathbf{q}_t + \mathbf{D}\mathbf{h}_t, \mathbf{\Sigma}_{\mathbf{y}}\right) d\mathbf{h}_t d\mathbf{q}_t \\
&= -\frac{1}{2}\left\{\begin{array}{c}
\ln\left|\mathbf{\Sigma}_{\mathbf{y}}\right| + \mathbf{y}_t^T\left(\mathbf{\Sigma}_{\mathbf{y}}\right)^{-1}\mathbf{y}_t \\
-\left[E\left\langle \mathbf{q}_t^T\right\rangle E\left\langle \mathbf{h}_t^T\right\rangle\right]\left[\mathbf{C}\ \mathbf{D}\right]^T\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}_t \\
-\mathbf{y}_t^T\mathbf{\Sigma}_{\mathbf{y}}^{-1}\left[\mathbf{C}\ \mathbf{D}\right]\left[E\left\langle \mathbf{q}_t^T\right\rangle E\left\langle \mathbf{h}_t^T\right\rangle\right]^T \\
+E\left\langle\left[\mathbf{q}_t^T\ \mathbf{h}_t^T\right]\left[\mathbf{C}\ \mathbf{D}\right]^T\mathbf{\Sigma}_{\mathbf{y}}^{-1}\left[\mathbf{C}\ \mathbf{D}\right]\left[\mathbf{q}_t^T\ \mathbf{h}_t^T\right]^T\right\rangle
\end{array}\right\}
\end{aligned} \tag{18}
$$

M-step: The model parameters can be updated by maximizing (14), which are expressed as:

$$
\begin{aligned}
\left(\mathbf{u}_0^{\mathbf{q}}\right)^{new} &= E\left\langle \mathbf{q}_1\right\rangle \\
\left(\mathbf{V}_0^{\mathbf{q}}\right)^{new} &= E\left\langle \mathbf{q}_1\mathbf{q}_1^T\right\rangle - E\left\langle \mathbf{q}_1\right\rangle E\left\langle \mathbf{q}_1^T\right\rangle
\end{aligned} \tag{19}
$$

$$
\mathbf{F}^{new} = \left(\sum_{t=2}^{T} E\left\langle \mathbf{q}_t\mathbf{q}_{t-1}^T\right\rangle\right)\left(\sum_{t=2}^{T} E\left\langle \mathbf{q}_{t-1}\mathbf{q}_{t-1}^T\right\rangle\right)^{-1} \tag{20}
$$

$$
\mathbf{\Sigma}_{\mathbf{q}}^{new} = \frac{1}{T-1}\cdot\sum_{t=2}^{T}\left\{E\left\langle \mathbf{q}_t\mathbf{q}_t^T\right\rangle - \mathbf{F}^{new}\cdot E\left\langle \mathbf{q}_{t-1}\mathbf{q}_t^T\right\rangle\right\} \tag{21}
$$

$$
\begin{aligned}
\mathbf{A}^{new} &= \left(\sum_{t=1}^{T}\mathbf{x}_t E\left\langle \mathbf{q}_t^T\right\rangle\right)\left(\sum_{t=1}^{T} E\left\langle \mathbf{q}_t\mathbf{q}_t^T\right\rangle\right)^{-1} \\
\mathbf{B}^{new} &= \left(\sum_{t=1}^{T}\mathbf{x}_t E\left\langle \mathbf{h}_t^T\right\rangle\right)\left(\sum_{t=1}^{T} E\left\langle \mathbf{h}_t\mathbf{h}_t^T\right\rangle\right)^{-1} \\
\mathbf{C}^{new} &= \left(\sum_{t=1}^{T}\mathbf{y}_t E\left\langle \mathbf{q}_t^T\right\rangle\right)\left(\sum_{t=1}^{T} E\left\langle \mathbf{q}_t\mathbf{q}_t^T\right\rangle\right)^{-1} \\
\mathbf{D}^{new} &= \left(\sum_{t=1}^{T}\mathbf{y}_t E\left\langle \mathbf{h}_t^T\right\rangle\right)\left(\sum_{t=1}^{T} E\left\langle \mathbf{h}_t\mathbf{h}_t^T\right\rangle\right)^{-1}
\end{aligned} \tag{22}
$$

$$
\begin{aligned}
\mathbf{\Sigma}_{\mathbf{x}}^{new} &= \frac{1}{T}\cdot\sum_{t=1}^{T}\left\{\mathbf{x}_t\mathbf{x}_t^T - \left(\mathbf{A}^{new}E\left\langle \mathbf{q}_t\right\rangle + \mathbf{B}^{new}E\left\langle \mathbf{h}_t\right\rangle\right)\mathbf{x}_t^T\right\} \\
\mathbf{\Sigma}_{\mathbf{y}}^{new} &= \frac{1}{T}\cdot\sum_{t=1}^{T}\left\{\mathbf{y}_t\mathbf{y}_t^T - \left(\mathbf{C}^{new}E\left\langle \mathbf{q}_t\right\rangle + \mathbf{D}^{new}E\left\langle \mathbf{h}_t\right\rangle\right)\mathbf{y}_t^T\right\}
\end{aligned} \tag{23}
$$

The iteration of E-step and M-step will stop until the log-likelihood function converges. The whole EM procedure of the CDLV and Ss-CDLV model is shown in Appendix B.

IV. SEMI-SUPERVISED EXTENSION AND SOFT SENSOR APPLICATION

Based on the proposed CDLV model, the Ss-CDLV model is carried out to cope with the insufficient label problem. In addition, the corresponding soft sensor is also given.

A. CONSTRUCTION OF Ss-CDLV

During data collection, not all training samples include quality variables, which results in the incomplete data issue among different samples, as shown in Figure 3. To better

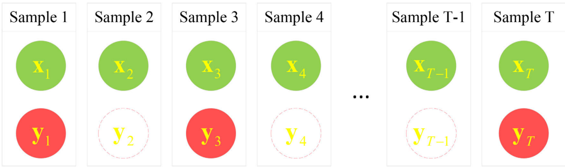


FIGURE 3. The schematic of incomplete data.

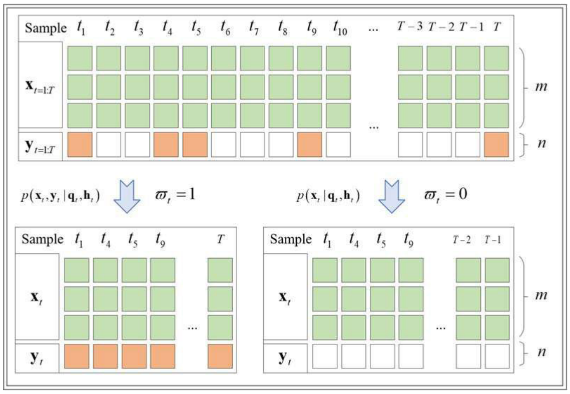


FIGURE 4. The schematic of semi-supervised data.

describe the specific missing locations of quality variables, a sampling indicator ϖ_t is introduced. When $\varpi_t = 1$, the quality variable value is effectively collected at time t. Otherwise, $\varpi_t = 0$. Therefore, the collected data can be naturally divided into labeled samples and unlabeled samples. According to the definition of CDLV in (3), the modified likelihood function of the Ss-CDLV model can be obtained as follows:

$$\ln p(\mathbf{q}_1) + \sum_{t=2}^{T} \ln p(\mathbf{q}_t \mid \mathbf{q}_{t-1}) + \begin{cases} \sum_{t=1}^{T} \ln p(\mathbf{x}_t \mid \mathbf{q}_t, \mathbf{h}_t) & \text{if } \varpi_t = 0 \\ \sum_{t=1}^{T} \ln p(\mathbf{x}_t, \mathbf{y}_t \mid \mathbf{q}_t, \mathbf{h}_t) & \text{if } \varpi_t = 1 \end{cases}$$
$$\Rightarrow \ln p(\mathbf{q}_1) + \sum_{t=2}^{T} \ln p(\mathbf{q}_t \mid \mathbf{q}_{t-1}) + \sum_{t=1}^{T} \ln p(\mathbf{x}_t \mid \mathbf{q}_t, \mathbf{h}_t) + \sum_{t=1}^{T} \varpi_t \ln p(\mathbf{y}_t \mid \mathbf{q}_t, \mathbf{h}_t) \tag{24}$$

Equation (24) is the semi-supervised learning extension of the proposed model. In order to better extract useful information from semi-supervised data, both quality-labeled and quality-unlabeled samples should be used for parameter learning of the model, as shown in the Figure 4. Therefore, (24) contains two terms, $p(\mathbf{x}_t|\mathbf{q}_t, \mathbf{h}_t)$ and $p(\mathbf{x}_t, \mathbf{y}_t|\mathbf{q}_t, \mathbf{h}_t)$, which can actually be regarded as an approximation of the model learning under full data. Since the data are fully utilized, the proposed semi-supervised model can extract the correlation relationship between process variables and quality variables to ensure the accuracy of online soft sensing.

B. PARAMETER SOLUTION OF THE Ss-CDLV MODEL

The EM algorithm is also introduced to obtain the Ss-CDLV parameters learning. Due to the insufficient quality information, those incomplete samples provide an adaptive posterior estimation for the latent variables. When the conditional distribution of the joint latent variable $\begin{bmatrix} \mathbf{q}_t \\ \mathbf{h}_t \end{bmatrix}$ in (9)-(11) will be modified as follows:

$$\begin{bmatrix} \mathbf{u}_t^{\mathbf{q}} \\ \mathbf{u}_t^{\mathbf{h}} \end{bmatrix} = \begin{bmatrix} \mathbf{F}\mathbf{u}_{t-1}^{\mathbf{q}} \\ \mathbf{0} \end{bmatrix} + \mathbf{K}_t \left[\mathbf{x}_t - \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \mathbf{F}\mathbf{u}_{t-1}^{\mathbf{q}} \right] \tag{25}$$

$$\begin{bmatrix} \mathbf{V}_t^{\mathbf{q}} & \hat{\Sigma}_t^{\mathbf{qh}} \\ \hat{\Sigma}_t^{\mathbf{hq}} & \mathbf{V}_t^{\mathbf{h}} \end{bmatrix} = \left[\mathbf{I} - \mathbf{K}_t \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \right] \begin{bmatrix} \mathbf{F}\mathbf{V}_{t-1}^{\mathbf{q}}\mathbf{F}^T + \Sigma_{\mathbf{q}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{26}$$

$$\mathbf{K}_t = \begin{bmatrix} \mathbf{F}\mathbf{V}_{t-1}^{\mathbf{q}}\mathbf{F}^T + \Sigma_{\mathbf{q}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}^T \times \left[\Sigma_{\mathbf{x}} + \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{F}\mathbf{V}_{t-1}^{\mathbf{q}}\mathbf{F}^T + \Sigma_{\mathbf{q}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}^T \right] \tag{27}$$

In M-step, all updated expressions related to quality variables are revised as:

$$\mathbf{C}^{new} = \left(\sum_{t=1}^{T} \varpi_t \mathbf{y}_t E\left\langle \mathbf{q}_t^T \right\rangle \right) \left(\sum_{t=1}^{T} \varpi_t E\left\langle \mathbf{q}_t \mathbf{q}_t^T \right\rangle \right)^{-1}$$
$$\mathbf{D}^{new} = \left(\sum_{t=1}^{T} \varpi_t \mathbf{y}_t E\left\langle \mathbf{h}_t^T \right\rangle \right) \left(\sum_{t=1}^{T} \varpi_t E\left\langle \mathbf{h}_t \mathbf{h}_t^T \right\rangle \right)^{-1} \tag{28}$$

$$\Sigma_{\mathbf{y}}^{new} = \frac{1}{T} \cdot \sum_{t=1}^{T} \varpi_t \left\{ \mathbf{y}_t \mathbf{y}_t^T - \left(\mathbf{C}^{new} E\langle \mathbf{q}_t \rangle + \mathbf{D}^{new} E\langle \mathbf{h}_t \rangle \right) \mathbf{y}_t^T \right\} \tag{29}$$

The other operation steps are consistent with the CDLV model.

C. SOFT SENSORS APPLICATION

The flow chart of soft sensing application can be shown in Figure 5 which is mainly composed of two parts including off-line stage and on-line stage. When a query sample $\mathbf{X}_{k=1,2,3,\cdots}^{query}$ appears, the trained parameters can be utilized as

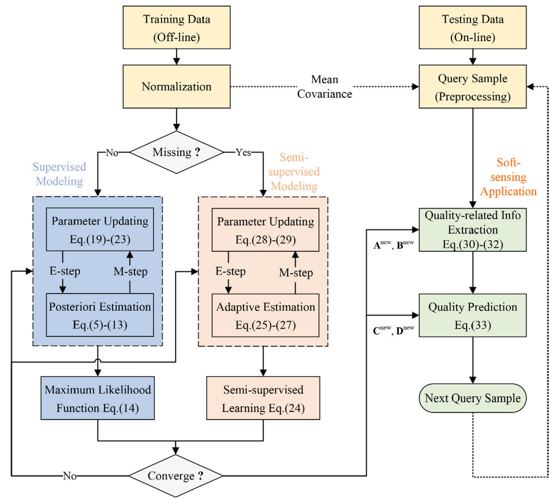


FIGURE 5. The soft sensor application flow chart.

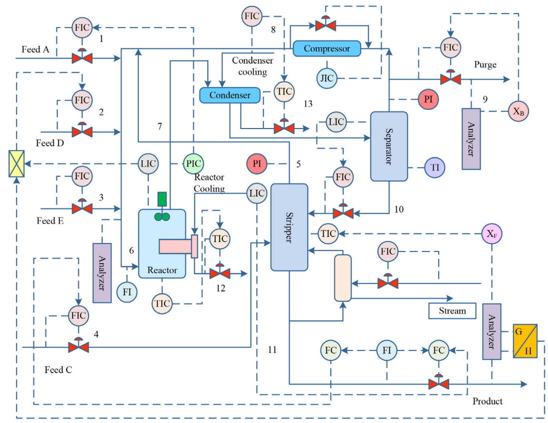


FIGURE 6. The TE process diagram.

follows:

$$\begin{bmatrix} \hat{\mathbf{u}}_k^{\mathbf{q}} \\ \hat{\mathbf{u}}_k^{\mathbf{h}} \end{bmatrix} = \begin{bmatrix} \mathbf{F}^{new}\hat{\mathbf{u}}_{k-1}^{\mathbf{q}} \\ \mathbf{0} \end{bmatrix} + \mathbf{K}_k \left[\mathbf{x}_k - \left[\mathbf{A}^{new} \quad \mathbf{B}^{new} \right] \mathbf{F}^{new}\hat{\mathbf{u}}_{k-1}^{\mathbf{q}} \right] \tag{30}$$

$$\begin{bmatrix} \hat{\mathbf{V}}_k^{\mathbf{q}} & \hat{\boldsymbol{\Sigma}}_k^{\mathbf{qh}} \\ \hat{\boldsymbol{\Sigma}}_k^{\mathbf{hq}} & \hat{\mathbf{V}}_k^{\mathbf{h}} \end{bmatrix} = \left[\mathbf{I} - \mathbf{K}_k \left[\mathbf{A}^{new} \quad \mathbf{B}^{new}\right]\right] \cdot \Psi_{k-1} \tag{31}$$

$$\mathbf{K}_k = \boldsymbol{\Psi}_{k-1} \left[\mathbf{A}^{new} \quad \mathbf{B}^{new}\right] \times \left[\boldsymbol{\Sigma}_{\mathbf{x}}^{new} + \left[\mathbf{A}^{new} \quad \mathbf{B}^{new}\right] \boldsymbol{\Psi}_{k-1} \left[\mathbf{A}^{new} \quad \mathbf{B}^{new}\right]^T \right] \tag{32}$$

where $\boldsymbol{\Psi}_{k-1} = \begin{bmatrix} \mathbf{F}^{new}\hat{\mathbf{V}}_{k-1}^{\mathbf{q}} (\mathbf{F}^{new})^T + \boldsymbol{\Sigma}_{\mathbf{q}}^{new} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$, $\hat{\mathbf{u}}_k^{\mathbf{q}}$, $\hat{\mathbf{u}}_k^{\mathbf{h}}$ and $\hat{\mathbf{V}}_k^{\mathbf{q}}$, $\hat{\mathbf{V}}_k^{\mathbf{h}}$ represent the means and covariances of the dual latent variables containing quality-related information, respectively. Then, the predicted value of quality variables can be calculated as follows:

$$\mathbf{y}_k^{predict} = \mathbf{C}^{new}\hat{\mathbf{u}}_k^{\mathbf{q}} + \mathbf{D}^{new}\hat{\mathbf{u}}_k^{\mathbf{h}} \tag{33}$$

V. CASE STUDY

In this section, two industrial cases are implemented to verify the soft sensing performance of the proposed methods, one of which is the Tennessee Eastman (TE) benchmark, and the other is a gasoline catalytic cracking process. For comparison, SFA [14], PPLS [13], SLDS [22] and their semi-supervised forms Ss-FA [15], Ss-PPLS [37], Ss-LDS [38] are also introduced. In addition, the regression evaluation index root means square error (RMSE), mean absolute error (MAE) and R-Squared ($\mathbf{R}^2$) are employed to evaluate prediction accuracy. Then the formula for calculating RMSE, MAE and $\mathbf{R}^2$ can be expressed as:

$$\mathbf{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_l)^2} \tag{34}$$

$$\mathbf{R}^2 = 1 - \frac{\sum_I \hat{y^{(i)}} - y^{(i)}}{\sum_I (\overline{y}^{(i)} - y^{(i)})^2} \tag{35}$$

$$\mathbf{MAE} = \frac{1}{m}\sum_{i=1}^{m} |y_i - \hat{y}_i| \tag{36}$$

A. TE BENCHMARK VALIDATION

TE benchmark is a chemical simulation platform, which is widely used to evaluate the process monitoring performance of different soft sensing models. The process is mainly composed of five units: reactor, condenser, compressor, separator, and stripper, involving 41 measurement variables and 12 control variables. The schematic of TE process can be shown in Figure 6 where different units organically cooperate with each other through different streams and connections. Since various control strategies are implemented, a part of the variables can be autocorrelated while others are not.

In this case, the liquid product H in stream 9 is considered as quality variable while 22 process variables XMEAS (1) – XMEAS (22) are selected as process variables which are shown in Table 2. For more detail about variable descriptions, please refer to [22]. To evaluate the performance of quality prediction, 960 samples are collected as training samples while another 960 samples are collected as the testing dataset in the same working condition. These two datasets are then normalized with unit variance and zero mean. In addition, the model hyperparameter including the initial parameters and latent dimensions can be determined according to the historical results and the PCA-based cumulative variance contribution plot [38].

TABLE 2. The description of the variables selected in the TE process.

Variables	Description	Variables	Description
$\mathbf{x}_1$	Feed flow component A(stream 1)	$\mathbf{x}_{13}$	Separator pressure in kPa gauge
$\mathbf{x}_2$	Feed flow component D(stream 2)	$\mathbf{x}_{14}$	Separator underflow in liquid phase (stream 10)
$\mathbf{x}_3$	Feed flow component E(stream 3)	$\mathbf{x}_{15}$	Stripper level
$\mathbf{x}_4$	Feed flow components A/B/C(stream 4)	$\mathbf{x}_{16}$	Stripper pressure
$\mathbf{x}_5$	Recycle flow to reactor from separator (stream 8)	$\mathbf{x}_{17}$	Stripper underflow (stream 11)
$\mathbf{x}_6$	Reactor feed rate (stream 6)	$\mathbf{x}_{18}$	Stripper temperature
$\mathbf{x}_7$	Reactor pressure	$\mathbf{x}_{19}$	Stripper steam flow
$\mathbf{x}_8$	Reactor level	$\mathbf{x}_{20}$	Compressor work
$\mathbf{x}_9$	Reactor temperature	$\mathbf{x}_{21}$	Reactor cooling water outlet temperature
$\mathbf{x}_{10}$	Purge flow rate (stream 9)	$\mathbf{x}_{22}$	Condenser cooling water outlet temperature
$\mathbf{x}_{11}$	Separator temperature	$\mathbf{x}_{23}$	Concentration of H in purge (stream 9)
$\mathbf{x}_{12}$	Separator level		

TABLE 3. The regression evaluation results of each model.

Index \ Model	CDLV	SFA	SLDS	PPLS
RMSE	**0.0785**	0.1135	0.0920	0.0974
MAE	**0.0608**	0.0897	0.0713	0.0752
$\mathbf{R}^2$	**0.7104**	0.3942	0.6023	0.5539

The prediction results of different models without unlabeled data are shown in Table 3 and Figure 7 where SFA model can partially extract the cross-correlation between process variables and quality variables by one single static latent variable. Unfortunately, it ignores the autocorrelation among training samples. Meanwhile, the SLDS model has made certain improvements because it considers the quality-related dynamic information for regression. However, neither of these two models can fully describe the quality-related information within the complex process. Although the PPLS model has divided the variable cross-correlations into two aspects to emphasize the quality-related and quality-unrelated information, the autocorrelation still has not been considered in quality prediction. Comparatively, the proposed CDLV has better soft sensing performance due to the involvement of dynamic and static latent variables so that both variable cross-correlation and sample autocorrelation can be fully explained.

In addition, another two datasets with same working condition are collected for soft sensing performance validation with insufficient sample labels. Both datasets contain 960 samples where the labels of 50 % training samples are missing. The four methods, namely Ss-CDLV, Ss-FA, Ss-PPLS and Ss-LDS, are involved and analyzed. Figure 8 shows the prediction scatter plot of each model, where the prediction of the proposed method lies very close to the diagonal line and the other three methods still have some points far from the diagonal line, such as point 1, 2 and 3. Other methods such as

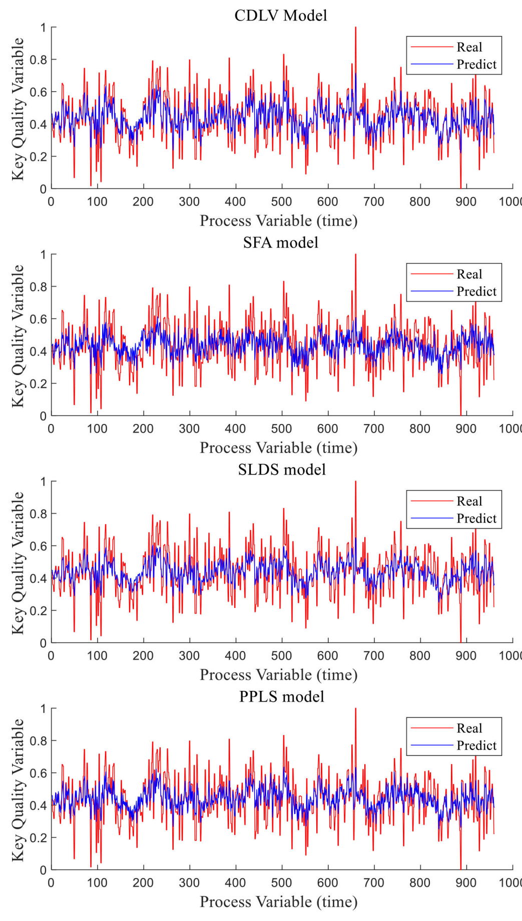


FIGURE 7. Quality prediction curve of each model.

Ss-FA and Ss-PPLS shows large distance from the diagonal line since the dynamic information within the process is not well considered while the Ss-LDS ignores the quality variable information in the quality prediction. Comparatively, the results of CDLV closely aligns with the diagonal line, indicating a higher degree of accuracy. The detailed performance

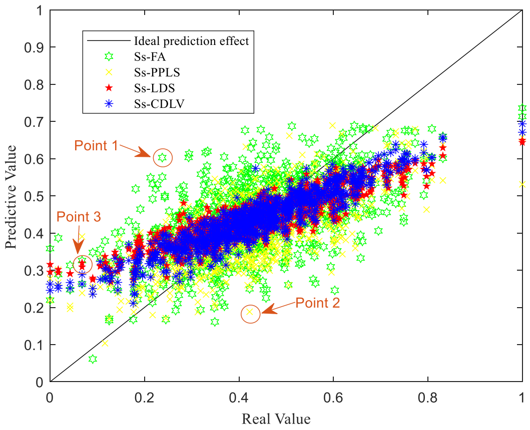


FIGURE 8. The error analysis of each semi-supervised model.

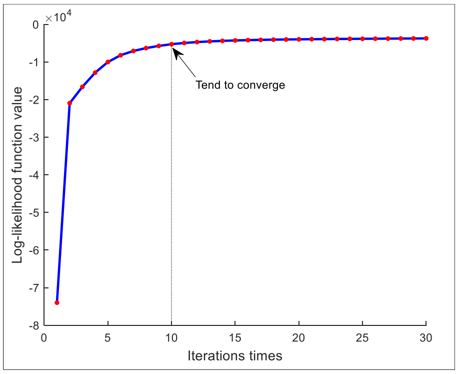


FIGURE 9. Likelihood function curve of the proposed model in TE process.

TABLE 4. The regression evaluation results of each semi-supervised model.

Index \ Model	Ss-CDLV	Ss-SFA	Ss-PPLS	Ss-SLDS
RMSE	**0.0803**	0.1213	0.1044	0.0944
MAE	**0.0626**	0.0955	0.0828	0.0738
R^2	**0.6977**	0.3076	0.4870	0.5806

of several methods can be seen in Table. 4. To demonstrate the convergence of corresponding algorithm, the convergence analysis of the proposed model in the TE process is shown in the Figure 9. It can be seen that the algorithm converges after 10 iteration times.

B. CATALYTIC CRACKING PROCESS

Nowadays, the catalytic cracking process is an important technique that converts heavy oil into gasoline, diesel and low-carbon olefin. However, current gasoline cleaning technology inevitably reduces the Research Octane Number (RON, an important indicator of the combustion performance of gasoline) in the process of desulfurization and olefin reduction, resulting in the deterioration of gasoline quality and combustion performance. Precise control of the RON in catalytic cracking gasoline refining unit can bring considerable economic benefits. Unfortunately, the reliable values of RON are usually difficult to measure online due to the complexity of the catalytic cracking gasoline production process. To guarantee high-quality gasoline production, it is necessary to perform soft sensing modeling for the catalytic cracking process to achieve real-time prediction of RON. In this case, a total of 325 samples were collected from Sinopec Gaoqiao Petrochemical real-time database and Laboratory Information Management System (LIMS). Meanwhile, 12 measurements that are highly correlated with RON are selected from the database as the process variables, including the raw material properties, product properties and the adsorbent properties of the catalytic cracking unit, as shown in Table 5. To evaluate the performance of quality prediction, the first 200 samples are used as training data and the other 125 samples serve as test data. These two datasets are then normalized with unit variance and zero mean.

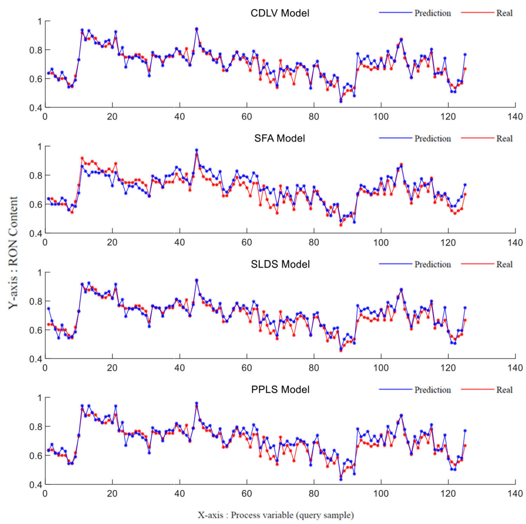


FIGURE 10. RON prediction curve of each model.

The prediction result of each comparison model can be listed in Table 6 where the proposed method shows great performance in accuracy and goodness of fit, while the prediction curves of each methods are shown in Figure 10. The prediction accuracy of the SFA and PPLS model is still unsatisfactory since the static modeling method cannot well describe the behavior of data autocorrelation within the catalytic cracking processes. Compared with the static models, the process dynamic information can be described by the latent variables of SLDS, hence better soft sensor performance can be obtained. However, a large amount of quality-related information that is more suitable for static description remains in the residual of SLDS, which has not been well extracted and reflected in the quality variable. To cope with both information, CDLV introduces two different latent variables which divide the complete quality-related information into dynamic and static parts. The dynamic latent

TABLE 5. The description of the variables selected in the gasoline catalytic cracking process.

Variables	Description	Variables	Description
$\mathbf{x}_1$	The Sulfur Feed	$\mathbf{x}_7$	Fluid Density
$\mathbf{x}_2$	The RON Feed	$\mathbf{x}_8$	The Sulfur Residue
$\mathbf{x}_3$	The Bromine Feed	$\mathbf{x}_9$	Active Carbon (U1)
$\mathbf{x}_4$	The Olefin Feed	$\mathbf{x}_{10}$	Adsorbent (U1)
$\mathbf{x}_5$	The Arene Feed	$\mathbf{x}_{11}$	Active Carbon (U2)
$\mathbf{x}_6$	Saturated hydrocarbon	$\mathbf{x}_{12}$	Adsorbent (U2)

TABLE 6. The regression evaluation results of each model.

Index \ Model	CDLV	SFA	PPLS	SLDS
RMSE	**0.0352**	0.0467	0.0410	0.0389
MAE	**0.0266**	0.0396	0.0323	0.0306
$\mathbf{R}^2$	**0.8729**	0.7765	0.8281	0.8451

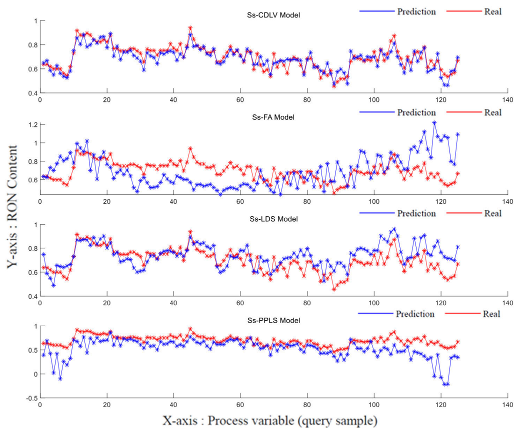


FIGURE 11. RON prediction curve of each semi-supervised model.

TABLE 7. The regression evaluation results of each semi-supervised model.

Index \ Model	Ss-CDLV	Ss-FA	Ss-LDS	Ss-PPLS
RMSE	**0.0474**	0.2002	0.0950	0.2295
MAE	**0.0388**	0.1643	0.0784	0.1737
$\mathbf{R}^2$	**0.7697**	< 0	0.7510	< 0

variable is used to extract quality-related dynamic information while the static latent variable can be a useful supplement to the description of quality-related information. Therefore, the soft sensing performance of CDLV has a great advantage over other quality-related models.

In the actual gasoline catalytic cracking process, the sampling time of the variables in Table 5 is 10 mins, while the sampling rate of RON depends on the experimental cost,

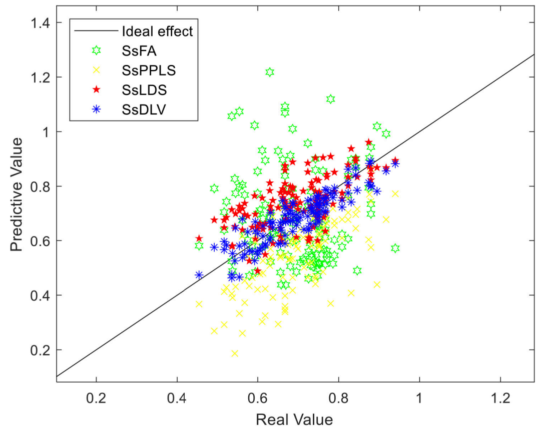


FIGURE 12. The error analysis of each semi-supervised model.

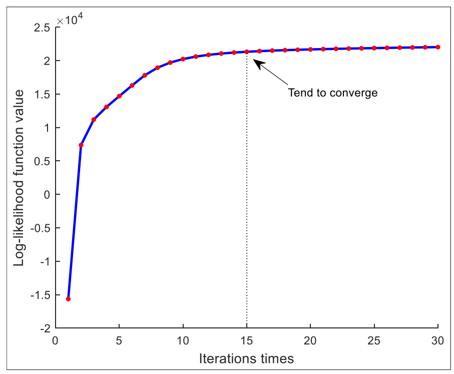


FIGURE 13. Likelihood function curve of the proposed model in Catalytic cracking process.

and it usually takes much longer time to obtain a precise measurement. To further verify the soft measurement performance of several semi-supervised models, the missing rate of quality variables in the training sample sets is set by 50 %. On this basis, the prediction results of Ss-CDLV, SSLDS and SSFA are recalculated, as shown in Table 7 where the proposed model outperforms other models in all metrics. The experiment results show that the soft sensing performance of Ss-CDLV is significantly better than Ss-FA and Ss-LDS, especially at some peak points, such as point 1 and point 2 in Figure 11. At the same time, Figure 12 provides a prediction scatter plot to intuitively reflect the prediction error of each

$$\frac{\partial E_{\mathbf{q}_1}\langle \ln p(\mathbf{q}_1)\rangle}{\partial \mathbf{u}_0^{\mathbf{q}}}=0 \Rightarrow \frac{\partial\left\{\left(\mathbf{u}_0^{\mathbf{q}}\right)^T\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}E\langle \mathbf{q}_1\rangle-2E\left\langle \mathbf{q}_1^T\right\rangle\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}\mathbf{u}_0^{\mathbf{q}}\right\}}{\partial \mathbf{u}_0^{\mathbf{q}}}=0$$

$$\Rightarrow \left(\mathbf{u}_0^{\mathbf{q}}\right)^{new}=E\langle \mathbf{q}_1\rangle \tag{A.1}$$

$$\frac{\partial E_{\mathbf{q}_1}\langle \ln p(\mathbf{q}_1)\rangle}{\partial \mathbf{V}_0^{\mathbf{q}}}=0 \Rightarrow \frac{\partial\left\{\ln\left|\mathbf{V}_0^{\mathbf{q}}\right|+E\left\langle \mathbf{q}_1^T\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}\mathbf{q}_1\right\rangle-\left(\mathbf{u}_0^{\mathbf{q}}\right)^T\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}E\langle \mathbf{q}_1\rangle-2E\left\langle \mathbf{q}_1^T\right\rangle\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}\mathbf{u}_0^{\mathbf{q}}\right\}}{\partial \mathbf{V}_0^{\mathbf{q}}}=0$$

$$\Rightarrow \left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}-\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}E\left\langle \mathbf{q}_1\mathbf{q}_1^T\right\rangle\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}+\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}E\langle \mathbf{q}_1\rangle E\left\langle \mathbf{q}_1^T\right\rangle\left(\mathbf{V}_0^{\mathbf{q}}\right)^{-1}=0$$

$$\Rightarrow \left(\mathbf{V}_0^{\mathbf{q}}\right)^{new}=E\left\langle \mathbf{q}_1\mathbf{q}_1^T\right\rangle-E\langle \mathbf{q}_1\rangle E\left\langle \mathbf{q}_1^T\right\rangle \tag{A.2}$$

$$\frac{\partial \sum_{t=2}^{T} E_{\mathbf{q}_t}\langle \ln p(\mathbf{q}_t|\mathbf{q}_{t-1})\rangle}{\partial \mathbf{F}}=0 \Rightarrow \frac{\partial\left\{\sum_{t=2}^{T}\left(E\left\langle \mathbf{q}_{t-1}^T\mathbf{F}^T\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{F}\mathbf{q}_{t-1}\right\rangle-E\left\langle \mathbf{q}_{t-1}^T\mathbf{F}^T\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{q}_t\right\rangle-E\left\langle \mathbf{q}_t^T\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{F}\mathbf{q}_{t-1}\right\rangle\right)\right\}}{\partial \mathbf{F}}=0$$

$$\Rightarrow 2\sum_{t=2}^{T}\left(\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{F}E\left\langle \mathbf{q}_{t-1}\mathbf{q}_{t-1}^T\right\rangle-\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}E\left\langle \mathbf{q}_t\mathbf{q}_{t-1}^T\right\rangle\right)=0$$

$$\Rightarrow \mathbf{F}^{new}=\left(\sum_{t=2}^{T}E\left\langle \mathbf{q}_t\mathbf{q}_{t-1}^T\right\rangle\right)\left(\sum_{t=2}^{T}E\left\langle \mathbf{q}_{t-1}\mathbf{q}_{t-1}^T\right\rangle\right)^{-1} \tag{A.3}$$

$$\frac{\partial \sum_{t=2}^{T} E_{\mathbf{q}_t}\langle \ln p(\mathbf{q}_t|\mathbf{q}_{t-1})\rangle}{\partial \boldsymbol{\Sigma}_{\mathbf{q}}}=0 \Rightarrow \frac{\partial\left\{(T-1)\ln\left|\boldsymbol{\Sigma}_{\mathbf{q}}\right|+\sum_{t=2}^{T}\begin{pmatrix} E\left\langle \mathbf{q}_t^T\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{q}_t\right\rangle-E\left\langle \mathbf{q}_{t-1}^T\mathbf{F}^T\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{q}_t\right\rangle \\ -E\left\langle \mathbf{q}_t^T\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{F}\mathbf{q}_{t-1}\right\rangle+E\left\langle \mathbf{q}_{t-1}^T\mathbf{F}^T\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{F}\mathbf{q}_{t-1}\right\rangle\end{pmatrix}\right\}}{\partial \boldsymbol{\Sigma}_{\mathbf{q}}}$$

$$=0$$

$$\Rightarrow (T-1)\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}+\sum_{t=2}^{T}\begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{F}E\left\langle \mathbf{q}_{t-1}\mathbf{q}_{t-1}^T\right\rangle\mathbf{F}\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}-\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}E\left\langle \mathbf{q}_t\mathbf{q}_t^T\right\rangle\boldsymbol{\Sigma}_{\mathbf{q}}^{-1} \\ +\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\mathbf{F}E\left\langle \mathbf{q}_t\mathbf{q}_{t-1}^T\right\rangle\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}+\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}E\left\langle \mathbf{q}_t\mathbf{q}_{t-1}^T\right\rangle\mathbf{F}^T\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\end{pmatrix}$$

$$=0$$

$$\Rightarrow \boldsymbol{\Sigma}_{\mathbf{q}}^{new}=\frac{1}{T-1}\cdot\sum_{t=2}^{T}\left\{E\left\langle \mathbf{q}_t\mathbf{q}_t^T\right\rangle-\mathbf{F}^{new}\cdot E\left\langle \mathbf{q}_{t-1}\mathbf{q}_t^T\right\rangle\right\} \tag{A.4}$$

$$\frac{\partial \sum_{t=1}^{T} E_{\mathbf{q}_t,\mathbf{h}_t}\langle \ln p(\mathbf{x}_t|\mathbf{q}_t,\mathbf{h}_t)\rangle}{\partial \mathbf{A}}=0 \Rightarrow \frac{\partial\left\{\sum_{t=1}^{T}\left(E\left\langle \mathbf{q}_t^T\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{A}\mathbf{q}_t\right\rangle-E\left\langle \mathbf{q}_t^T\right\rangle\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_t-\mathbf{x}_t^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{A}E\langle \mathbf{q}_t\rangle\right)\right\}}{\partial \mathbf{A}}=0$$

$$\Rightarrow \sum_{t=1}^{T}\left(2\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{A}E\left\langle \mathbf{q}_t\mathbf{q}_t^T\right\rangle-\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_tE\left\langle \mathbf{q}_t^T\right\rangle-\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_tE\left\langle \mathbf{q}_t^T\right\rangle\right)=0$$

$$\Rightarrow \mathbf{A}^{new}=\left(\sum_{t=1}^{T}\mathbf{x}_tE\left\langle \mathbf{q}_t^T\right\rangle\right)\left(\sum_{t=1}^{T}E\left\langle \mathbf{q}_t\mathbf{q}_t^T\right\rangle\right)^{-1} \tag{A.5}$$

$$\frac{\partial \sum_{t=1}^{T}\left\{E_{\mathbf{q}_t,\mathbf{h}_t}\langle \ln p(\mathbf{x}_t|\mathbf{q}_t,\mathbf{h}_t)\rangle\right\}}{\partial \boldsymbol{\Sigma}_{\mathbf{x}}}=0 \Rightarrow \frac{\partial\sum_{t=1}^{T}\left\{\frac{1}{T}\ln\left|\boldsymbol{\Sigma}_{\mathbf{x}}\right|+\sum_{t=1}^{T}\begin{pmatrix}E\left\langle \mathbf{q}_t^T\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{A}\mathbf{q}_t\right\rangle+\mathbf{x}_t^T(\boldsymbol{\Sigma}_{\mathbf{x}})^{-1}\mathbf{x}_t \\ -E\left\langle \mathbf{q}_t^T\right\rangle\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_t-\mathbf{x}_t^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{A}E\langle \mathbf{q}_t\rangle\end{pmatrix}\right\}}{\partial \boldsymbol{\Sigma}_{\mathbf{x}}}=0$$

$$\Rightarrow T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}+\sum_{t=1}^{T}\begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{A}E\left\langle \mathbf{q}_t\mathbf{q}_t^T\right\rangle\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_t\mathbf{x}_t^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \\ -\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{A}E\langle \mathbf{q}_t\rangle\mathbf{x}_t^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}-\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_tE\left\langle \mathbf{q}_t^T\right\rangle\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\end{pmatrix}=0$$

$$\Rightarrow \boldsymbol{\Sigma}_{\mathbf{x}}^{new}=\frac{1}{T}\cdot\sum_{t=1}^{T}\left\{\mathbf{x}_t\mathbf{x}_t^T-\left(\mathbf{A}^{new}E\langle \mathbf{q}_t\rangle+\mathbf{B}^{new}E\langle \mathbf{h}_t\rangle\right)\mathbf{x}_t^T\right\} \tag{A.6}$$

Algorithm 1 The EM Procedure of the CDLV and Ss-CDLV Model

Input: training dataset $\{x_{1:T}, y_{1:T}\}$

1: initialization parameter;
2: *E-step*:calculating the posteriori distribution of dual-latent variables.
3: Define the augmented form of the latent variables according to (5)-(8).
4: **if** the data is complete **then**
5: Executing (9)-(11)based on complete Kalman filter algorithm;
6: **else**
7: **for** $t = 1 : T$ **do**
8: **if** ϖ_t=0 **then**
9: Executing (25)-(27) by adaptive posterior estimation;
10: **if** ϖ_t=1 **then**
11: Executing (9)-(11) by adaptive posterior estimation;
12: Deriving the adaptive posteriori expectations of latent variables by (12)-(13);
13: *M-step*:solving model parameters by maximizing the log-likelihood functions;
14: **if** the data is complete **then**
15: Executing (19)-(23) for parameter updates;
16: **else**
17: Use (28)-(29) to replace the corresponding parameters in (19)-(23);
18: The trained parameters are used to determine whether the objective function converges. $L(\theta^{new}) - L(\theta^{previous}) < \epsilon$ represents convergence, where ϵ is the preset error value.
19: If the function converges, stop training; otherwise, go back to step 2.

model. The convergence analysis of the proposed model in Catalytic cracking process is shown in the Figure 13.

VI. CONCLUSION

It is significant and challenging to simultaneously consider both dynamic and static information in the quality prediction of modern industry. In this paper, a quality-related CDLV structure is proposed for soft sensing applications where two latent variables are designed to describe the dynamic and static parts in quality-related information respectively. Compared with traditional single-latent variable structure, the CDLV model improves the data expression for the relationship between process variables and quality variables. In addition, the CDLV is extended to the Ss-CDLV model to solve the insufficient quality information problem. The performance of the involved methods is demonstrated through two cases including the TE process and the catalytic cracking process. Compared with related works, the CDLV has shown high accuracy in predicting results where the values of RMSE and MAE reduce at least 10 % while the tracking performance index R^2 increases about 3 %. However, the nonlinearity is another feature that can affect the prediction performance. Therefore, this problem may be solved by kernel-based method or deep learning method, which will be discussed in our future works.

APPENDIX A

The calculation and derivation process of the equations (19) to (23) are expressed in (A.1)–(A.5), as shown at the top of the previous page.

Because of the structural symmetry among A, B, C, D, the calculation process of B, C, D will not be repeated here in (A.6), as shown at the bottom of the previous page.

According to the symmetry between $\mathbf{\Sigma_x}$ and $\mathbf{\Sigma_y}$, the $\mathbf{\Sigma_y}^{new}$ can also be calculated easily.

APPENDIX B

See Algorithm 1.

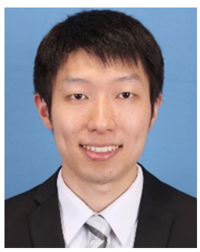
REFERENCES

[1] Z. Ying, Y. Wang, Y. He, and J. Wang, "Virtual sensing techniques for nonlinear dynamic processes using weighted probability dynamic dual-latent variable model and its industrial applications," *Knowl.-Based Syst.*, vol. 235, Jan. 2022, Art. no. 107642.
[2] P. Wang, Y. Yin, W. Bai, X. Deng, and W. Shao, "A unified just-in-time learning paradigm and its application to adaptive soft sensing for nonlinear and time-varying chemical process," *Chem. Eng. Sci.*, vol. 258, Aug. 2022, Art. no. 117753.
[3] K. Peng, K. Zhang, B. You, and J. Dong, "Quality-related prediction and monitoring of multi-mode processes using multiple PLS with application to an industrial hot strip mill," *Neurocomputing*, vol. 168, pp. 1094–1103, Nov. 2015.
[4] Q. Sun and Z. Ge, "A survey on deep learning for data-driven soft sensors," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 5853–5866, Sep. 2021.
[5] L. Yao and Z. Ge, "Refining data-driven soft sensor modeling framework with variable time reconstruction," *J. Process Control*, vol. 87, pp. 91–107, Mar. 2020.
[6] P. Kadlec, R. Grbić, and B. Gabrys, "Review of adaptation mechanisms for data-driven soft sensors," *Comput. Chem. Eng.*, vol. 35, no. 1, pp. 1–24, Jan. 2011.
[7] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: The role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.
[8] L. Yao and Z. Ge, "Big data quality prediction in the process industry: A distributed parallel modeling framework," *J. Process Control*, vol. 68, pp. 1–13, Aug. 2018.
[9] 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.
[10] 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.
[11] 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. 1–3, Apr. 2020.
[12] 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. Design*, vol. 95, pp. 113–132, Mar. 2015.
[13] X. Yang, X. Liu, and C. Xu, "Robust mixture probabilistic partial least squares model for soft sensing with multivariate Laplace distribution," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
[14] W. Xie, J.-S. Wang, C. Xing, S.-S. Guo, M.-W. Guo, and L.-F. Zhu, "Adaptive hybrid soft-sensor model of grinding process based on regularized extreme learning machine and least squares support vector machine optimized by golden sine Harris Hawk optimization algorithm," *Complexity*, vol. 2020, pp. 1–26, May 2020.

[15] Z. Li, H. Jin, S. Dong, B. Qian, B. Yang, and X. Chen, "Semi-supervised ensemble support vector regression based soft sensor for key quality variable estimation of nonlinear industrial processes with limited labeled data," *Chem. Eng. Res. Design*, vol. 179, pp. 510–526, Mar. 2022.
[16] L. Yao and Z. Ge, "Dynamic features incorporated locally weighted deep learning model for soft sensor development," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
[17] Y. Qin and C. Zhao, "Comprehensive process decomposition for closed-loop process monitoring with quality-relevant slow feature analysis," *J. Process Control*, vol. 77, pp. 141–154, May 2019.
[18] 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.
[19] J. Wang, W. Shao, X. Zhang, and Z. Song, "Dynamic variational Bayesian student's T mixture regression with hidden variables propagation for industrial inferential sensor development," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5314–5324, Aug. 2021.
[20] S. S. Rangapuram, M. W. Seeger, J. Gasthaus, L. Stella, Y. Wang, and T. Januschowski, "Deep state space models for time series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–10.
[21] J. M. B. Haslbeck, L. F. Bringmann, and L. J. Waldorp, "A tutorial on estimating time-varying vector autoregressive models," *Multivariate Behav. Res.*, vol. 56, no. 1, pp. 120–149, Jan. 2021.
[22] 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.
[23] 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.
[24] L. Zhou, G. Li, Z. Song, and S. J. Qin, "Autoregressive dynamic latent variable models for process monitoring," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 1, pp. 366–373, Jan. 2017.
[25] N. Chen, F. Hu, J. Chen, Z. Chen, W. Gui, and X. Li, "A process monitoring method based on dynamic autoregressive latent variable model and its application in the sintering process of ternary cathode materials," *Machines*, vol. 9, no. 10, p. 229, Oct. 2021.
[26] B. Shen and Z. Ge, "Weighted nonlinear dynamic system for deep extraction of nonlinear dynamic latent variables and industrial application," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3090–3098, May 2021.
[27] R. Guo and H. Liu, "Semisupervised dynamic soft sensor based on complementary ensemble empirical mode decomposition and deep learning," *Measurement*, vol. 183, Oct. 2021, Art. no. 109788.
[28] X. Yuan, S. Qi, Y. Wang, and H. Xia, "A dynamic CNN for nonlinear dynamic feature learning in soft sensor modeling of industrial process data," *Control Eng. Pract.*, vol. 104, p. 104614, 2020.
[29] 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, Nov. 2008.
[30] J. L. Schafer and J. W. Graham, "Missing data: Our view of the state of the art.," *Psychol. Methods*, vol. 7, no. 2, pp. 147–177, 2002.
[31] 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.
[32] W. Shao, Z. Ge, and Z. Song, "Semi-supervised mixture of latent factor analysis models with application to online key variable estimation," *Control Eng. Pract.*, vol. 84, pp. 32–47, Mar. 2019.
[33] S. Liao, X. Jiang, and Z. Ge, "Weakly supervised multilayer perceptron for industrial fault classification with inaccurate and incomplete labels," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 2, pp. 1192–1201, Apr. 2022.
[34] L. Yao and Z. Ge, "Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1490–1498, Feb. 2018.
[35] L. Yao and Z. Ge, "Cooperative deep dynamic feature extraction and variable time-delay estimation for industrial quality prediction," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 3782–3792, Jun. 2021.
[36] 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.
[37] G. Mu, T. Liu, J. Chen, C. Shang, and C. Zhong, "Variational PLS-based calibration model building with semi-supervised learning for moisture measurement during fluidized bed drying by NIR spectroscopy," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.
[38] Z. Ying, Y. Wang, J. Wang, and Y. He, "A dynamic latent structure with time-varying parameters for virtual sensing of industrial process with irregular missing data," *IEEE Access*, vol. 9, pp. 77017–77027, 2021.
[39] W. Shao, Y. Li, W. Han, and D. Zhao, "Block-wise parallel semisupervised linear dynamical system for massive and inconsecutive time-series data with application to soft sensing," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.
[40] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc., B Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.
[41] N. M. Nasrabadi, "Pattern recognition and machine learning," *J. Electron. Imag.*, vol. 16, no. 4, 2007, Art. no. 049901.
[42] R. J. Little and D. B. Rubin, *Statistical Analysis With Missing Data*, vol. 793. Hoboken, NJ, USA: Wiley, 2019.

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

ZE YING is currently pursuing the Ph.D. degree with the College of Information Science and Engineering, Northeastern University, Shenyang, China. His research interests include probability modeling, fault monitoring, soft sensors, and their applications in industrial processes.



YUCHEN HE (Member, IEEE) received the Ph.D. degree from Zhejiang University, China, in 2017. After that, he joined with 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.