IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 9, 2019, accepted April 17, 2019, date of publication April 24, 2019, date of current version May 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2912977

Special Probabilistic Prediction Model for Temperature Characteristics of Dynamic Fluid Processes

HONGYING DENG, YANG ZHANG, BOCHENG CHEN, YI LIU, (Member, IEEE), AND SHENGCHANG ZHANG

Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou 310023, China

Corresponding author: Yi Liu (yliuzju@zjut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61873241 and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY18F030024.

ABSTRACT Accurately predicting the temperature characteristics of a dynamic discharge process in different transportation conditions can improve the performance of reciprocating multiphase pumps in practice. However, an accurate model for the description of the complicated behavior is not available because of the unknown interphase interaction mechanisms and infeasible experiments. A probabilistic modeling method of automatically selecting prediction models is proposed for the dynamic discharge process. First, candidate computational fluid dynamics (CFD) models are empirically utilized to provide the training data for candidate Gaussian process models (GPMs). Then, a posterior probability index is proposed to assess the uncertainty of trained GPMs when the actual values are not available. With this information, the most suitable GPM and CFD models are selected sequentially for each new sample. Consequently, the developed special GPM (SGPM) can capture the main temperature characteristics. Moreover, the selection results of prediction models can provide useful information for the recognition of complicated flow patterns. The advantages of the proposed SGPM are demonstrated using a reciprocating multiphase pump under different transportation conditions.

INDEX TERMS Probabilistic modeling, gaussian process model, computational fluid dynamics, multiphase pump.

I. INTRODUCTION

Multiphase pumps, as the key facilities of close-line transportation systems, can efficiently increase oil and gas productions in the crude oil drilling [1]–[6]. Generally, under multiphase transportation conditions, the heat is generated during the gas compression, backflow, and mechanical friction processes [7]–[12]. The fluid temperature of the pump cavity will rise in the discharge process, especially for those conditions with smaller suction pressure and higher gas volume fraction [12]. A higher temperature will cause pump damages for the heat deformation of pump parts and the failure of sealing components [1]. Therefore, the temperature characteristics in different multiphase transportation conditions should be described for the reliability of multiphase pumps.

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

A. MOTIVATION

Previously, several mechanism models for describing the thermodynamic characteristics of multiphase pumps were proposed, based on mass and energy conservation equations, and variable mass thermodynamic law [9]–[12]. Unfortunately, they are not enough to describe the complex behavior of multi-component unstable flows, and the coupling phenomenon of heat and mass transfer. Additionally, it is difficult to accurately explain the interphase interaction including the resistance of bubble or particle, heat and mass transfer, etc. It is not easy to solve these high-order and nonlinear thermodynamic models [12]–[16]. Especially for a newly used reciprocating multiphase pump, the thermodynamic mechanism models are more difficult to be constructed and thus rarely involved.

Alternatively, computational fluid dynamics (CFD) models utilized in multiphase flows can provide useful information [5]–[8], [11], [12], [17], [18]. A few studies investigated the temperature characteristics of twin-screw multiphase

pumps [12]. However, little research was conducted for the reciprocating multiphase pumps. On one hand, only using a single CFD model for a working cycle is not enough to describe the time-varying flow states for the migration phenomena of the multiphase flow interface [5]. On the other hand, some CFD modeling procedures, such as the selection of multiphase and turbulence models, dynamic grid technique and user-defined functions, affect the reliability and accuracy of results [19]. Generally, large quantities of computational resources and time are required for a whole CFD simulation. Additionally, using the existing test technology, the quick and nonlinear temperature characteristics of the multiphase pumps are not easy to be accurately measured online [7]–[10], [12]. Therefore, a feasible modeling method should be developed for better description of temperature characteristics.

Recently, data-driven empirical modeling methods have been increasingly applied to process industries [4]–[6], [20]–[35]. As a nonlinear probabilistic modeling method, the Gaussian process model (GPM) can be developed without deep understanding of the process. Generally, GPM can be trained simpler and faster than the mechanism and CFD models. Additionally, compared with other data-driven models (e.g., support vector regression and deep neural networks [22], [23], [25]–[27]), GPM can evaluate the uncertainty of predictions. With this interesting property, different kinds of GPMs have been proposed for multiphase flows [4]–[6], [29] and other chemical processes [30]–[35].

B. CONTRIBUTION

In this work, a probabilistic modeling method integrating both advantages of CFD and GPM is developed to predict the temperature characteristics in different multiphase transportation conditions. First, in view of the difficulty in collecting enough experimental data, different CFD models are adopted to provide training data for the construction of several local GPMs. Then, using the Bayesian inference, a probabilistic index is developed to assess the uncertainty of local GPMs. With this information, the most suitable GPM can be adopted from the candidates for online prediction of a new sample. Meanwhile, for this sample, the corresponding CFD model is also the most appropriate one because it generates the training data for the selected GPM. Sequentially, for all samples of the discharge process, a special GPM (SGPM) can be constructed to predict the temperature characteristics of multiphase transportation conditions. Due to different CFD models having their reliable domains, the proposed method also explores that an assembled CFD model is more suitable to complicated flow patterns.

This work is organized as follows. In Section II, the probabilistic modeling and prediction method for the temperature characteristics is proposed. The prediction results of the SGPM for several new conditions are discussed and analyzed in Section III. The comparison studies of the SGPM and GPM are also investigated. Finally, the work is summarized in Section IV.

II. SGPM FOR TEMPERATURE CHARACTERISTICS PREDICTION

A. CFD MODEL FOR TEMPERATURE CHARACTERISTICS

As aforementioned, the quick and nonlinear temperature characteristics of the multiphase pumps are not easy to be accurately measured online. Several common CFD models are constructed to provide initial modeling data for GPMs. To ensure the fairness of the modeling data, some efforts are undertaken in the CFD numerical calculation. First, select the structure parameters of the CFD calculation model. To assure the calculation accuracy and reduce the computational time, they are consistent with the test pump shown in Fig. 1. Then, construct the three-dimensional geometric model using the SolidWorks software. Due to the symmetrical structure of the test pump, as shown in Fig. 2(a), half of the cavity, suction and discharge valves is adopted here. Additionally, conduct grids in the whole model using the Gambit software. Due to the complexity of the model, multiple grids are adopted and

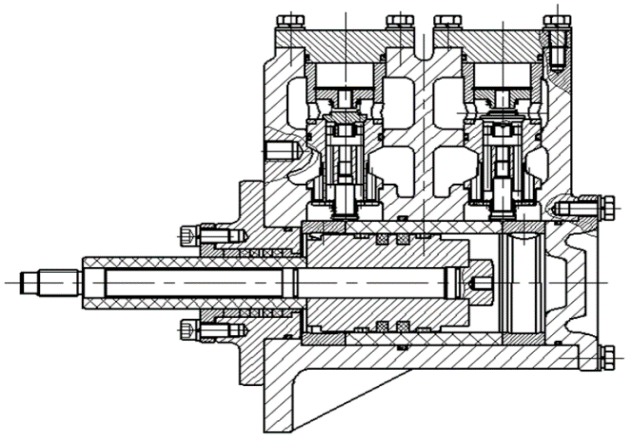


FIGURE 1. The structure of the reciprocating multiphase pump.

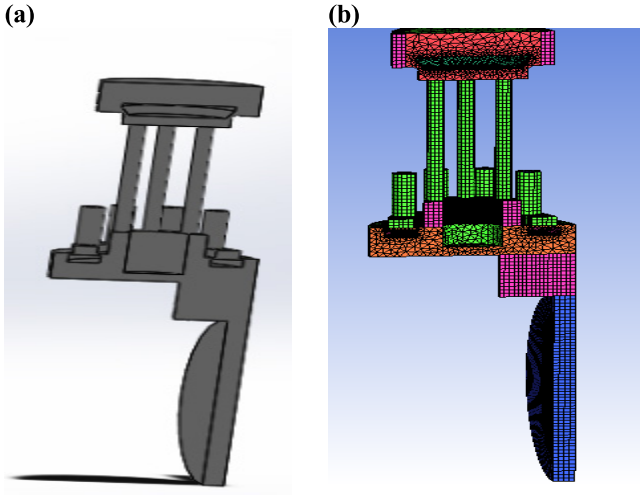


FIGURE 2. (a) The three dimensional geometric model of the reciprocating multiphase pump (b) The three dimensional numerical model of the reciprocating multiphase pump in the caption.

shown in Fig. 2(b). The grid interface, remeshing technologies and the grid independence test are carried out to ensure the meshing quality. The results show that the influence of grid size on calculation results is very small when the number of model grids reaches 287528. Finally, the Fluent software is employed for the numerical calculation.

In the numerical calculation, the mixture model is used to study the interphase coupling and pulsation characteristics of complex multiphase flows in the reciprocating multiphase pump. Additionally, the user-defined function is adopted for valves' motion before the iteration of each step. Moreover, due to the difficulty and importance for the selection of turbulence model, the RNG k-ε model (simply denoted as RNG), the SST k-ω model (simply denoted as SST), and the standard k-ε model (simply denoted as STA), are utilized here to simulate the multiphase process. The three models are widely used to describe the complicated turbulence flows. From the CFD simulation results, the y plus values for three turbulence models empirically validate their effectiveness. Finally, according to the operational environment of an oilfield in China, the liquid and gas phases are set as crude oil and methane, respectively. Therefore, based on three turbulence models, three sets of CFD simulation data (i.e., CFD_1 or CFD_{RNG}, CFD_2 or CFD_{SST}, CFD_3 or CFD_{STA}) are obtained as the training data for GPMs.

B. GPM CANDIDATES

The pump speed, the suction pressure, the discharge pressure, the gas volume fraction, and the crank angle are several important factors for the temperature characteristics [1], [7]–[12]. For practical use, the samples in the same pump speed are considered as one subclass. Consequently, the training samples from three candidate CFD models (i.e., CFD_1, CFD_2, CFD_3) can be divided into L, M and Z subclasses and denoted as $\mathbf{S} = (\mathbf{S}_1, \ldots, \mathbf{S}_l)^{\mathrm{T}}, l = 1, \ldots, L$,$\mathbf{P} = (\mathbf{P}_1, \ldots, \mathbf{P}_m)^{\mathrm{T}}, m = 1, \ldots, M$, and $\mathbf{Q} = (\mathbf{Q}_1, \ldots, \mathbf{Q}_z)^{\mathrm{T}}, z = 1, \ldots, Z$, respectively.

As an example, for the data of CFD_1 model, the lth training subclass with N_l samples is denoted as $\mathbf{S}_l = \{\mathbf{X}_l, \mathbf{y}_l\} = \left\{\mathbf{x}_{l,i}, y_{l,i}\right\}_{i=1}^{N_l}$. For each subclass, GPM provides a prediction of the output variable for an input sample through the Bayesian inference. For an output variable $\mathbf{y}_l$, GPM can be described a discrete form [36]:

$$\mathbf{y}_l = \left(y_{l,1}, \ldots, y_{l,N_l}\right)^{\mathrm{T}} \sim G\left(0, \mathbf{C}_l\right) \tag{1}$$

where $\mathbf{C}_1$ is the $N_l \times N_l$ covariance matrix with the ij-th element $C_l\left(\mathbf{x}_{l,i}, \mathbf{x}_{l,j}\right)$ defined by the covariance function below [36].

$$C_l\left(\mathbf{x}_{l,i}, \mathbf{x}_{l,j}\right) = a_{l,0} + a_{l,1}\sum_{d=1}^{D} x_{l,id}x_{l,jd} + v_{l,0}\exp\left(-\sum_{d=1}^{D} w_{l,d}\left(x_{l,id} - x_{l,jd}\right)^2\right) + \delta_{l,j}b_l \tag{2}$$

where $x_{l,id}$ is the d-th component of the vector $\mathbf{x}_{l,i}$. $\delta_{ij} = 1$ if $i = j$, otherwise, it is equal to zero. $\boldsymbol{\theta}_l = [a_{l,0}, a_{l,1}, v_{l,0}, w_{l,1}, \ldots, w_{l,d}, b_l]^{\mathrm{T}}$ are the model parameters.

Using the Bayesian method to train the lth GPM, the parameters $\boldsymbol{\theta}_l$ can be obtained [36]. Finally, for the test subclass with N_t input samples $\mathbf{X}_t = \left\{\mathbf{x}_{t,i}\right\}_{i=1}^{N_t}, t = 1, \ldots, T$, the predicted output of $y_{t,i}$ (i.e., $\hat{y}_{l,ti}$) and its variance ($\sigma^2_{\hat{y}_{l,ti}}$) can be calculated below [36].

$$\hat{y}_{l,ti} = \mathbf{k}_{l,ti}^{\mathrm{T}}\mathbf{C}_l^{-1}\mathbf{y}_l \tag{3}$$

$$\sigma^2_{\hat{y}_{l,ti}} = k_{l,ti} - \mathbf{k}_{l,ti}^{\mathrm{T}}\mathbf{C}_l^{-1}\mathbf{k}_{l,ti} \tag{4}$$

where $\mathbf{k}_{l,ti} = [C_l\left(\mathbf{x}_{t,i}, \mathbf{x}_{l,1}\right), C_l\left(\mathbf{x}_{t,i}, \mathbf{x}_{l,2}\right), \ldots, C_l(\mathbf{x}_{t,i}, \mathbf{x}_{l,N_l})]^{\mathrm{T}}$ is the covariance vector between the new input and the training data, and $k_{l,ti} = C\left(\mathbf{x}_{t,i}, \mathbf{x}_{t,i}\right)$ is the covariance of the new input. Additionally, Eq. (4) provides a confidence level on the prediction.

Consequently, several GPMs, denoted as $\mathrm{GPM}_l^1, l = 1, \ldots, L$, can be built offline for L subclasses using the Eq. (1) and Eq. (2). For a test subclass, the online prediction and its variance can be calculated using the Eq. (3) and Eq. (4), respectively. Using the same method, $\mathrm{GPM}_m^2, m = 1, \ldots, M$ for CFD_2 and $\mathrm{GPM}_z^3, z = 1, \ldots, Z$ for CFD_3 can be built, respectively.

C. CONSTRUCTION OF SGPM

For a new sample of the test subclass, it is important to judge which GPM and CFD models are the most suitable. To this end, a probability index based on the Bayesian method is proposed to evaluate the relationship between a single GPM and a test sample $\mathbf{x}_{t,i}$.

To calculate the probability of each sample $\mathbf{x}_{t,i}$ with each GPM_l^1 or GPM_m^2, GPM_z^3 model, the posterior probability $P\left(\mathrm{GPM}_l^1 \middle| \mathbf{x}_{t,i}\right)$, $P\left(\mathrm{GPM}_m^2 \middle| \mathbf{x}_{t,i}\right)$, and $P\left(\mathrm{GPM}_z^3 \middle| \mathbf{x}_{t,i}\right)$ using the Bayesian inference is proposed. Taking GPM_l^1 as an example, $P\left(\mathrm{GPM}_l^1 \middle| \mathbf{x}_{t,i}\right)$ is calculated as follows [5], [33]:

$$P\left(\mathrm{GPM}_l^1 | \mathbf{x}_{t,i}\right) = \frac{P\left(\mathbf{x}_{t,j} | \mathrm{GPM}_l^1\right) P\left(\mathrm{GPM}_l^1\right)}{P\left(\mathbf{x}_{t,i}\right)} = \frac{P\left(\mathbf{x}_{t,i} | \mathrm{GPM}_l^1\right) P\left(\mathrm{GPM}_l^1\right)}{\sum_{l=1}^{L}\left[P\left(\mathbf{x}_{t,i} | \mathrm{GPM}_l^1\right) P\left(\mathrm{GPM}_l^1\right)\right]}, \quad l = 1, \ldots, L \tag{5}$$

where $P\left(\mathrm{GPM}_l^1\right)$ and $P\left(\mathbf{x}_{t,i} \middle| \mathrm{GPM}_l^1\right)$ are the prior probability and conditional probability, respectively. The prior probability for each GPM_l^1 can be simply defined as follows [5]:

$$P\left(\mathrm{GPM}_l^1\right) = \frac{N_l}{N}, \quad l = 1, \ldots, L \tag{6}$$

where $N = \sum_{l=1}^{L} N_l$ is the number of all training samples. To determine the other terms in Eq. (5), a relative prediction variance item of the test sample for each GPM is defined [5].

$$v_{l,\mathbf{x}_{t,i}} = \frac{\sigma_{\hat{y}_{l,ti}}}{\left|\hat{y}_{l,ti}\right|} \times 100\%, \quad l = 1, \ldots, L \tag{7}$$

where the actual value of $y_{t,i}$ is unknown and it is replaced by its predicted value $\hat{y}_{l,ti}$. The item $\sigma_{\hat{y}_{l,ti}}$ describes the prediction uncertainty for a test sample with this model. The value of $\sigma_{\hat{y}_{l,ti}}$ is relatively large if the test sample $\mathbf{x}_{t,i}$ is predicted with an inapposite model. Consequently, a larger value of $v_{l,\mathbf{x}_{t,i}}$ generally means a larger uncertainty when the GPM_l^1 is utilized for prediction. In such a situation, the conditional probability $P\left(\mathbf{x}_{t,i}\middle|\,\mathrm{GPM}_l^1\right)$ is defined [5].

$$P\left(\mathbf{x}_{t,i}\middle|\,\mathrm{GPM}_l^1\right)=\frac{1}{v_{l,\mathbf{x}_{t,i}}},\quad l=1,\ldots,L \tag{8}$$

Consequently, Eq. (5) becomes:

$$P\left(\mathrm{GPM}_l^1\middle|\,\mathbf{x}_{t,i}\right)=\frac{N_l}{v_{l,\mathbf{x}_{t,i}}\sum_{l=1}^{L}\left(N_l/v_{l,\mathbf{x}_{t,i}}\right)},\quad l=1,\ldots,L \tag{9}$$

Using the probabilistic analysis approach, the GPM_l^1 is more suitable to predict the new test sample $\mathbf{x}_{t,i}$ if the value of $P\left(\mathrm{GPM}_l^1\middle|\,\mathbf{x}_{t,i}\right), l=1,\ldots,L$ is larger. Similarly, $P\left(\mathrm{GPM}_m^2\middle|\,\mathbf{x}_{t,i}\right), m=1,\ldots,M$ and $P\left(\mathrm{GPM}_z^3\middle|\,\mathbf{x}_{t,i}\right), z=1,\ldots,Z$ can be obtained. Consequently, Eq. (9) provides a feasible method to evaluate which GPM is most suitable for a new test sample $\mathbf{x}_{t,i}$.

Generally, in three selected GPM_l^1, GPM_m^2 and GPM_z^3 models for the prediction of the same test sample $\mathbf{x}_{t,i}$, the one with the largest posterior probability index, namely $\max\{\max\left[P\left(\mathrm{GPM}_l^1\middle|\,\mathbf{x}_{t,i}\right)\right],\max\left[P\left(\mathrm{GPM}_m^2\middle|\,\mathbf{x}_{t,i}\right)\right],\max\left[P\left(\mathrm{GPM}_z^3\middle|\,\mathbf{x}_{t,i}\right)\right]\}$ is most suitable to. The corresponding CFD model is also the most appropriate one to describe the test sample $\mathbf{x}_{t,i}$ because it generates an initial set of training data for the most suitable GPM. Thus, the prediction uncertainty of the CFD model can also be obtained using the posterior probability index. As a result, a most suitable GPM among all candidates can be sequentially selected for each sample of the test subclass $\mathbf{X}_t=\left\{\mathbf{x}_{t,i}\right\}_{i=1}^{N_t}$ and thus the temperature characteristics of a discharge process is obtained using the proposed SGPM method. Meanwhile, based on the superiority of different turbulence models in the description of different flow patterns, the selected CFD models can assist the analysis of internal flow fields. The obtained information will assist the designers optimize the structure of the multiphase pumps more efficiently.

In summary, two main stages with several implemented steps of the SGPM-based probabilistic modeling method are illustrated in Fig. 3. As a useful evaluation index, the uncertainty of the GPM and CFD models can be assessed by Eq. (9). From a practical viewpoint, this method can be simply implemented to predict the temperature characteristic of a multiphase pump in different multiphase transportation conditions.

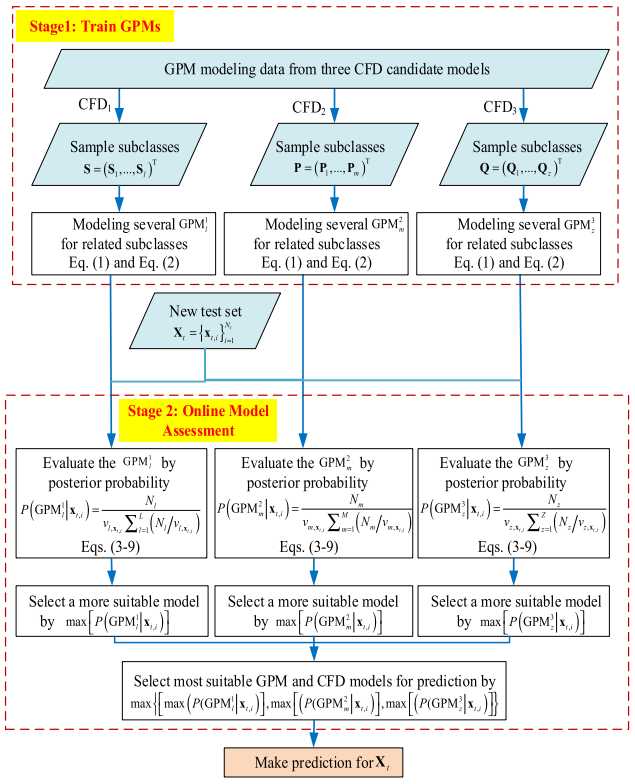


FIGURE 3. The probabilistic modeling method flowchart for the temperature characteristics prediction of the reciprocating multiphase pump.

III. ONLINE PREDICTION OF TEMPERATURE CHARACTERISTICS

A. TRAINING AND TEST SETS SELECTION

Considering different work conditions of the test pump, the CFD simulations are conducted in different pump speeds ($n=100, 120, 148, 160, 180$ r/min), suction pressures ($P_s=0.2, 0.25, 0.3, 0.35, 0.4$ MPa), discharge pressures ($P_d=1.0, 1.5, 2, 2.5, 3.0$ MPa), and gas volume fractions ($\beta=40, 60, 70, 80, 90$ %), respectively. With the change of crank angle ($\theta=180^\circ\sim360^\circ$), the transient temperature values of the pump cavity in a discharge process are obtained from three CFD models, respectively. That is to say, for an example of CFD_1 samples, the lth subclass with N_l samples can be represented as $\mathbf{s}_{l,i}=\left\{\mathbf{x}_{l,i}=\left[n_{l,i},P_{sl,i},P_{dl,i},\beta_{l,i},\theta_{l,i}\right]^{\mathrm{T}},\ y_{l,i}=T_{l,i}\right\}_{i=1}^{N_l}$. About 5760 samples of 15 operational conditions are collected from each CFD model. The former 9 sets (i.e., $\mathbf{S}_1,\ldots,\mathbf{S}_9$) are used for training and the remaining 6 sets (i.e., $\mathbf{S}_{10},\ldots,\mathbf{S}_{15}$) are for test.

Generally, the transient temperature will rise obviously with the decrease of the suction pressure and the increase of the gas volume fraction. To validate the reliability of the proposed method and provide meaningful information for engineering applications, 3 test sets with different multiphase transportation conditions (i.e., S_{10}, S_{11}, S_{12}) are selected for the detailed analysis. And their input varies are $\mathbf{X}_{10}=\left\{\mathbf{x}_{10,i}=[148,0.4,2,40,180+i]^{\mathrm{T}}\right\}_{i=1}^{180}$, $\mathbf{X}_{11}=\left\{\mathbf{x}_{11,i}=[148,0.3,2,60,180+i]^{\mathrm{T}}\right\}_{i=1}^{180}$, and $\mathbf{X}_{12}=\left\{\mathbf{x}_{20,i}=[148,0.2,2,80,180+i]^{\mathrm{T}}\right\}_{i=1}^{180}$, respectively.

TABLE 1. The posterior probability index comparisons of 9 GPMs for the 20th test sample with $\theta = 200°$ (The selected GPM with the largest posterior probability index is bolded and underlined).

No.		Posterior probability index								
		1	2	3	4	5	6	7	8	9
S_{10}	GPM_l^1	0.153	0.286	0.008	0.022	**0.480**	0.028	0.022	0	0.001
	GPM_m^2	0.080	0	0.003	**0.768**	0.145	0.001	0	0	0.003
	GPM_z^3	0.192	0.124	0.006	0.143	0.161	0.001	0.002	**0.470**	0.001
S_{11}	GPM_l^1	0.020	0.114	0.174	0.002	0.060	0.002	0.259	**0.350**	0.019
	GPM_m^2	0.022	0.241	**0.266**	0.003	0.133	0.003	0.072	0.245	0.015
	GPM_z^3	0.021	0.198	0.122	0.203	0.127	0.002	0.003	**0.306**	0.018
S_{12}	GPM_l^1	0.051	0.028	0.029	**0.579**	0.233	0.018	0.019	0.016	0.018
	GPM_m^2	0.191	0.011	0.022	**0.453**	0.245	0.012	0.020	0.002	0.045
	GPM_z^3	0.096	0	0.007	**0.853**	0.040	0	0	0	0.003
S_{13}	GPM_l^1	0.029	**0.698**	0.148	0.041	0.002	0.075	0.002	0.004	0.001
	GPM_m^2	0.065	**0.631**	0.130	0.014	0.068	0.086	0.002	0.003	0.001
	GPM_z^3	0.091	**0.801**	0.028	0.006	0.003	0.063	0.001	0.002	0.005
S_{14}	GPM_l^1	0.014	0.056	0.008	0.209	0.091	0.022	0.155	0.021	**0.425**
	GPM_m^2	0.008	0.012	**0.578**	0.234	0.017	0.019	0.107	0.009	0.016
	GPM_z^3	0.031	0.019	0.011	**0.734**	0.060	0.005	0.029	0.001	0.110
S_{15}	GPM_l^1	0.022	0.221	0.066	0.103	0.033	0.003	0.072	**0.470**	0.010
	GPM_m^2	0.116	0.009	0.020	**0.361**	0.195	0.010	0.017	0.269	0.003
	GPM_z^3	0.031	0.008	0.010	0.223	0.018	0.019	0.007	**0.675**	0.009

A common performance index, namely the relative root-mean-square error (simply denoted as RE), is adopted to evaluate the proposed method. For the t-th test subclass, RE_t can be defined as follows

$$RE_t = \sqrt{\sum_{i=1}^{N_t} \left[\left(\hat{y}_{t,i} - y_{t,i}\right)/y_{t,i}\right]^2 / N_t} \times 100\% \ , t = 1, \ldots, T \tag{10}$$

where $y_{t,i}$ comes from the CFD model, $\hat{y}_{t,i}$ denotes the prediction of $y_{t,i}$, and N_t is the sample number of the t-th test subclass.

B. RESULTS AND DISCUSSION

As shown in Fig. 3, each test sample can automatically select its most reliable GPM for the prediction based on the posterior probability index, respectively. Taking the 20th sample with $\theta = 200°$ of S_{10} as an example, its posterior probability indices predicted by 9 GPMs are listed in Table 1. It shows that the posterior probability index of GPM_5^1 is larger than other GPMs, which are trained by the data from CFD_1. Consequently, GPM_5^1 is the most suitable model for online prediction of the 20th sample of S_{10}. Similarly, GPM_4^2 and GPM_8^3 trained by the data from CFD_2 and CFD_3, are the most appropriate ones for the 20th sample of S_{10}, respectively. Moreover, the posterior probability index of GPM_4^2 is larger than the ones of GPM_5^1 and GPM_8^3. In such a situation, GPM_4^2 and corresponding CFD_2 are selected as the most suitable GPM and CFD models for the predictions of the 20th sample of S_{10}, respectively. Similarly, as also tabulated in Table 1, using the posterior probability indices of the 20th sample, suitable GPM and CFD models for $\mathbf{S}_{11}, \ldots, \mathbf{S}_{15}$ are chosen as follows: GPM_8^1 and CFD_1 for S_{11}, GPM_4^3 and CFD_3 for S_{12}, GPM_2^3 and CFD_3 for S_{13}, GPM_4^3 and CFD_3 for S_{14}, GPM_8^3 and CFD_3 for S_{15}, respectively.

As an illustrated case, compared with the simulation result of corresponding CFD models, the RE values of the 20th sample of S_{10} predicted by 9 GPMs are listed in Table 2. The RE value of GPM_5^1 is smaller than that of other models trained by the data from the CFD_1 model, which indicates GPM_5^1 can obtain better prediction performance. Similarly, the smaller RE values of GPM_4^2 and GPM_8^3 trained by the data from the CFD_2 and CFD_3 models show they also predict better than other GPMs. Additionally, the RE value of GPM_4^2 is smaller than that of GPM_5^1 and GPM_8^3. This indicates that GPM_4^2 is more suitable than GPM_5^1 and GPM_8^3 for online prediction of the 20th sample of S_{10}. Consequently, the corresponding CFD_2 model is validated as the most appropriate one for prediction of the 20th sample of S_{10}. Similarly, as also listed in Table 2, GPM_8^1 and corresponding CFD_1 GPM_4^3 and corresponding CFD_3 GPM_2^3 and corresponding CFD_3 GPM_4^3 and corresponding CFD_3 GPM_8^3 and corresponding CFD_3, can be validated as the most appropriate models for prediction of the 20th sample of $\mathbf{S}_{11}, \ldots, \mathbf{S}_{15}$, respectively.

TABLE 2. The RE index comparisons of 9 GPMs for the 20th test samples with $\theta = 200°$ (The selected GPM with the smallest RE index is bolded and underlined).

No.		RE index (%)								
		1	2	3	4	5	6	7	8	9
S_{10}	GPM_l^1	1.467	0.718	2.297	4.441	**0.382**	1.559	1.472	3.090	2.317
	GPM_m^2	0.571	1.443	2.089	**0.036**	0.482	1.794	0.585	26.794	4.062
	GPM_z^3	2.012	4.133	5.686	3.492	2.947	5.293	3.998	**1.208**	7.753
S_{11}	GPM_l^1	4.588	3.694	2.173	3.830	4.263	4.259	2.002	**1.795**	5.749
	GPM_m^2	8.025	2.318	**2.198**	4.505	2.823	3.141	3.614	2.604	3.131
	GPM_z^3	4.225	2.659	5.017	2.341	3.076	3.106	3.525	**1.997**	7.958
S_{12}	GPM_l^1	2.113	3.942	3.744	**1.954**	1.995	8.135	8.095	15.456	8.101
	GPM_m^2	2.915	4.954	3.506	**2.055**	2.805	4.209	3.542	28.615	3.085
	GPM_z^3	1.710	10.220	4.058	**0.170**	1.856	5.984	5.907	5.055	4.954
S_{13}	GPM_l^1	3.406	**0.102**	0.769	4.409	5.045	3.306	2.220	12.975	6.161
	GPM_m^2	6.807	**0.129**	3.488	6.902	9.328	5.144	3.655	19.570	9.529
	GPM_z^3	0.645	**0.030**	3.586	4.249	6.522	2.424	4.025	24.855	5.530
S_{14}	GPM_l^1	4.745	4.792	3.385	1.266	3.215	4.310	1.913	5.697	**1.242**
	GPM_m^2	6.217	5.789	**1.187**	2.572	3.174	3.429	2.792	4.711	4.911
	GPM_z^3	2.480	4.793	15.719	**1.005**	2.123	7.216	3.747	17.207	1.858
S_{15}	GPM_l^1	4.006	0.448	2.633	0.672	3.735	5.443	1.323	**0.150**	4.172
	GPM_m^2	8.687	15.486	12.599	**1.245**	7.446	13.368	12.797	4.518	15.870
	GPM_z^3	10.963	16.020	11.410	5.150	11.733	11.170	23.716	**0.108**	12.880

For a new sample in different multiphase transportation conditions, which model is more appropriate to describe its characteristics is unknown before the actual value is available. All the prediction results listed in Table 1 and Table 2 validate that the posterior probability index suits to evaluate the GPM-based and CFD candidate models. Generally, a GPM-based candidate model with a larger posterior probability value can have a smaller RE value when it is applied for online prediction of a new sample. Thus, a CFD candidate model, generating training data for a selected GPM, can better describe the transient characteristics.

TABLE 3. The selection results of the SGPM for all test subclasses.

Test subclass	GPM_l^1	GPM_m^2	GPM_z^3
$\mathbf{S}_{10}$	154	26	0
$\mathbf{S}_{11}$	175	1	4
$\mathbf{S}_{12}$	156	1	23
$\mathbf{S}_{13}$	146	34	0
$\mathbf{S}_{14}$	166	3	11
$\mathbf{S}_{15}$	143	2	35

Consequently, a most suitable GPM among all candidates can be sequentially selected for each sample of all test sets. The selection results for 6 test sets (i.e., $\mathbf{S}_{10}, \ldots, \mathbf{S}_{15}$) are listed in Table 3. It indicates that, if only using a single turbulence model, RNG k-ε may be the most appropriate one for the description of 6 test sets, mainly because most of GPM_l^1 are selected for the online prediction. However, only using a single turbulence model is often not enough. With the decrease of the suction pressure and the increase of the gas volume fraction, more and more samples are captured using the standard k-ε turbulence model for the increasing selection of GPM_z^3. Consequently, several GPMs, trained using the data from different CFD models, can be integrated to better track the temperature characteristics of a whole discharge process in different multiphase transportation conditions.

As aforementioned, the SGPM can integrate better prediction results for the discharge process. To further analyze the internal flow field of a discharge process, the prediction results for different samples of the SGPM are denoted as $SGPM_{RNG}$, $SGPM_{SST}$, and $SGPM_{STA}$, respectively. The $SGPM_{RNG}$, $SGPM_{SST}$ and $SGPM_{STA}$ exhibit that the SGPM is obtained by the data from the CFD_1, CFD_2, and CFD_3 models, respectively. Their detailed prediction results of S_{10}, S_{11}, S_{12}, obtained by the SGPM and GPMs trained by the data from three CFD models, are compared with their corresponding CFD test data (i.e., CFD_{MIX}, $CFDR_{NG}$, CFD_{SST}, CFD_{STA}), respectively. Notice that CFD_{MIX} means that it has several turbulence models during the discharge process.

The RE indices in Table 4 show that SGPM obtains better prediction performance than three GPMs for S_{10}. Its detailed prediction results shown in Fig. 4 also indicate that it can better track the main characteristics of S_{10}. This implies the SGPM based on the proposed posterior probability index

TABLE 4. The RE index comparisons of the SGPM, GPM prediction models and their CFD models for S_{10}, S_{11}, and S_{12}. (The smallest RE index is bolded and underlined).

	RE index (%)		
Model	S_{10}	S_{11}	S_{12}
GPM^1	0.369	1.231	12.498
GPM^2	0.624	2.520	11.006
GPM^3	0.834	4.207	7.719
SGPM	**0.032**	**0.610**	**5.215**

can replace costly experiments and complicated CFD modeling processes to realize the prediction of S_{10}. Additionally, 26 samples in the opening lag stage of the discharge valve corresponding to the rising stage select GPM_m^2 as the better prediction models. And the remaining 154 samples mainly in the opening stage of the discharge valve corresponding to the declining and steady stages choose GPM_l^1 as the better ones. Due to the superiority of the SST k-ω turbulence model in the description of free flows near the walls, there may be a large amount of gas flows near the discharge valve in the opening lag stage [19]. Similarly, a lot of vortex flows appear in the opening stage of the discharge valve for the selected RNG k-ε turbulence model. The main reason may be that the mixture containing continually and highly compressed gas flows out of the pump rapidly to reach the maximum flow rate, and then the flow rate reduces and vibrates for the opening lag characteristics of the discharge valve [4]–[6]. Consequently,

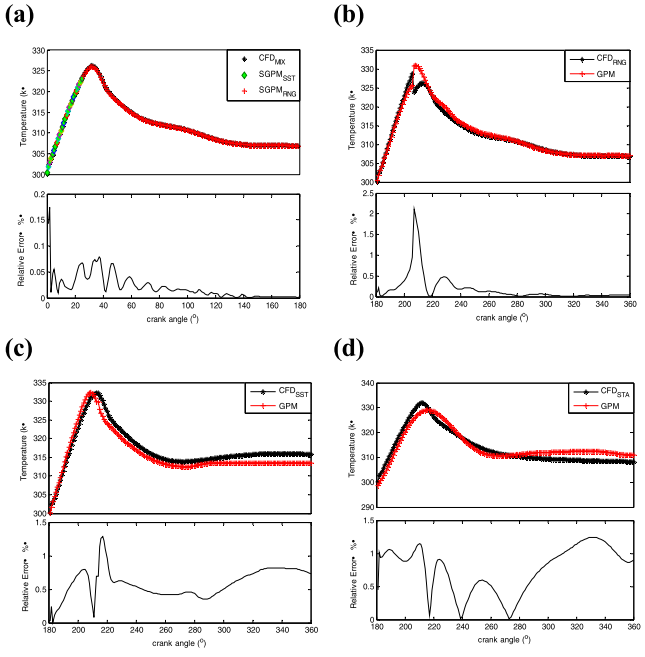


FIGURE 4. (a) Online prediction results of the test subclass S_{10} with SGPM (b) Online prediction results of the test subclass S_{10} with GPM trained by the data from the CFD RNG k-ε transient model (c) Online prediction results of the test subclass S_{10} with GPM trained by the data from the CFD SST k-ω transient model (d) Online prediction results of the test subclass S_{10} with GPM trained using the data from the CFD.

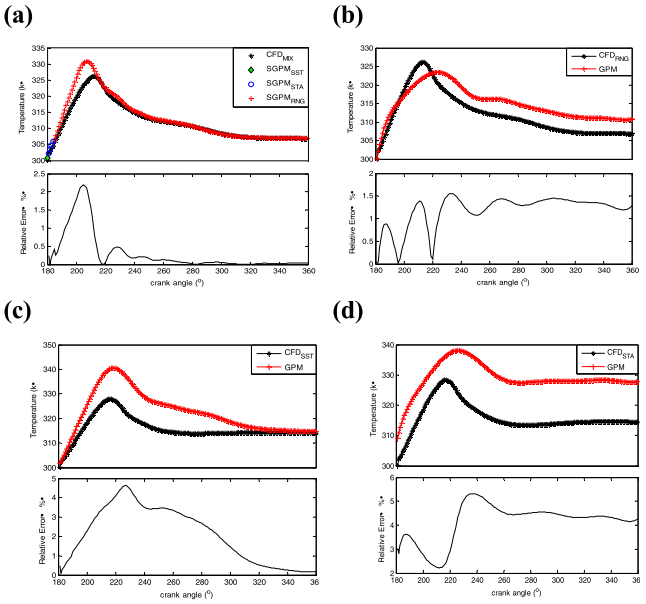


FIGURE 5. (a) Online prediction results of the test subclass S_{11} with SGPM (b) Online prediction results of the test subclass S_{11} with GPM trained by the data from the CFD RNG k-ε transient model (c) Online prediction results of the test subclass S_{11} with GPM trained by the data from the CFD SST k-ω transient model (d) Online prediction results of the test subclass S_{11} with GPM trained using the data from the CFD.

the selection results of GPMs can provide useful information for the recognition of complicated flow patterns.

Similarly, the RE indices listed in Table 4 show SGPM exhibits better prediction performance of S_{11} than other GPMs. Its detailed prediction results shown in Fig. 5 also indicate it tracks the main characteristics of S_{11} more suitably. Additionally, most of samples select GPM_l^1 as the better prediction models, except for 5 samples in the opening lag stage of the discharge valve. And one of 5 samples chooses GPM_m^2 as the better one, the remaining 4 samples adopt GPM_z^3. This indicates that new flow patterns may be generated except for free flows near the walls and vortex flows, and they can be better tracked using the standard k-ε turbulence model. Moreover, compared with S_{10}, the operating condition of S_{11} presents relatively small suction pressure $P_s = 0.3$ MPa and high gas volume fraction $\beta = 60\%$. Thus, the resulting larger opening lag angle of the discharge valve cause more complicated internal flow.

Similarly, as shown in Table 4 and Fig. 6, SGPM can be further validated as the most appropriate model for the prediction of S_{12}. In view of the smaller suction pressure $P_s = 0.2$ MPa and the higher gas volume fraction $\beta = 80\%$ for the operating condition of S_{12}, most of samples still select GPM_l^1 as the better prediction models, except for 24 samples in the opening lag stage of the discharge valve. One of 24 samples chooses GPM_m^2 as the better one, and the remaining 23 samples select GPM_z^3.

From the above analysis, it can be summarized that SGPM has better prediction results for all test sets. This implies that it is difficult to capture all information with only a single

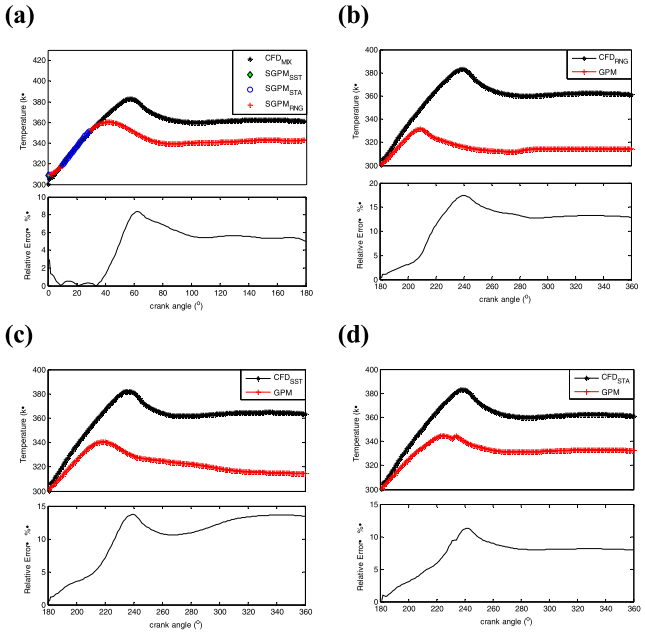


FIGURE 6. **(a) Online prediction results of the test subclass S_{12} with SGPM (b) Online prediction results of the test subclass S_{12} with GPM trained by the data from the CFD RNG k-ε transient model (c) Online prediction results of the test subclass S_{12} with GPM trained by the data from the CFD SST k-ω transient model (d) Online prediction results of the test subclass S_{12} with GPM trained using the data from the CFD standard k-ε transient model.**

GPM model. Similarly, it also indicates that a single CFD turbulence model is inadequate for depicting the temperature characteristics of multiple multiphase transportation conditions and different stages of a discharge process. Additionally, most of samples select GPM_l^1 as the better prediction models, except for a few samples in the opening lag stage of the discharge valve. With the decrease of the suction pressure and the increase of the gas volume fraction, more and more samples in the opening lag stage choose GPM_z^3 as the better ones. Due to the turbulence models having their reliable applications in the description of different flow patterns, the selection results of GPMs can also help explore the complicated internal flows.

Although more computational resources and time (hours to days) are needed for several CFD modeling processes, the proposed method can better describe the main temperature characteristics of new conditions than a single CFD model. The prediction time of SGPM is much less than CFD. Consequently, the proposed SGPM method shows better prediction performance compared with only using a GPM and more efficient implementations than the traditional CFD modeling method.

IV. CONCLUSION

A probabilistic modeling method is proposed to predict the temperature characteristics for a dynamic discharge process of reciprocating multiphase pumps. Its main advantages can be summarized in three aspects. First, using the proposed posterior probability index, suitable GPM and CFD models are automatically selected for a new sample without knowing its actual value. Second, compared with only using a single GPM, SGPM can better describe the main temperature characteristics of new conditions. Third, for practical applications, the selection results of GPM and CFD models can provide useful information for better recognition of complicated flow patterns in the pump cavity.

One of our future research directions is to construct a model for better description of the characteristics near the maximum temperature of dynamic fluid processes. Application of deep neural networks [22], [23], [25], [26] to dynamic fluid processes is also an interesting topic.

REFERENCES

[1] J. Falcimaigne and S. Decarre, *Multiphase Production: Pipeline Transport, Pumping and Metering, Editions Technip*. Paris, France: 2008.

[2] G. Hua, G. Falcone, C. Teodoriu, and G. L. Morrison, "Comparison of multiphase pumping technologies for subsea and downhole applications," *Oil Gas Facilities*, vol. 1, pp. 36–46, Feb. 2012.

[3] A. H. Dogru, A. A. Hamoud, and S. G. Barlow, "Multiphase pump recovers more oil in a mature carbonate reservoir," *J. Petroleum Technol.*, vol. 56, pp. 64–67, Feb. 2004.

[4] H. Deng, Y. Liu, P. Li, and S. Zhang, "Active learning for modeling and prediction of dynamical fluid processes," *Chemometrics Intell. Lab. Syst.*, vol. 183, pp. 11–22, Dec. 2018.

[5] H. Deng, Y. Liu, P. Li, Y. Ma, and S. Zhang, "Integrated probabilistic modeling method for transient opening height prediction of check valves in oil-gas multiphase pumps," *Adv. Eng. Softw.*, vol. 118, pp. 18–26, Apr. 2018.

[6] H. Deng, Y. Liu, P. Li, and S. Zhang, "Hybrid model for discharge flow rate prediction of reciprocating multiphase pumps," *Adv. Eng. Softw.*, vol. 124, pp. 53–65, Oct. 2018.

[7] J. Duan, J. Gong, H. Yao, T. Deng, and J. Zhou, "Numerical modeling for stratified gas–liquid flow and heat transfer in pipeline," *Appl. Energy*, vol. 115, pp. 83–94, Feb. 2014.

[8] J. Wang, H. Zha, J. M. Mcdonough, and D. Zhang, "Analysis and numerical simulation of a novel gas–liquid multiphase scroll pump," *Int. J. Heat Mass Transf.*, vol. 91, pp. 27–36, Dec. 2015.

[9] C. Y. Nakashima, S. de Oliveira, Jr., and E. F. Caetano, "Heat transfer in a twin-screw multiphase pump: Thermal modeling and one application in the petroleum industry," *Energy*, vol. 31, no. 15, pp. 3415–3425, 2006.

[10] X. Yang, C. Hu, Y. Hu, and Z. Qu, "Theoretical and experimental study of a synchronal rotary multiphase pump at very high inlet gas volume fractions," *Appl. Therm. Eng.*, vol. 110, pp. 710–719, Jan. 2017.

[11] O. T. Kajero, T. Chen, Y. Yao, Y.-C. Chuang, and D. S. H. Wong, "Metamodelling in chemical process system engineering," *J. Taiwan Inst. Chem. Engineers*, vol. 73, pp. 135–145, Apr. 2017.

[12] A. Patil, "Performance evaluation and CFD simulation of multiphase twin-screw pumps," Ph.D. dissertation, Mech. Eng., Texas A&M Univ., Texas, TX, USA, 2013.

[13] U. Kadri, R. F. Mudde, and R. V. A. Oliemans, "Influence of the operation pressure on slug length in near horizontal gas–liquid pipe flow," *Int. J. Multiphase Flow*, vol. 36, pp. 423–431, May 2010.

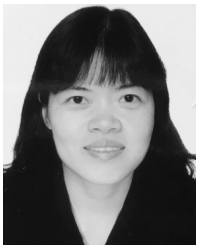
[14] Y. Zhao, G. Chen, C. Ye, and Q. Yuan, "Gas–liquid two-phase flow in microchannel at elevated pressure," *Chem. Eng. Sci.*, vol. 87, pp. 122–132, Jan. 2013.

[15] L. Liu, "A new method for evaluating drag reduction in gas–liquid two-phase flow based on energy dissipation," *Chem. Eng. Sci.*, vol. 95, pp. 54–64, May 2013.

[16] A. C. Bannwart, O. M. H. Rodriguez, F. E. Trevisan, F. F. Vieira, and C. H. M. de Carvalho, "Experimental investigation on liquid–liquid–gas flow: Flow patterns and pressure-gradient," *J. Petroleum Sci. Eng.*, vol. 65, pp. 1–13, Mar. 2009.

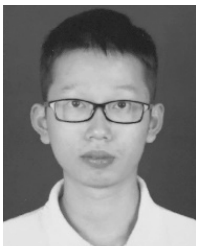
[17] J.-W. Suh, J.-W. Kim, Y.-S. Choi, J.-H. Kim, W.-G. Joo, and K.-Y. Lee, "Development of numerical Eulerian-Eulerian models for simulating multiphase pumps," *J. Petroleum Sci. Eng.*, vol. 162, pp. 588–601, Mar. 2018.

[18] J. H. Kim, H. C. Lee, J. H. Kim, Y. K. Lee, and Y. S. Choi, "Reliability verification of the performance evaluation of multiphase pump," *Int. J. Mech. Aerosp. Ind. Mech.*, vol. 24, pp. 1782–1786, Nov. 2014.
[19] R. Löhner, *Applied Computational Fluid Dynamics Techniques: An Introduction Based on Finite Element Methods*. Hoboken, NJ, USA: Wiley, 2008.
[20] 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.
[21] Y. Yao and F. Gao, "A survey on multistage/multiphase statistical modeling methods for batch processes," *Annu. Rev. Control*, vol. 33, no. 2, pp. 172–183, 2009.
[22] Y. Liu, Y. Fan, and J. Chen, "Flame images for oxygen content prediction of combustion systems using DBN," *Energy Fuels*, vol. 31, no. 8, pp. 8776–8783, 2017.
[23] Y. Liu, C. Yang, Z. Gao, and Y. Yao, "Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes," *Chemometrics Intell. Lab. Syst.*, vol. 174, pp. 15–21, Mar. 2018.
[24] 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.
[25] Q. Xuan *et al.*, "Automatic pearl classification machine based on a multistream convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 65, pp. 6538–6547, Aug. 2018.
[26] Q. Xuan, Z. Z. Chen, Y. Liu, H. M. Huang, G. J. Bao, and D. Zhang, "Multiview generative adversarial network and its application in pearl classification," *IEEE Trans. Ind. Electron.*, to be published. doi: 10.1109/TIE.2018.2885684.
[27] Y. Liu and J. Chen, "Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multigrade processes," *J. Process Control*, vol. 23, no. 6, pp. 793–804, 2013.
[28] W. J. Zheng, Y. Liu, Z. Gao, and J. Yang, "Just-in-time semi-supervised soft sensor for quality prediction in industrial rubber mixers," *Chemometrics Intell. Lab. Syst.*, vol. 180, pp. 36–41, Sep. 2018.
[29] J. Kocijan and B. Likar, "Gas-liquid separator modelling and simulation with Gaussian process models," *Simul. Model. Pract. theory*, vol. 16, pp. 910–922, Sep. 2008.
[30] T. Chen and J. Ren, "Bagging for Gaussian process regression," *Neurocomputing*, vol. 72, pp. 1605–1610, Mar. 2009.
[31] Z. Ge, T. Chen, and Z. Song, "Quality prediction for polypropylene production process based on CLGPR model," *Control Eng. Pract.*, vol. 19, no. 5, pp. 423–432, 2011.
[32] H. Jin, X. Chen, L. Wang, K. Yang, and L. Wu, "Adaptive soft sensor development based on online ensemble Gaussian process regression for nonlinear time-varying batch processes," *Ind. Eng. Chem. Res.*, vol. 54, no. 30, pp. 7320–7345, Oct. 2015.
[33] Y. Liu, T. Chen, and J. Chen, "Auto-switch Gaussian process regression-based probabilistic soft sensors for industrial multigrade processes with transitions," *Ind. Eng. Chem. Res.*, vol. 54, no. 18, pp. 5037–5047, 2015.
[34] L. Zhou, J. Chen, and Z. Song, "Recursive Gaussian process regression model for adaptive quality monitoring in batch processes," *Math. Problems Eng.*, vol. 2015, 2015, Art. no. 761280.
[35] Y. Liu, Q.-Y. Wu, and J. Chen, "Active selection of informative data for sequential quality enhancement of soft sensor models with latent variables," *Ind. Eng. Chem. Res.*, vol. 56, no. 16, pp. 4804–4817, Jun. 2017.
[36] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.



HONGYING DENG received the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2019. She is currently a Lecturer with the Zhejiang University of Technology, Hangzhou. She has published several international journal papers. Her research interests include modeling, control, and optimization of industrial processes.

YANG ZHANG received the bachelor's degree from the College of Mechanical Engineering, North China University of Water Resources and Electric Power, in 2016. He is currently pursuing the master's degree with the College of Mechanical Engineering, Zhejiang University of Technology. His research interest includes the design and optimization of reciprocating multiphase pumps.



BOCHENG CHEN received the bachelor's degree from the College of Mechanical Engineering, University of South China, in 2016. He is currently pursuing the master's degree with the College of Mechanical Engineering, Zhejiang University of Technology. His research interests include generative adversarial networks, deep learning algorithms, and data mining.

YI LIU (M'11) received the Ph.D. degree in control theory and engineering from Zhejiang University, Hangzhou, China, in 2009. From 2012 to 2013, he was a Postdoctoral Researcher with the Department of Chemical Engineering, Chung Yuan Christian University. He is currently an Associate Professor with the Zhejiang University of Technology, Hangzhou. He has published more than 30 international journal papers. His research interests include data intelligence with applications to modeling, control, and optimization of industrial processes.

SHENGCHANG ZHANG received the bachelor's degree in hydraulic machinery from the Lanzhou University of Technology, Lanzhou, China, in 1981. He is currently a Professor with the Zhejiang University of Technology, Hangzhou, China. He has published about 20 international journal papers and holds 35 patents. His research interest includes the design and optimization of multiphase pumps.