Optimized Forecast Components-SVM-Based Fault Diagnosis With Applications for Wastewater Treatment

HONGCHAO CHENG[1], YIQI LIU[ID][1], (Member, IEEE), DAOPING HUANG[1], AND BIN LIU[ID][2], (Member, IEEE)

[1]School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China
[2]Department of Management Science, University of Strathclyde, Glasgow G1 1XQ, U.K.

Corresponding authors: Yiqi Liu (aulyq@scut.edu.cn) and Daoping Huang (audhuang@scut.edu.cn)

ABSTRACT Process monitoring of wastewater treatment plant (WWTP) is a challenging industrial problem, due to its exposure to the hostile working environment and significant disturbances. This paper proposed a novel fault diagnosis method, termed as optimization forecast components-support vector machine (OFC-SVM). The method firstly improved the forecastable component analysis (ForeCA) for feature extraction. Secondly, in order to further enhance the method, the quadratic Grid Search (GS) algorithm is utilized to optimize the parameters of the proposed method. Thirdly, to properly evaluate the method performance, a new evaluation index is proposed, named Pre Alarm Rate (PAR), aiming to achieve the quantitative trade-off between false alarm rate (FAR) and missed alarm rate(MAR). Then, the new ROC curve can be further derived by PAR. Finally, the performance of OFC-SVM is strictly compared with other five methods as well as validated by a Monte Carlo model and a full-scale WWTP.

INDEX TERMS Fault diagnosis, grid search (GS), feature extraction, forecastable component analysis, support vector machine, wastewater treatment.

I. INTRODUCTION

With the increasing complexity of industrial system, process monitoring strategy of industrial process have become a hotspot [1]–[3]. WWTP is a complex industrial system with mixture of physical, chemical and biological reactions. Also, the hostile working environment and significant disturbances further add further difficulty for process monitoring. If not monitored well, a WWTP will not only bring about economic losses, but also cause the secondary pollution of rivers. So we conduct an in-depth study on monitoring the common sensor fault in wastewater treatment plants. Data-driven process monitoring is one of the most popular industrial monitoring methods, because it does not require the prior knowledge and can achieve better performance by comparison with mathematical models [4]. During recent decades, SVM receives more and more attentions, which is based on the theory of VC (Vapnik-Chervonenkis Theory). SVM can not only

trade-off the complexity and learning capability in the case of exposing to small sample [5], but also can use structural risk minimization to effectively compensate for the over-fitting problem [6]. Relying on the above advantages and excellence generalization ability, SVM has quickly attracting the researcher attention, thus resulting in successfully application to the diagnosis of abnormal events in machinery, medicine and other fields subsequently [7]–[10]. Although SVM has been successfully applied in many different fields, it is rarely used in wastewater treatment processes monitoring. Also, uncertain disturbances, large time-delay and multi-variable coupling make building an effective SVM model difficult, mainly due to the fact that SVM is sensitive to noises. In addition, process redundant features will not only greatly reduce the diagnostic accuracy, but also increase the storage overload.

In the feature-based fault diagnosis, it is necessary to consider model performance degradation caused by useless features [6]. Chang et al. proposed to use PCA to extract effective features [11]. However, industrial data are generally

---

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

non-linear, and PCA cannot effectively extract the non-linear characteristics. Zhang et al. proposed that adaptive KPCA to extract useful features of nonlinear data, and then combined with SVM for fault diagnosis of high-voltage circuit breaker [9]. Although KPCA can extract the non-linear characteristics, the diagnosis performance will degrade when the data do not follow Gaussian distribution. Cao et al. compared the performance of SVM, PCA-SVM, KPCA-SVM and ICA-SVM, showing that effective feature extraction can indeed improve SVM performance significantly [12]. In this light, a new feature extraction method is proposed by improved the ForeCA, and then combine it with SVM to diagnose the fault of WWTPs. ForeCA is a novel dimension reduction technique able to take full use of temporally dependent signals [13]. This method firstly transforms the time domain signal into frequency domain signal, and then uses Shannon entropy to reformulate the optimization problem. Finally, the optimization problem is solved by EM-like (Expectation Maximization-like) algorithm. Due to this property, both of time and frequency domains can be taken into account to make the feature extraction sensitively. Also, ForeCA is able to exploit the unpredictability of white noise which is usually hided by other disturbances. By proper recognition of these underlying noises, noises and disturbances can be removed more efficiently. Traditionally, PCA is used to measure the information according to the size of variance with the philosophy that the larger the variance, the more information. However, the variance is mainly to describe the degree of discreteness, and cannot be directly used to capture the uncertainty of information. Unlike the previous methods, ForeCA is capable of measuring the information uncertainty by information entropy and is not premised on the assumption that the data follows Gaussian distribution [14], thus making feature extraction more efficiently.

Typically, the free parameters (penalize parameter C and kernel parameter g) are set up manually, which makes it converge into sub-optimal performance. Recently, to further improve the performance of a diagnosis model, genetic algorithm (GA) and particle swarm optimization (PSO) are widely used to optimize the model parameters [10], [15], [16]. However, PSO and GA usually locate the optimal value through iterative search procedure. This procedure could not only lead to unacceptable uncertainty, but also easily come to a pseudo-global optimal solution. In this light, this paper introduces the quadratic GS algorithm to address this issue, which can generate the corresponding "grid" by arrangement and combination of parameters. And use the cross-validation method to select free parameters. Unlike previous GS, Quadratic GS sets the initial parameter interval twice by trading off the running time and diagnosis accuracy. By setting the parameters by interval rather than point sampling according to the specific data prior knowledge, the consumed time of optimal Quadratic GS will be greatly reduced and achieve better performance.

Based on the above discussion, ForeCA algorithm and quadratic GS algorithm can effectively improve the SVM performance. So we proposed a new method of OFC-SVM by combining the above two algorithm.

Industrial processes often evaluate the performance by rating missed alarm rate (MAR) and false alarm rate (FAR) [1], [17]–[20]. Although these indicators can assess the monitoring methods from bi-direction, the costs of false alarm and missed alarm are different. Therefore, a novel comprehensive evaluation index of pre-alarm rate (PAR) is proposed by merging false alarm and missed alarm. The weight parameter of PAR formulas can set by requirements. At the same time, a new ROC (Receiver Operating Characteristic) curve is further re-derived, which is aim to assist PAR method to make decision.

The structured of this paper as follows. In section II, the basic theory of SVM is introduced. Section III presents the framework and the theories of the proposed method. In section IV, OFC-SVM is strictly compared with five methods and validated by two experimental data sets. The experimental results are analyzed and discussed in detail in section V and the conclusions are come to Section VI.

II. SUPPORT VECTOR MACHINE

In the 1860s, Vapnik et al. explored the relationship between the large number theorem of general function space and the learning process. And then proposed the SVM model [21]. The essence of SVM in the field of fault diagnosis is the use of its classification capability. We assumed that the training data set of $\tilde{X} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T \in R^{n*k}$ is linearly separable, n represents the number of samples, k represents the number of features. In the training set, $y_i = +1$ respresents the fault data label, whereas $y_i = -1$ represents the normal work condition data label. The SVM method needed to use the training set to find the optimal classification hyperplane, in such a way that the test set can be accurately classified sequentially. Assuming that the classification hyperplane is $(\bar{w}, b)$, and the corresponding classification formula is shown as follows:

$$\begin{cases} \bar{w}^T x_i + b \geq +1 & y_i = +1 \\ \bar{w}^T x_i + b \leq -1 & y_i = -1 \end{cases} \tag{1}$$

where $\bar{w}^T$ is normal vector and b is a scalar. If the sample is closest to the hyperplane, named "support vector", and the "margin" is formulated as $r = 2/\|\bar{w}\|$. To find the maximum spacing by resorting to the hyperplane, the above solution becomes an optimization problem.

$$\max_{\bar{w},b} \frac{2}{\|\bar{w}\|}$$
$$s.t. \ y_i(\bar{w}^T x_i + b) \geq 1 \tag{2}$$

When there is a linear inseparable point in the sample, we can relax the constraints and formulate the optimized equations as (3).

$$\min_{\bar{w},b,\xi} \frac{1}{2}\|\bar{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$
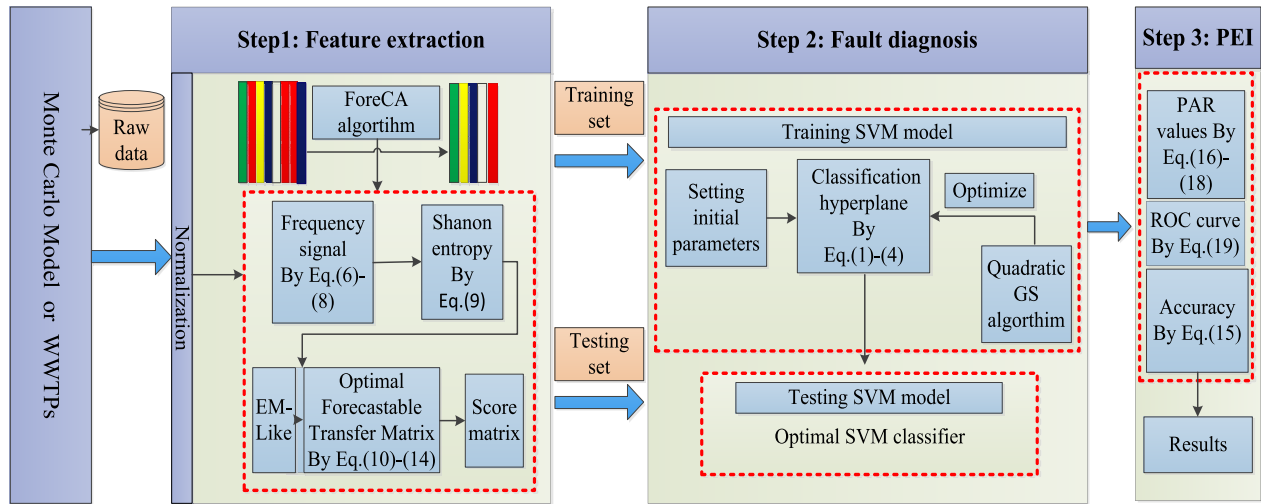$$s.t. \ y_i\left(\bar{w}^T x_i + b\right) \geq 1 - \xi_i \quad (\xi_i \geq 0) \tag{3}$$

FIGURE 1. Process monitoring framework based on the proposed OFC-SVM method.

ξ_i is the relaxation parameter and C is the regularization parameter. When the data is not linearly separable, the kernel function $K\left(x_i, x_j\right) = \emptyset\left(x_i\right)^T \emptyset\left(x_i\right)$ is used to replace x_i. So the classification decision function formula can be further get as follows:

$$\text{f}(\text{x}) = sign\left(\sum_{i=1}^{n} y_i a_i K\left(x, x_j\right) + b\right) \quad (4)$$

where $0 \le a_i \le C$, x represent the input vector.

III. OFC-SVM THEORY AND MONITORING FRAMEWORK
Generally, SVM is suitable to small samples issue but is very sensitive to noise [22]. Additionally, redundant feature information could seriously affect the SVM performance. Thus, we proposed a novel fault monitoring framework, termed OFC-SVM.

As shown in Fig. 1, the main steps can be summarized as follows: (1) Data sampling; (2) Normalization; (3) Feature extraction (As showed in algorithm 1); (4) Fault diagnosis (Model training and testing); (5) Evaluating model by PEI (Performance Evaluation Index) PAR.

A. FORECASTABLE COMPONENT ANALYSIS (ForeCA)
ForeCA is a new statistical signal processing method, which mixes the advantages of Fourier transform, spectral density, shannon entropy and EM-like theory. The core of ForeCA is to find the optimal forecastable transfer matrix W. Assuming that the collected data set is $X \in R^{n \times m}$ (*n* and *m* represent the number of samples and the variables number, respectively), and $W \in R^{m \times m}$ is the optimal transfer matrix of forecast components.

$$Y = WX^T \quad (5)$$

where $Y^T \in R^{n \times m}$ is the corresponding score matrix, and the ForeCA algorithm needs to estimate the optimal W and Y by the data set X. Goerg pointed out in his paper that the first step is to transform the time domain signal into the frequency

---

Algorithm 1 Forecastable Component Analysis

Input: Training dataset $X_{train} = \left\{ x_1 \cdots x_n \right\}$.
      m process variables, *n* sampling.
      Control limit *threshold* = 90%

Process:
S1: Normalize the raw data set X_{train}
S2: The time domain signal is transformed into frequency domain signal by the formula (7).
S3: The spectral density $\rho_z(\lambda)$ is derived by the formula (8).
S4: Compute the Information entropy of x_i by $\rho_z(\lambda)$ and the formula (9).
S5: The predictability values Φ_i ($i = 1 \cdots m$) and eigenvectors w $= \{w_i\}_1^m \in R^m$ are obtained by the EM-like algorithm and formula (10)-(12).
S6: the number of forecast principal components is computed by the formula (14). The principal forecast components score matrix can be further obtained.
Output: The forecast principal components number K, the forecast components transfer matrix W. The principal forecast components score matrix $\tilde{X} = (W_k^T X^T)^T \in R^{n*k}$

---

domain signals [13], so we take the univariate stationary time series z_t as an example. Assuming that the corresponding mean and variance are u_z and σ_z^2, respectively. The auto-covariance function (ACVF) can be obtained as follows.

$$r_z(l) = E(z_t - u_z)(z_{t-l} - u_z) \quad (6)$$

Since the stationary time series is independent of the initial point, $r_z(l) = r_z(-l)$. The frequency spectrum of the univariate time series can be obtained by the following formula:

$$P_z(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} r_z(j) e^{2\pi ij\lambda} \quad \lambda \in [-\pi, \pi] \quad (7)$$

$P_z(\lambda)$ represents the nonnegative real function, i is the imaginary unit. Particularly, when $j = 0$, $r_z(0) = \sigma_z^2$. According to the inverse Fourier transform, $r_z(l) = \int_{-\pi}^{\pi} P_z(\lambda) e^{-2\pi i l \lambda} d\lambda$, so $r_z(0) = \int_{-\pi}^{\pi} P_z(\lambda) e^0 d\lambda = \sigma_z^2$. The standard spectral density function can be derived as follows:

$$\rho_z(\lambda) = \frac{P_z(\lambda)}{\sigma_z^2} \tag{8}$$

$\rho_z(\lambda)$ is a special spectral density function, which is used to analogize the probability density function of random variables in a unit circle, so $\int_{-\pi}^{\pi} \rho_z(\lambda) d\lambda = 1$. Shannon entropy can be used to estimate the uncertainty level of $\rho_z(\lambda)$, so the information entropy formula can be organized as follows:

$$S_{P,a}(z_t) := \int_{-\pi}^{\pi} \rho_z(\lambda) log_a \rho_z(\lambda) d\lambda \tag{9}$$

Remark 1: Assuming that ε_t is white noise with zero mean and finite variance, when $\neq 0$, $r_\varepsilon(l) = 0$ and $P_\varepsilon(\lambda) = (\sigma_\varepsilon^2)/2\pi$ can be obtained by the formula (6) - (8). Generally, the flat spectrum represents the sequence is most unpredictable [13], white noise is a special flat spectrum. The time series has the larger information entropy. so the information entropy of ε_t is maximum.

The information entropy of z_t is not greater than ε_t, so the following formula can be obtained:

$$
\begin{aligned}
S_{P,a}(z_t) &\leq S_{P,a}(\varepsilon_t) = -\int_{-\pi}^{\pi} \rho_\varepsilon(\lambda) log_a \rho_\varepsilon(\lambda) d\lambda \\
&= -\int_{-\pi}^{\pi} \frac{1}{2\pi} log_a \frac{1}{2\pi} d\lambda = log_a 2\pi
\end{aligned}
\tag{10}
$$

According to formula (10), the important index for solving forecastable transfer matrix is derived-$\Phi(\cdot)$.

$$
\begin{aligned}
\Phi(z_t) &= 1 - \frac{S_{P,a}(z_t)}{log_a 2\pi} = 1 - S_{P,2\pi}(z_t) \\
\Phi &: z_t \longmapsto [0, \infty]
\end{aligned}
\tag{11}
$$

$\Phi(\cdot)$ is an important index for evaluating the predictability of z_t. If $X_t \in R^m$ is a multivariate time series, the univariate time series z_t can be obtained by calculating $z_t = w^T X_t$. But w may not be optimal, that is, it cannot be used to maximize the value of $\Phi(\cdot)$. Therefore, finding W becomes an optimization problem shown as follows:

$$
\begin{aligned}
max_w \, \Phi(z_t) &= \Phi\left(w^T X_t\right) \\
&= max_w \left(1 + \frac{\int_{-\pi}^{\pi} \frac{w^T P_X(\lambda) w}{w^T \Sigma_X w} log_a \frac{w^T P_X(\lambda) w}{w^T \Sigma_X w} d\lambda}{log_a 2\pi}\right)
\end{aligned}
$$
$$s.t. \; w^T \Sigma_X w = 1 \tag{12}$$

where $P_X(\lambda) = 1/2 \sum_{j=-\infty}^{\infty} \Gamma_X(j) e^{2\pi ij\lambda}$, $\Gamma_X(j)$ is ACVF of X_t, $\Gamma_X(j) = \Gamma_X(-j)^T$. Because the number of data set is finite. Therefore, the optimization of formula (12) requires spectral density estimation first. In this paper, [13], [23]

(WOSA) is used to estimate the spectral density. So the formula can be further simplified as follows:

$$w^* = argmax_w (1/T) \sum_{j=1}^{T-1} w^T P_X(\omega_j) w \, log_a w^T P_X(\omega_j) w \tag{13}$$

Let $L(w; \omega_j) = log_a w^T P_X(\omega_j) w$, w^* represents that the forecastable vector is optimal. An optimal transfer matrix W and the corresponding predictability value Φ can be obtained by cyclically calling the EM-like algorithm. Goerg gives a detailed derivation process of the EM- like algorithm [13].

Motivated by the cumulative variance contribution rate of PCA [24]. The cumulative predictable contribution rate is used to determine the number of forecast principal components.

$$\tau(\Phi) = \frac{\sum_{i=1}^{k} \Phi_i}{\sum_{i=1}^{m} \Phi_i} \geq threshold(k \leq m) \tag{14}$$

m represents the number of variables, and k represents the number of forecast principal components. The threshold is usually set by experience. The predictability values of Φ_i are arranged from large to small, therefore, $W_k \in R^{m*k}$ and $\tilde{X} = (W_k^T X^T)^T \in R^{n*k}$ can be further derived. $W_k \in R^{m*k}$ represents the optimal forecast principal components transfer matrix, $\tilde{X} = (W_k^T X^T)^T \in R^{n*k}$ represents the principal forecast components score matrix.

B. PARAMETER OPTIMIZATION-QUADRATIC GS ALGORITHM

The classification ability of SVM is mainly affected by the kernel parameter g and penalty parameter C [25]. So far, there is still no generally accepted optimal algorithm. Different from the standard GS algorithm, this paper proposed the quadratic GS algorithm to deal with the parameters optimization. The quadratic GS firstly sets up a large initial parameter interval, and then performs a second adjustment according to the verification data.

Firstly, quadratic GS divides data set into training set and testing set. Secondly, the parameter interval are arranged to generate the corresponding "grid". Thirdly, the cross-validation method is used to select the parameter C, g. Finally, parameters intervals are re-adjusted according to OFC-SVM performance. In some degree, Quadratic GS algorithm can avoid over-fitting drawback by secondary adjustment strategy. However, the experiment found that more than one group of parameters C, g can simultaneously maximize the SVM classification accuracy. For this situation, this paper chooses the small value of C. Since the larger value of C, the easier occurrence of over-fitting problem [26]. Meanwhile, this paper considers the computational cost problem and chooses the most commonly used K-fold Cross Validation (K-CV) method for cross-validation. K-CV divides the data set into K groups equally, and establishes K models respectively (the verification set is a set from the K group data, and the remaining K-1 group are used as training set),

Then, the average classification accuracy of K models is used as the performance index of K-CV.

In order to strictly verify the performance of quadratic GS algorithm, the parameter interval is not changed once it is set by second adjustment. At the same time, PSO and GA are compared with the quadratic GS algorithm in this paper. Initial parameter of three optimization algorithm can refer to Supplementary Information A(SIA).

C. PERFORMANCE EVALUATION INDEX-PAR

Performance evaluation index (PEI) is an important tool for evaluating and selecting monitoring models. Although the diagnosis accuracy can roughly evaluate the model, it cannot meet the plant real-time requirements. For example, the actual WWTP always use false alarm and missed alarm to evaluate the performance of the monitoring system [17]. Because the work intensity is directly affected by the false alarm and missed alarm. When the FAR of WWTP is high, that is, an alarm flood occurs under normal work conditions. The monitoring systems will continuously false alarm commands, which will waste a lot of manpower and resources. The missed alarm is a more serious condition than the false alarm. If the monitoring system missed the fault, the engineer cannot take effective measures to troubleshoot in time. So we proposed a comprehensive PEI of Pre-alarm rate (PAR) by mixing false alarm and missed alarm. The formula of PAR, MAR and Accuracy are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (15)$$

$$M_{AR} = \text{Fr(Normal|Fault)} = \frac{FN}{FN + TP} \quad (16)$$

$$F_{AR} = \text{Fr(Fault|Normal)} = \frac{FP}{FP + TN} \quad (17)$$

Note that the "Normal" represents the normal work conditions. Fr(|) represents the conditional frequency. TP is true positive, this article represents the sample correctly classified as fault. TN is true negative, and it represents the sample correctly classified as normal. FP (False Positive) indicates that the sample incorrectly classified as normal. M_{AR} represents the formula of MAR, F_{AR} represents the formula of FAR. The formula of PAR as follows:

$$P_{AR} = \gamma M_{AR} + (1 - \gamma)F_{AR} \quad (18)$$

γ represents the weight parameter ($0 \leq \gamma \leq 1$), the plant can adjust γ according to the demand signal.

Definition 1 (PAR in Some Different Situations):

1. When $\gamma = 0$, $P_{AR} = F_{AR}$. PAR degenerates into FAR, so let Type I PAR = FAR.

2. When $\gamma = 1$, $P_{AR} = M_{AR}$. PAR degenerates into MAR, so let Type II PAR = MAR.

3. When $\gamma > 0.5$, the missed alarm is more important for plant.

4. When $\gamma < 0.5$, the plant is more value its false alarm.

In the experiment, we find a special scenario that the several methods have the same PAR value. To solve this
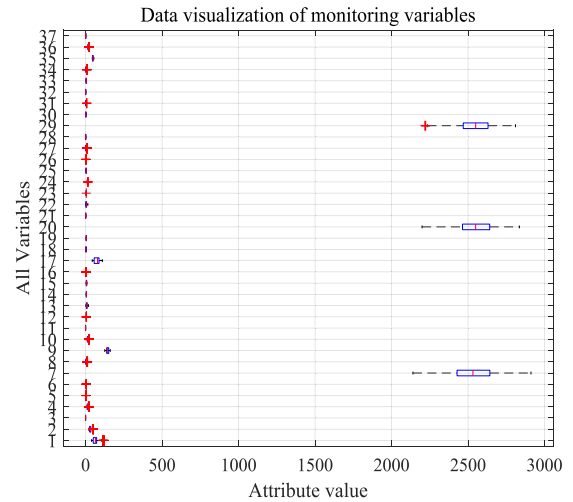

FIGURE 2. The data box-plot.

problem, Type-I PAR and Type-II PAR are used to derive the corresponding ROC curve, and then using this index to assist making decision. ROC is originated from the radar signal detection, which is a powerful tool for evaluating the learner generalization performance [27]. The *x*-axis and *y*-axis of ROC curve are composed of false positive rate (FPR) and true positive rate (TPR), respectively. In statistics, the FPR is equal to the Type-I PAR. The TPR and the Type-II PAR can be linked, TPR formula is $T_{PR} = TP/(FN + TP)$, which indicates the proportion of predicted correctly under the fault condition. Therefore, the $M_{AR} = 1 - T_{PR}$ can be further derived. As given above, the ROC curve can be derived by Type-I PAR and Type-II PAR. If the ROC curve of a method completely covers another, it can directly assert that the former is superior. But when the two curves intersect, we need to calculate the area. So the area of ROC curve is as follows:

$$\text{AUC} = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1}^{PAR} - x_i^{PAR})(y_i^{PAR} + y_{i+1}^{PAR}) \quad (19)$$

(x_i^{PAR}, y_i^{PAR}) represents the coordinate points of ROC curve. Generally, the larger area under ROC curve indicates the stronger generalization performance

Remark 2: In this paper, ROC curve is derived by Type-I PAR and Type-II PAR. A special point needed to be stress, ROC curve in this paper is not a global evaluate index, but an auxiliary PAR to evaluate the performance of the monitoring model.

IV. CASE STUDY
A. CASE 1: MONTE CARLO SIMULATION
1) BACKGROUNDS AND SCENARIO DEFINITION
Monte Carlo model is a numerical simulation widely used in the field of processes monitoring. More information about this model can referred to [28], [29]. In this paper, the sensor

TABLE 1. Diagnosis results of X4.

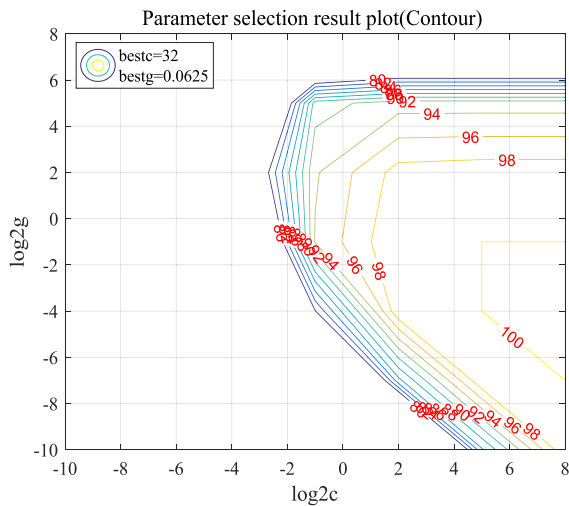| Evaluation Index	X4					
	SVM	PCA-SVM	ForeCA-SVM	OFC-SVM	ForeCA-GA-SVM	ForeCA-PSO-SVM
Type-I PAR	8.000%	6.500%	5.500%	3.500%	4.000%	3.500%
Type-II PAR	18.000%	19.000%	17.000%	15.000%	28.000%	16.000%
PAR	12.000%	11.500%	10.100%	8.100%	13.600%	8.500%
Accuracy	88.667% (266/300)	89.333% (268/300)	90.667% (272/300)	92.667% (278/300)	88.000% (264/300)	92.333% (277/300)

FIGURE 3. Quadratic GS parameter optimal results plot.
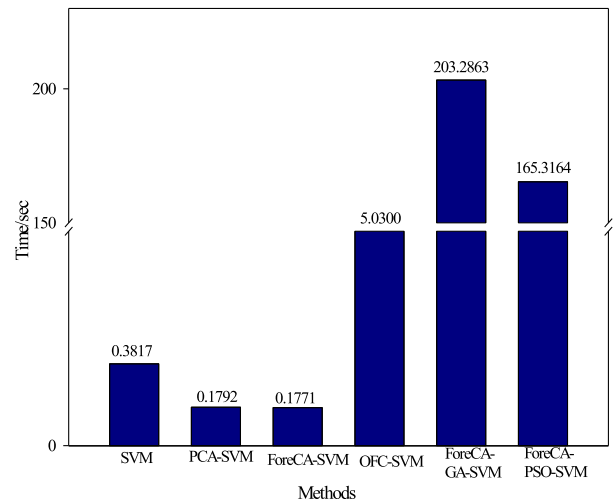

FIGURE 4. Six methods time consuming plot.

fault is simulation by the following Monte Carlo model:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_{18} \end{bmatrix} = A_{18 \times 6} * \begin{bmatrix} t_1 \\ \vdots \\ t_6 \end{bmatrix} + \Psi_{noise} \qquad (20)$$

where x_i $(i = 1 \cdots 18)$ represents different types of sensors and t_i $(i = 1 \cdots 6)$ represents the input variable which subject to normal distribution, $(t_1 \sim N(16, 0.3), t_2 \sim N(18, 0.6), t_3 \sim N(13, 0.5), t_4 \sim N(15, 0.1), t_5 \sim N(16, 0.4) t_6 \sim N(8, 0.6)$. The noise of Ψ_{noise} in this process is assumed to be normally distributed with the mean of zero and the standard deviation of 0.2. To obtain the suitable data set, the simulation period is set to 1000 and the sampling interval is 1 sec. A sensor fault is happened after the 600th sample and the fault amplitude is 0.35. The matrix of $A_{18 \times 6}$ can refer to Supplementary Information B.

2) EXPERIMENTAL VERIFICATION
To demonstrate the effectiveness of the proposed method, the original random noise is expanded four times randomly and the fault sensor X_4 is selected. OFC-SVM and another five methods are trained by the raw data of 1-300 samples and 601-900 samples, which is from the normal working condition and the fault working condition respectively. The remaining data are used to train the model. Both of PCA

and ForeCA have the same control limit 0.9.The core initial parameters of the six methods can refer to Table S1-S4 (Table S1-S4 in the Supplementary Information A)

Table 1 shows the six methods fault diagnosis results. The PAR weight parameter is $\gamma = 0.6$. We can see that the OFC-SVM method achieves the lowest PAR value, which reduces the PAR value of SVM method by 32.5%. At the same time, the OFC-SVM diagnosis accuracy is 92.667%, which is much higher than that of other five methods. It is important to notice that ForeCA-GA-SVM PAR value is larger than ForeCA-SVM and SVM. Due to the performance of GA is heavily sensitive the initial parameters setting, thus leading to obvious uncertainty and then resulting in the worse classification relatively. As tabulated in Table 1, ForeCA and PCA can improve the SVM performance by effectively extracting features. Moreover, the program running time is greatly reduced after feature extraction. Fig. S1 shows the six methods program running time (Supplementary Information D) after the feature extraction, SVM running time is 0.34 sec. The feature is extracted effectively by ForeCA algorithm, the program running time is shortened by 35.3% with it being 0.22 sec. Since the optimization algorithm is very time consuming, the ForeCA-GA-SVM and ForeCA-PSO-SVM programs run with the consumed time being 278.94 sec and 222.98 sec, respectively. But the proposed OFC-SVM running time only costs 5.67 sec.
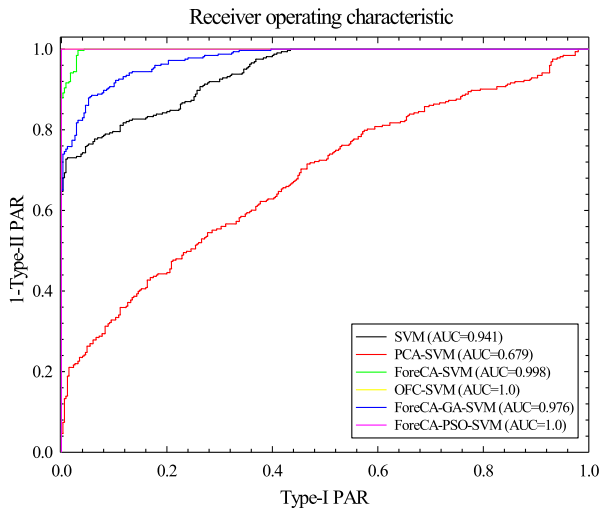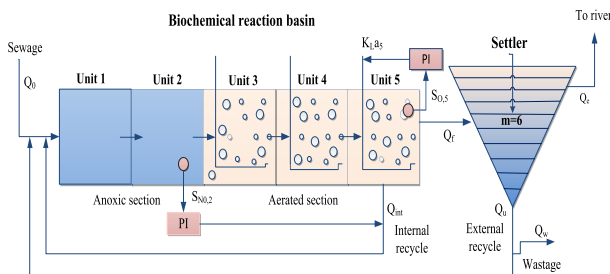
FIGURE 5. Six methods ROC curve plot.

FIGURE 6. Schematic diagram of a full scale wastewater plant under BSM1.

B. CASE 2: A FULL-SCALE WWTPs

1) BACKGROUNDS

A widely recognized full-scale wastewater treatment platform (benchmark simulation model 1 BSM1) is selected in this paper, as depicted in Fig. 6. It is proposed within the framework of COST Actions 682/ 624 by the IWA (International Water Association), aiming to remove C/N. The BSM1 wastewater treatment plant full considers the complexity of biochemical reactions and physical phenomena, so it is design five biochemical reactions and a complete settler tanks. BSM1 is set three weather event (dry weather, rain weather, storm weather) and some default parameters, every weather event simulation last 14 days. Also, the data is collected every 15 min. The detailed introduction can refer to the website (http://www.benchmarkWWTP.org).

2) SCENARIO DEFINITION AND MONITORING VARIABLES SELECTION

In this case study, the raw data set is collected on dry weather days in BSM1, consisting of 37 process variables. (Variables are depicted in Table S5 of the Supplementary Information C). The empirically selected monitoring variables cover the entire process of the WWTP.

To verify the effectiveness of the method, the faulty sensor is selected according to the following two principles: (1) The

selected fault sensor must envelop the whole WWTP process, i.e., including influent wastewater, five biochemical reaction tanks, and secondary settling tank and effluent wastewater. (2) The variables selected are based on the importance of the variables to WWTP and the frequency of failures.

3) EXPERIMENTAL VERIFICATION

In this section, the faulty senor $S_O - 1$ is taken into account. $S_O - 1$ sensor is usually used to measure oxygen content in No. 1 anoxic tank, which is an important indicate if the denitrification reaction can proceed normally in anoxic tank. By analyzing the data box-plot in Fig. 2, we can easily find that the data need to be normalized. Table 2 shows diagnosis results of six methods after data normalization. The PAR weight parameter is $\gamma = 0.6$, which is based on the fact that missed alarm will bring about more serious losses to WWTP. The other important parameters are consistent with Case 1. The result of quadratic GS parameter optimization is profiled as Fig. 3. The optimal parameters are $C = 32$, $g = 0.0625$.

From the diagnostic results of Table 2, we can see that the PAR value with respect to ForeCA-SVM is significantly lower than SVM, which is reduced by 67.294% compared with the PAR value of SVM. This demonstrates that the performance of SVM can be indeed improved by using ForeCA algorithm to feature extraction. But the PCA-SVM PAR value is significantly higher than SVM. This is mainly because not all the data collected by WWTPs follow normal distribution. In addition, the PCA-SVM method is better than SVM from the result of Case 1, this is mainly due to the data obey normal distribution.

To further assess the performance, the running time for each algorithm is also considered. Fig. 4 shows program running time of six methods. ForeCA-SVM indicates that the SVM running time after feature extraction is 0.177 sec, whereas the SVM running time without feature extraction is 0.3817 sec. The feature extraction can indeed reduce the consuming time. OFC-SVM diagnosis accuracy is 100%, which achieve the best performance than the remaining five methods. At the same time, the time consuming is much lower than ForeCA-GS-SVM and ForeCA-PSO-SVM.

In addition, the diagnosis results with respect to six monitoring methods are given in Fig. 7. To make the diagnosis more visual, we have added three units to the predict label. It can be seen directly that the OFC-SVM performs best. And the six methods ROC curve in Fig. 5 further confirms the proposed method is superior.

V. RESULTS AND DISCUSSION

Through comparing the PAR values of case 1 with case 2, we can make an initial conclusion that the OFC-SVM method is optimal. However, the following two special scenarios have happened in the experiment:

a). As shown in section III-C, PAR formula is constructed by the FAR and MAR. So there are two special conditions can make several methods have the same PAR values. Firstly, when $M_{AR}^i = M_{AR}^j$, $F_{AR}^i = F_{AR}^j$ i $\neq$ j and i, j $\in \{1 \cdots 6\}$,

TABLE 2. Diagnosis results of $S_O - 1$.

Evaluation Index	S_O-1					
	SVM	PCA-SVM	ForeCA-SVM	OFC-SVM	ForeCA-GA-SVM	ForeCA-PSO-SVM
Type-I PAR	6.000%	8.286%	1.429%	0.000%	2.857%	0.000%
Type-II PAR	23.529%	68.731%	8.050%	0.000%	19.195%	0.000%
PAR	16.517%	44.553%	5.401%	0.000%	12.660%	0.000%
Accuracy	85.587% (576/673)	62.704% (422/673)	95.394% (642/673)	100% (673/673)	89.302% (601/673)	100% (673/673)

FIGURE 7. SO-1 diagnosis results visual plot.
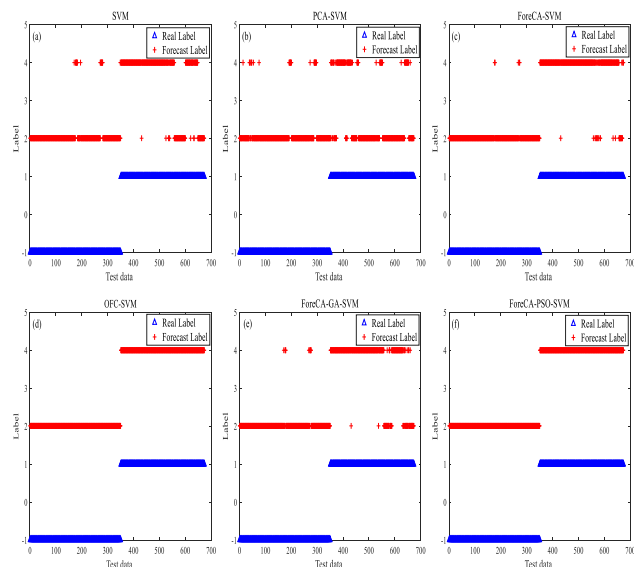

FIGURE 8. ROC curve plot of SALK-5.

the PAR have the same values. Secondly, when $M_{AR}^i > M_{AR}^j$, $F_{AR}^i < F_{AR}^j$, or $M_{AR}^i < M_{AR}^j$, $F_{AR}^i > F_{AR}^j$, at the same time, the weight parameter γ ($0 < \gamma < 1$) make several different methods occurred the situation of $P_{AR}^i = P_{AR}^j$.

b). The proposed OFC-SVM method and the five basic methods are presented to monitor the industrial process, as discussed above, the industrial system mix physical, chemical and biological reactions. Also, the significant disturbance is further added the difficult of processes monitoring. So the same method will show inconsistent performance when it under the different working conditions. At the same time, GA and PSO algorithms have uncertainty in finding parameters [16]. As showed in case 1, the performance of ForeCA-GA-SVM method is weaker than ForeCA–SVM. But in some special conditions, the parameters are searched by GA algorithm will be better than GS and PSO algorithm (The special conditions can refer the supplement information D-Table S6 (the fault sensor of X6)). So the performance of OFC-SVM needed to be further strictly confirmed.

A. SCENARIO 1: SEVERAL METHODS HAVE THE SAME PAR VALUES

To solve the problem that several methods have the same PAR values, a novel ROC curve is developed to assist PAR decision-making. Taking the S_{ALK}-5 fault as an example,
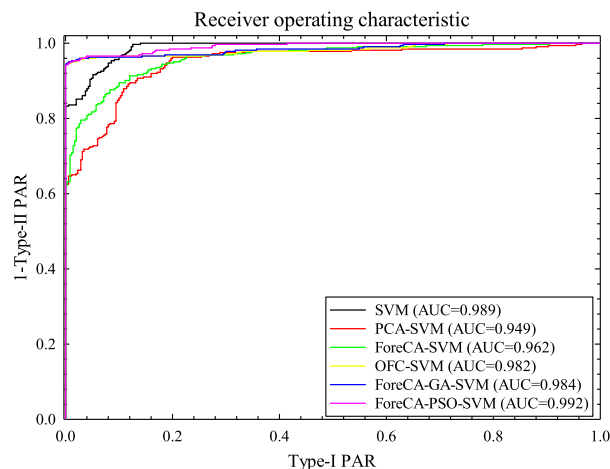
OFC-SVM, ForeCA-GA-SVM and ForeCA-PSO-SVM PAR values are 5%, 4%, and 4%, respectively. ForeCA-GA-SVM and ForeCA-PSO-SVM have the same PAR values, the performance of the two methods cannot be judged by the PAR values. So the ROC curve derived by type I PAR and Type II as shown in Fig. 8, the AUC of ForeCA-PSO- SVM (AUC = 0.992) is larger than that of ForeCA-GA-SVM (AUC = 0.984). So the ForeCA-PSO-SVM method is much better than ForeCA-GA -SVM, which is consistent with the diagnosis accuracy (96.73% > 96.29%).

B. SCENARIO 2: OFC-SVM METHOD IS NOT ALWAYS OPTIMAL

As shown in Scenario 1, the PAR value of OFC-SVM is large than ForeCA-GA-SVM and ForeCA-PSO-SVM. Since the quadratic GS sets the parameter interval by trading off the running time and diagnosis accuracy, the optimal parameter may be outside the set interval. In order to rigorously verify the method performance, the 10 sensor fault diagnosis results of the Monte Carlo model show in Table S6. Table S7 presents the 20 sensor faults diagnosis results of WWTP (Table S6-S7 in the Supplementary Information D). The PAR weight parameter is still set 0.6, and all parameters are consistent with the previous setting. Fig. 9 and Fig. 10 are the averaging PAR and average diagnostic accuracy for Tables S6 and S7. It can be seen from the Fig 8 and Fig 9 that the OFC-SVM average PAR is the lowest on all platforms. Taking Fig. 10 as an example, the proposed OFC-SVM method average PAR
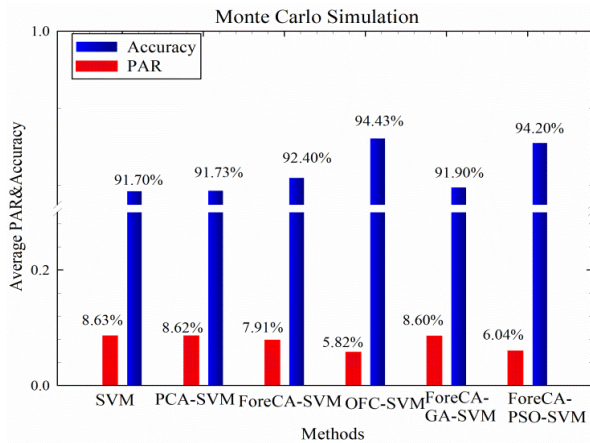
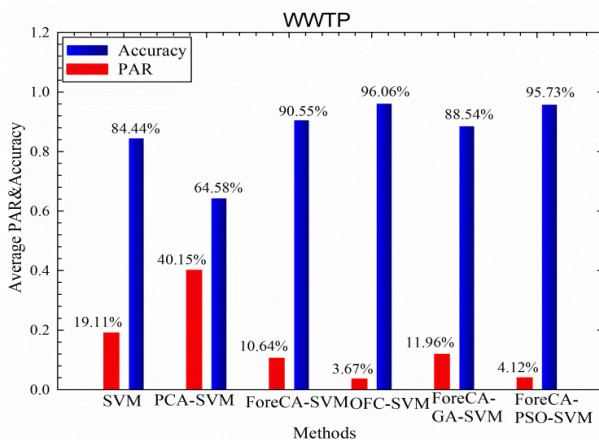**FIGURE 9.** Fault diagnosis results of Monte carlo model.

FIGURE 10. Fault diagnosis results of WWTP.

value is 3.67%, which is lower than the SVM by 80.8%. It is also lower than ForeCA-SVM (65.5%). Moreover, the OFC-SVM diagnosis accuracy is the highest among the six methods, which is 6.1% higher than ForeCA-SVM method and 13.76% higher than SVM method. Therefore, the average PAR and diagnosis accuracy simultaneously verify that the OFC-SVM method is optimal.

VI. CONCLUSION

In this work, a novel fault diagnosis method namely OFC-SVM together with a following comprehensive performance evaluation index PAR is proposed. The contribution herein mainly includes the following three aspects: Firstly, ForeCA algorithm is improved to extract the feature, the ForeCA algorithm can compensate for the weakness of PCA under Non-Gaussian data. Also, use of ForeCA algorithm for feature extraction can effectively reduce the time consumption of SVM model. Secondly, to further improve the model performance, a simple quadratic GS algorithm is used to derive the optimal parameters. The quadratic GS is able to achieve better results through seeking parameters than PSO and GA with less time consumption. Thirdly, a novel

comprehensive index PAR is proposed. And further derivation a new ROC curve to assist PAR decision. The proposed OFC-SVM method is compared with SVM, PCA-SVM, ForeCA-SVM, ForeCA-GA-SVM, ForeCA-PSO-SVM, with being validated by Monte Carlo model and a full-scale WWTP. By compare the results of all methods, OFC-SVM is more superior in terms of PAR values. In the meanwhile, the diagnostic accuracy further validates the conclusion.

REFERENCES

[1] Y. Liu, B. Liu, X. Zhao, and M. Xie, "A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6478–6486, Aug. 2018.

[2] J. Zhou, Y. Yang, S. X. Ding, Y. Zi, and M. Wei, "A fault detection and health monitoring scheme for ship propulsion systems using SVM technique," *IEEE Access*, vol. 6, pp. 16207–16215, 2018.

[3] X. Zhu, J. Xiong, and Q. Liang, "Fault diagnosis of rotation machinery based on support vector machine optimized by quantum genetic algorithm," *IEEE Access*, vol. 6, pp. 33583–33588, 2018.

[4] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 10, pp. 3543–3562, 2013.

[5] V. Vapnik, *The Nature of Statistical Learning Theory*. 1995.

[6] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, 2007.

[7] A. Widodo, B.-S. Yang, and T. Han, "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors," *Expert Syst. Appl.*, vol. 32, no. 2, pp. 299–312, 2007.

[8] I. Yélamos, G. Escudero, M. Graells, and L. Puigjaner, "Performance assessment of a novel fault diagnosis system based on support vector machines," *Comput. Chem. Eng.*, vol. 33, pp. 244–255, Jan. 2009.

[9] J. Ni, C. Zhang, and S. X. Yang, "An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBs," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1960–1971, Jul. 2011.

[10] J. J. F. Chen and D. P. Visco, Jr., "Identifying novel factor XIIa inhibitors with PCA-GA-SVM developed vHTS models," *Eur. J. Med. Chem.*, vol. 140, pp. 31–41, Nov. 2017.

[11] Y.-W. Chang, Y.-C. Wang, T. Liu, and Z.-J. Wang, "Fault diagnosis of a mine hoist using PCA and SVM techniques," *J. China Univ. Mining Technol.*, vol. 18, pp. 327–331, Sep. 2008.

[12] L. Cao, K. Chua, W. Chong, H. Lee, and Q. Gu, "A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine," *Neurocomputing*, vol. 55, nos. 1–2, pp. 321–336, 2003.

[13] G. M. Goerg, "Forecastable component analysis (ForeCA)," 2012, *arXiv:1205.4591*. [Online]. Available: https://arxiv.org/abs/1205.4591

[14] A. Dionisio, M. Rui, and D. A. Mendes, "Entropy and uncertainty analysis in financial markets," Tech. Rep., 2007.

[15] S. Yin, Y. Jiang, Y. Tian, and O. Kaynak, "A data-driven fuzzy information granulation approach for freight volume forecasting," *IEEE Trans. Ind. Electron.*, vol. 64, no. 2, pp. 1447–1456, Feb. 2017.

[16] V. Roberge, M. Tarbouchi, and G. Labonte, "Comparison of parallel genetic algorithm and particle swarm optimization for real-time UAV path planning," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 132–141, Feb. 2013.

[17] Y. Liu, Y. Pan, Q. Wang, and D. Huang, "Statistical process monitoring with integration of data projection and one-class classification," *Chemometrics Intell. Lab. Syst.*, vol. 149, pp. 1–11, Dec. 2015.

[18] H. Xiao, Y. Liu, and D. Huang, "Semiadaptive fault diagnosis via variational Bayesian mixture factor analysis with application to wastewater treatment," *J. Control Sci. Eng.*, vol. 2016, Mar. 2016, Art. no. 13.

[19] A. Tulsyan, F. Alrowaie, and B. Gopaluni, "Design and assessment of delay timer alarm systems for nonlinear chemical processes," *AIChE J.*, vol. 64, no. 1, pp. 77–90, 2018.

[20] J. Xu, J. Wang, I. Izadi, and T. Chen, "Performance assessment and design for univariate alarm systems based on FAR, MAR, and AAD," *IEEE Trans. Autom. Sci. Eng.*, vol. 9, no. 2, pp. 296–307, Apr. 2012.

[21] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[22] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using continuous wavelet transform," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2300–2312, 2011.

[23] A. Nuttall and G. Carter, "A generalized framework for power spectral estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 3, pp. 334–335, Jun. 1980.

[24] H. Cheng, J. Wu, Y. Liu, and D. Huang, "A novel fault identification and root-causality analysis of incipient faults with applications to wastewater treatment processes," *Chemometrics Intell. Lab. Syst.*, vol. 188, pp. 24–36, May 2019.

[25] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, 2003.

[26] R. Izmailov, V. Vapnik, and A. Vashist, "Multidimensional splines with infinite number of knots as SVM kernels," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2013, pp. 1–7.

[27] Y. Cai, M.-Y. Chow, W. Lu, and L. Li, "Evaluation of distribution fault diagnosis algorithms using ROC curves," in *Proc. IEEE PES Gen. Meeting*, Jul. 2010, pp. 1–6.

[28] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring," *Automatica*, vol. 45, no. 7, pp. 1593–1600, Jul. 2009.

[29] J. Xuan, Z. Xu, and Y. Sun, "Incipient sensor fault diagnosis based on average residual-difference reconstruction contribution plot," *Ind. Eng. Chem. Res.*, vol. 53, no. 18, pp. 7706–7713, 2012.

HONGCHAO CHENG was born in Xinyang, Henan, China. He received the B.E. degree in statistics from Kashi University, in 2016. He is currently pursuing the Ph.D. degree in control science and engineering with the Department of Automation, South China University of Technology, Guangzhou, China.

His research interests include statistical theory research, fault diagnosis, soft sensor, process monitoring, and wastewater treatment.

YIQI LIU was born in Haikou, China, in 1983. He received the B.S. and M.S. degrees in control engineering from the Chemical University of Technology, Beijing, in 2009, and the Ph.D. degree in control engineering from the South China University of Technology, Guangzhou, China, in 2013. From 2013 to 2016, he was a Lecturer with the South China University of Technology, where he has been an Associate Professor with the Department of Automation, since 2016. He is the author of more than 50 articles. His research interests include soft sensors, fault diagnosis, and wastewater treatment. He holds three patents. He was a recipient of the Hong Kong Scholar Award, in 2016, the Chinese Scholarship Council Award, in 2011, and the Deutscher Akademischer Austausch Dienst Award, in 2015. He is an Associate Editor of IEEE Access.

DAOPING HUANG received the B.Eng. degree in chemical automation and instruments and the M.Eng. and Ph.D. degrees in automatic control theory and applications from the South China University of Technology (SCUT), Guangzhou, China, in 1982, 1986, and 1998, respectively.

In 1982, he started his Teaching and Research Career with the Department of Automation, SCUT. From 1995 to 1996, he was with the University of Gent, Belgium, as a Government Sponsored Visiting Scholar. He is currently a Full Professor with the School of Automation Science and Engineering. He has authored three academic books and more than 150 conference and journal articles. His research interests include intelligent detection and control and soft-sensing technology, as well as fault diagnosis and accident prediction for industrial processes. He serves as the Vice Director for the Education Committee and a member of the Process Control and Application Committees at the Chinese Association of Automation. He has directed more than ten research projects. He was granted a Third Prize from the National Education Committee, in 1992, and a Second Prize from the Guangdong Provincial Government, in 2005, for his contributions to science and technology.

BIN LIU received the B.S. degree in automation from Zhejiang University, China, and the Ph.D. degree in industrial engineering from the City University of Hong Kong, Hong Kong. He was a Postdoctoral Fellow with the University of Waterloo, Canada. He is currently a Lecturer with the Department of Management Science, University of Strathclyde, Glasgow, U.K. His research interests include risk analysis, reliability and maintenance modeling, decision making under uncertainty, and data analysis.

• • •