Received July 2, 2020, accepted July 4, 2020, date of publication July 15, 2020, date of current version July 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009562

Multiple Elastic Networks With Time Delays for Early Fault Detection and Prognostics

DONGLIANG GUO[1], WEN YANG[2], FENGBO TAO[1], BING SONG[2], HUI LIU[3], LEI SUN[1], AND JIALE WANG[2]

[1]State Grid Jiangsu Electric Power Company Ltd. Research Institute, Nanjing 211103, China
[2]Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China
[3]Shanghai Power & Energy Storage Battery System Engineering Technology Company Ltd., Shanghai 200241, China

Corresponding authors: Bing Song (songbing@ecust.edu.cn) and Dongliang Guo (tiankongyiran90@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61703161, in part by the Fundamental Research Funds for the Central Universities under Grant 222201714031, and in part by the National Natural Science Foundation of Shanghai under Grant 19ZR1473200.

ABSTRACT Aiming at the problem of fault prognostics for the energy storage power station, this paper proposes a novel data-driven method named multiple elastic networks with time delays (MEN-TD). The proposed method can learn the status of the energy storage power station in advance and provide early detection of the fault. First, through the correlation analysis and the mechanism knowledge, the energy storage power station key parameter and corresponding key factors affecting the parameter are determined. Secondly, in order to predict the trend of the key parameter over a period of time and improve the prediction accuracy, the MEN-TD model is constructed. Then, based on the predicted values of the key parameter, compared with the control limit in the healthy status, the fault can be pre-warned in advance. Finally, through testing on the practical energy storage power station in Zhenjiang of China, the effectiveness and superiority of the proposed MEN-TD method are demonstrated.

INDEX TERMS Fault detection, data analysis, fault prognostics, fault diagnosis, process monitoring.

I. INTRODUCTION

With the further requirement of the system safety of energy storage power stations, not only do we hope to be able to detect and isolate faults when the status deteriorates, but also require that we can provide early fault detection and prognostics before the system deteriorates, to ensure that there is enough time to take measures to prevent deterioration and avoid unnecessary losses [1]–[4]. Early fault detection and prognostics is an area where fault detection and event prediction technology are combined. The key is to accurately predict the operating status, estimate the future change trend [5], [6], and then detect whether the fault occurs or not.

Trend prediction methods can be divided into the mechanism model based method and data-driven method. The mechanism model based method refers to the establishment of a mathematical model in the form of differential equations or algebraic equations based on material balance and energy conservation. This type of method is based on the solid process mechanism analysis, so its mathematical model is accurate and the prediction results are accurate. Many scholars have done a lot of research in this field, formed the mature framework, and obtained successful application cases [7]. However, because the operating system of energy storage power stations often has strong nonlinearity, strong coupling, and instability, if the system mechanism cannot be fully modeled, it is easy to cause model mismatch. With the rapid development of the information perception technology and the application of advanced sensing equipment, a large amount of historical data has been accumulated during the operation of energy storage power stations, providing a solid foundation for data-driven trend prediction methods. In addition, mechanism model based method and data-driven method have been fused. Given that the effect of multiple causes, Cai *et al.* proposed the hybrid physics-model-based and data-driven method for remaining useful life estimation, and obtained satisfactory performance [8]. Data-driven prediction methods mainly use historical data and current data to mine valuable information such as periodicity, basic trends and association rules of data, and construct mathematical models to predict future trends. At present, data-driven prediction methods mainly include the statistical learning based method, the regression based methods, and the neural network based methods.

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

The statistical learning based method builds the statistical model that can accurately reflect the dynamic dependencies contained in the data to achieve the purpose of prediction. Most of these methods require data to meet the assumptions of stationary, time series correlation, and data variance cannot be too large. They mainly include auto regressive (AR) [9], moving average (MA) [10], auto regressive moving average (ARMA) [11], auto regressive integrated moving average (ARIMA) [12] and seasonal auto regressive moving average (SARMA) [13]. However, the above models are suitable for univariate, same-variance prediction scenarios. For multivariate and heteroscedasticity prediction problems, the auto regressive conditional heteroskedasticity (ARCH) [14] was proposed. Due to the randomness and instability of data, the applicability of the statistical learning based method is limited. In addition, the statistical learning based method is suitable for short and medium term prediction, and it has certain limitations for long term prediction.

The prediction process is essentially closely related to regression analysis. Typical linear regression methods include partial least squares (PLS), multivariable linear regression (MLR), principal component regression (PCR), independent component regression (ICR), ridge regression (RR) and so on [15]–[18]. According to different application scenarios, scholars have studied different improved algorithms, such as improved PLS (IPLS) [19] and kernel PLS (KPLS) [20]. In addition to linear regression methods, support vector machine (SVM) and Bayesian network are also successfully used in the prediction work. Gestel *et al.* proposed applying the Bayesian evidence framework to the least squares support vector machine to predict the model parameters and related volatility of the data [21]. This method mainly includes three steps, where the first step is to infer the parameters of the time series model using the least squares support vector machine, the second step is to infer the hyperparameters related to the variance of the regularized noise, and the third step is to evaluate the time series model evidence. Bayesian network prediction is mainly to learn the Bayesian network structure for a given dataset [22]. This part is the basis for building the entire prediction model, and its purpose is to find a network structure that is most suitable for the data. Chen *et al.* designed a dynamic Bayesian network based on the kernel method to solve the problem of incomplete input time series prediction [23]. On the basis of object-oriented Bayesian networks, the fault detection and diagnosis method for complex systems with repetitive structures was developed, where both historical data and expert knowledge were processed [24]. The Bayesian network fault diagnosis method was developed for three-phase inverters, where two output line-to-line voltages are measured and the faults are detected based on the Bayesian networks [25]. The bibliographic review of Bayesian network has been presented for probabilistic knowledge representation and inference [26].

For more complex and highly nonlinear processes, the neural network based model is often used for prediction. Neural network has non-linear, self-organizing, self-learning, self-adaptive ability characteristics, and is an ideal rule and pattern learner. It converts complex mapping relationships into network parameters such as weights and thresholds, and converts model parameter identification problems into network parameters optimization problems, and then uses intelligent optimization algorithms to train parameters for obtaining reliable prediction performance. The neural network based method has been successfully applied to trend prediction [27]–[31].

With the continuous improvement of computer hardware and the rise of artificial intelligence, deep neural network models with complex structures and strong expression capabilities have emerged. Some deep neural network models have been applied for prediction successfully. The deep belief network with restricted boltzmann machines has been used to deal with time series prediction problems [32]. Yuan *et al.* proposed the deep learning based feature representation for top-level product quality prediction [33]. The prediction performance between the deep belief network and the integrated denoising autoencoder are compared [34]. In many deep neural network models, the hidden layer of the recurrent neural network (RNN) not only receives the signal of the neuron of the previous layer, but also receives the signal of the node of the previous hidden layer, considering the dependence between the nodes, which greatly improves the ability to predict the trend of data. Long short term memory (LSTM) [35] is an optimized variant of RNN, which is first proposed by Hochreater and Schmidhuber. The LSTM model makes up for the problems of gradient disappearance and gradient explosion of RNN, making it possible to use long distance time series information effectively [36].

Aiming at the problem of early detection and prognostics of the fault for the energy storage power station, this paper proposes the multiple elastic networks with time delays (MEN-TD) method. First of all, considering that accurate prediction models only relying on quantitative calculation methods are not enough, quantitative calculation methods based on the mutual information and qualitative methods based on the mechanism knowledge are combined to accurately determine the key factors which affect the energy storage power station key parameter. Secondly, in order to obtain the trend of the energy storage power station key parameter for a period of time, multiple elastic networks with time delays are established for prediction. Thirdly, to obtain the control limit of the energy storage power station parameter, the kernel density estimation (KDE) is used to analyze historical normal data. Finally, on the basis of the predicted values and the estimated control limit, the energy storage power station operating status can be evaluated, and the fault can be pre-warned or detected in advance. The effectiveness of the proposed method is tested and applied on the practical energy storage power station in Zhenjiang of China.

The main contributions of our work in this manuscript include: 1) In order to improve the prediction accuracy, both the quantitative calculation and the qualitative analysis are used to determine the key factors affecting the key parameter to be predicted. 2) A novel cumulative information percentage is proposed to select the number of key factors affecting the key parameter. 3) A novel multiple elastic networks with time delays method is proposed to predict the key parameter over a period of time simultaneously for early fault detection and prognostics. 4) Compared with the state of the art-the EN method with fixed delay, the proposed MEN-ED method can achieve higher accuracy through setting different delays. 5) For the purpose of improving the system safety, the proposed method has been applied to the practical energy storage power station. 6) The advantage of the proposed MEN-ED method is tested and proved through comparing with the EN method with fixed delay, the MEN-ED method without key factors selection, and the widely used PLS method.

The remaining of this work is listed as follows. Section II introduces the original EN algorithm. Section III elaborates the developed multiple elastic networks with time delays based fault prognostics method. Section IV introduces the testing results and analysis of the proposed method under the energy storage power station in Zhenjiang of China. Section V concludes this article.

II. PRELIMINARIES

Elastic network (EN) is a regression method with biased estimation, which can be used for collinear data analysis [37]. It uses L1 and L2 regularization at the same time. Thence, the regression prediction model not only has some non-zero sparse parameters, but also retains some conventional attributes.

Suppose $\boldsymbol{y} \in \boldsymbol{R}^{N\times 1}$ is the historical normal dataset of the key parameter to be predicted, $\boldsymbol{X} = [\boldsymbol{x}_1\boldsymbol{x}_2\cdots\boldsymbol{x}_m] \in \boldsymbol{R}^{N\times m}$ is the historical normal operation dataset, where N is the number of sample and m is the number of variables in the operation data.

$$\boldsymbol{X} = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_m(1) \\ x_1(2) & x_2(2) & \cdots & x_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N) & x_2(N) & \cdots & x_m(N) \end{bmatrix} \in \boldsymbol{R}^{N\times m} \quad (1)$$

$$\boldsymbol{y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} \in \boldsymbol{R}^{N\times 1}, \quad (2)$$

where $x_i(j)$ is the $j-th$ sample of $\boldsymbol{x}_i$, and $y(j)$ is the $j-th$ sample of $\boldsymbol{y}$.

In the EN model, the regression coefficient $\boldsymbol{\beta}^0 = \left[\beta_1^0, \beta_2^0, \cdots, \beta_m^0\right]^T$ between $\boldsymbol{X}$ and $\boldsymbol{y}$ is calculated as follows:

$$\boldsymbol{\beta}^0 = \arg\min\left[\sum_{j=1}^{N}\left(y(j) - \sum_{i=1}^{m}\beta_i^0 x_i(j)\right)^2\right] \cdot \varphi\sum_{i=1}^{m}\left|\beta_i^0\right| + (1-\varphi)\sum_{i=1}^{m}\left(\beta_i^0\right)^2 \leq t \quad (3)$$

where φ is the hybrid parameter of the elastic network, which controls the ratio of L1 regularization and L2 regularization.

Once the regression coefficient $\boldsymbol{\beta}^0 = \left[\beta_1^0, \beta_2^0, \cdots, \beta_m^0\right]^T$ is obtained, the predicted value $\hat{y}(T)$ of the key parameter at time T can be calculated as follows:

$$\hat{y}(T) = x_1(T) * \beta_1^0 + x_2(T) * \beta_2^0 + \cdots + x_m(T) * \beta_m^0. \quad (4)$$

In order to realize the simultaneous prediction of the key parameter for a period of time M, the elastic network with time delay M needs to be established, where the independent data and the data to be predicted are constructed as follows

$$\dot{\boldsymbol{X}} = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_m(1) \\ x_1(2) & x_2(2) & \cdots & x_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N-M) & x_2(N-M) & \cdots & x_m(N-M) \end{bmatrix} \in \boldsymbol{R}^{(N-M)\times m} \quad (5)$$

$$\dot{\boldsymbol{y}} = \begin{bmatrix} y(M+1) \\ y(M+2) \\ \vdots \\ y(M+N) \end{bmatrix} \in \boldsymbol{R}^{(N-M)\times 1}. \quad (6)$$

Similar to the equation in (3), after the regression coefficient $\boldsymbol{\beta}^M = \left[\beta_1^M, \beta_2^M, \cdots, \beta_m^M\right]^T$ is obtained, the predicted value $\left[\hat{y}(T+1), \hat{y}(T+2), \cdots, \hat{y}(T+M)\right]$ for a period of time M at time T can be calculated as follows

$$\begin{aligned} \hat{y}(T+1) &= x_1(T-M+1) * \beta_1^M + \cdots \\ &\quad + x_m(T-M+1) * \beta_m^M \\ \hat{y}(T+2) &= x_1(T-M+2) * \beta_1^M + \cdots \\ &\quad + x_m(T-M+2) * \beta_m^M \\ &\vdots \\ \hat{y}(T+M) &= x_1(T) * \beta_1^M + \cdots + x_m(T) * \beta_m^M. \end{aligned} \quad (7)$$

III. MULTIPLE ELASTIC NETWORKS WITH TIME DELAYS BASED FAULT PROGNOSTICS METHOD

A. SELECTION OF KEY FACTORS

Considering that some variables in the energy storage power station operating data are not related to the key parameter to be predicted, if these variables are added to the prediction model, the accuracy would be decreased. By using variables that are correlated with the key parameter to be predicted, redundant information and noise can be effectively removed. Therefore, it is necessary to select key factors that are related with the key parameter to be predicted for establishing the accurate prediction model.

Let $\boldsymbol{y} \in \boldsymbol{R}^{N\times 1}$ be the historical normal dataset of the key parameter to be predicted, $\boldsymbol{X} = [\boldsymbol{x}_1\boldsymbol{x}_2\cdots\boldsymbol{x}_m] \in \boldsymbol{R}^{N\times m}$ be the historical normal operation dataset. For the purpose of eliminating interference in the historical normal dataset $\boldsymbol{y} \in \boldsymbol{R}^{N\times 1}$ and $\boldsymbol{X} = [\boldsymbol{x}_1\boldsymbol{x}_2\cdots\boldsymbol{x}_m] \in \boldsymbol{R}^{N\times m}$, three-sigma rule is used. For convenience, the historical dataset after eliminating interference is still recorded as $\boldsymbol{y} \in \boldsymbol{R}^{N\times 1}$ and $\boldsymbol{X} = [\boldsymbol{x}_1\boldsymbol{x}_2\cdots\boldsymbol{x}_m] \in \boldsymbol{R}^{N\times m}$.

First, the mutual information (MI) is used to calculate the correlation between variables $\boldsymbol{x}_i (i = 1, 2, \cdots, m)$ and $\boldsymbol{y}$. MI is an information measure algorithm based on the information entropy. Suppose the joint probability density of $\boldsymbol{x}_i$ and $\boldsymbol{y}$ is denoted as $p(\boldsymbol{x}_i, \boldsymbol{y})$, and $p(\boldsymbol{x}_i)$ and $p(\boldsymbol{y})$ are the corresponding edge probability density, respectively. Then, the mutual information of $\boldsymbol{x}_i$ and $\boldsymbol{y}$ is calculated as follows:

$$I(\boldsymbol{x}_i, \boldsymbol{y}) = H\langle \boldsymbol{y} \rangle - H\langle \boldsymbol{y} \mid \boldsymbol{x}_i \rangle = \sum_{\boldsymbol{x}_i} \sum_{\boldsymbol{y}} p(\boldsymbol{x}_i, \boldsymbol{y}) \log \frac{p(\boldsymbol{x}_i, \boldsymbol{y})}{p(\boldsymbol{x}_i) p(\boldsymbol{y})}. \tag{8}$$

Sort the mutual information between variable $\boldsymbol{x}_i (i = 1, 2, \cdots, m)$ and $\boldsymbol{y}$ in the descending order as

$$\boldsymbol{I}(\boldsymbol{y}) = \begin{bmatrix} I(\boldsymbol{x}_1, \boldsymbol{y}) \\ I(\boldsymbol{x}_2, \boldsymbol{y}) \\ \vdots \\ I(\boldsymbol{x}_m, \boldsymbol{y}) \end{bmatrix} \in \boldsymbol{R}^{m \times 1}, \tag{9}$$

where $I(\boldsymbol{x}_1, \boldsymbol{y}) > I(\boldsymbol{x}_2, \boldsymbol{y}) > \ldots > I(\boldsymbol{x}_m, \boldsymbol{y})$.

The number of variables related to $\boldsymbol{y}$ is recorded as k. In order to determine k, this paper proposes the cumulative information percentage method as follows

$$\sum_{i=1}^{k} I(\boldsymbol{x}_i, \boldsymbol{y}) / \sum_{i=1}^{m} I(\boldsymbol{x}_i, \boldsymbol{y}) \times 100\% \geq ratio, \tag{10}$$

where *ratio* represents the minimum cumulative information ratio. This parameter can be determined by the cross-validation method. If the value is too large, redundant information and noise will be selected as the input to the prediction model. When the value is 1, all variables in the energy storage power station operating data will be determined as key factors. If the value is too small, the key factors are not comprehensively selected. When the value is 0, no variable is determined as the key factor. Whether the value is too large or too small, the accuracy of the prediction model will be affected.

According to the correlation with $\boldsymbol{y}$, $\boldsymbol{X}$ is divided into two sub-blocks, which are denoted as related dataset $\boldsymbol{X}_r$ and unrelated dataset $\boldsymbol{X}_u$ respectively.

Given that quantitative calculation is not enough for accurate prediction, by analyzing the system mechanism of the energy storage power station, the series and parallel structure of the battery and the experience of experts, the key factors which affect the key parameter to be predicted are determined as $\boldsymbol{X}_R$. Finally, the factor dataset $\boldsymbol{X}_{Rr}$ affecting the key parameter $\boldsymbol{y}$ is determined as follows:

$$\boldsymbol{X}_{Rr} = \boldsymbol{X}_R \cap \boldsymbol{X}_r = \left[\tilde{\boldsymbol{x}}_1 \tilde{\boldsymbol{x}}_2 \cdots \tilde{\boldsymbol{x}}_p\right] \in \boldsymbol{R}^{N \times p}. \tag{11}$$

B. PREDICTION MODEL CONSTRUCTION

On the basis of historical dataset $\boldsymbol{X}_{Rr} = \left[\tilde{\boldsymbol{x}}_1 \tilde{\boldsymbol{x}}_2 \cdots \tilde{\boldsymbol{x}}_p\right] \in \boldsymbol{R}^{N \times p}$ and $\boldsymbol{y} \in \boldsymbol{R}^{N \times 1}$, the regression model is established to predict the future trend of $\boldsymbol{y}$. In order to realize the simultaneous prediction of the key parameter for a period of time M and improve the prediction accuracy, multiple elastic networks with different time delays are established. Suppose the delay value is set as $L (L = 1, 2, \cdots, M)$. In the prediction model with delay L, the independent data matrix and the data matrix to be predicted are established as follows:

$$\dot{\boldsymbol{X}}_{Rr} = \begin{bmatrix} \tilde{x}_1(1) & \tilde{x}_2(1) & \cdots & \tilde{x}_p(1) \\ \tilde{x}_1(2) & \tilde{x}_2(2) & \cdots & \tilde{x}_p(2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_1(N-L) & \tilde{x}_2(N-L) & \cdots & \tilde{x}_p(N-L) \end{bmatrix} \in \boldsymbol{R}^{(N-L) \times p} \tag{12}$$

$$\dot{\boldsymbol{y}} = \begin{bmatrix} y(L+1) \\ y(L+2) \\ \vdots \\ y(N) \end{bmatrix} \in \boldsymbol{R}^{(N-L) \times 1}, \tag{13}$$

where $\tilde{x}_i(j)$ is the $j-th$ sample of $\tilde{\boldsymbol{x}}_i$, and $y(j)$ is the $j-th$ sample of $\mathbf{y}$.

The regression coefficient $\tilde{\boldsymbol{\beta}}^L = \left[\tilde{\beta}_1^L, \tilde{\beta}_2^L, \cdots, \tilde{\beta}_p^L\right]^T$ between $\dot{\boldsymbol{X}}_{Rr}$ and $\dot{\boldsymbol{y}}$ is calculated as follows:

$$\tilde{\boldsymbol{\beta}}^L = \arg\min \left[\sum_{j=1}^{N-L} \left(y(L+j) - \sum_{i=1}^{p} \tilde{\beta}_i^L \tilde{x}_i(j)\right)^2\right] \cdot \varphi \sum_{i=1}^{p} \left|\tilde{\beta}_i^L\right| + (1-\varphi) \sum_{i=1}^{p} \left(\tilde{\beta}_i^L\right)^2 \leq t \tag{14}$$

After the regression coefficient $\tilde{\boldsymbol{\beta}}^L = \left[\tilde{\beta}_1^L, \tilde{\beta}_2^L, \cdots, \tilde{\beta}_p^L\right]^T$ is obtained, the $L-th$ predicted value $\hat{y}(T+L)$ at time T can be calculated as follows:

$$\hat{y}(T+L) = \tilde{x}_1(T) * \tilde{\beta}_1^L + \tilde{x}_2(T) * \tilde{\beta}_2^L + \cdots + \tilde{x}_p(T) * \tilde{\beta}_p^L. \tag{15}$$

By establishing multiple prediction models with different time delays $L = 1, 2, \cdots, M$, the predicted parameter values $\left[\hat{y}(T+1), \hat{y}(T+2), \cdots, \hat{y}(T+M)\right]$ within a period of time M can be obtained simultaneously as follows

$$\begin{aligned} \hat{y}(T+1) &= \tilde{x}_1(T) * \tilde{\beta}_1^1 + \tilde{x}_2(T) * \tilde{\beta}_2^1 + \cdots + \tilde{x}_p(T) * \tilde{\beta}_p^1 \\ \hat{y}(T+2) &= \tilde{x}_1(T) * \tilde{\beta}_1^2 + \tilde{x}_2(T) * \tilde{\beta}_2^2 + \cdots + \tilde{x}_p(T) * \tilde{\beta}_p^2 \\ &\vdots \\ \hat{y}(T+M) &= \tilde{x}_1(T) * \tilde{\beta}_1^M + \tilde{x}_2(T) * \tilde{\beta}_2^M + \cdots + \tilde{x}_p(T) * \tilde{\beta}_p^M. \end{aligned} \tag{16}$$

Comparions: To predict the parameter for a period of time, there are mainly three differences between the proposed MEN-TD method and the EN method listed in (5), (6), and (7). First, the EN method uses the data at time $T-M+1, T-M+2, \cdots, T$ as the input, and the proposed MEN-TD method only uses the data at time T as the input. Second, the EN method constructs one model $\boldsymbol{\beta}^M$, and the proposed MEN-ED method constructs M models $\tilde{\boldsymbol{\beta}}^1, \tilde{\boldsymbol{\beta}}^2, \cdots, \tilde{\boldsymbol{\beta}}^M$. Third, the EN method uses m variables of the operation data, and the proposed MEN-TD method uses p variables which are related to the predicted key parameter.

For the complexity of the algorithm, in the offline modeling stage, due to the need to establish multiple models with different delays in the MEN-TD method, the computational complexity of the proposed method is higher than the EN method. However, in the online prediction stage, since the model has been established, the computational complexity of the proposed method is close to the EN method.

Since the data at the current time T contains newest information compared with the data at $T-M+1$, $T-M+2, \cdots, T$ and the model accuracy of $\tilde{\boldsymbol{\beta}}^1, \tilde{\boldsymbol{\beta}}^2, \cdots, \tilde{\boldsymbol{\beta}}^M$ is high than $\boldsymbol{\beta}^M$, it can be concluded that the proposed MEN-TD method can achieve the higher prediction accuracy.

Remark: The historical offline normal dataset is used to construct the prediction model in the offline modeling stage. Once the prediction model is constructed, the predicted value $\left[\hat{y}(T+1), \hat{y}(T+2), \cdots, \hat{y}(T+M)\right]$ in the period M after the current time T can be obtained in case of importing the real-time data $[x_1(T)x_2(T)\cdots x_m(T)]$.

C. CONTROL LIMIT DETERMINATION

The historical normal dataset $\boldsymbol{y} \in \boldsymbol{R}^{N\times 1}$ of the key parameter can provide data distribution information in the normal status and characterize the normal status of the key parameter. In order to judge whether the predicted value $\left[\hat{y}(T+1), \hat{y}(T+2), \cdots, \hat{y}(T+M)\right]$ is under the normal status and detect the fault in advance, it is necessary to determine the control limit of the key parameter under normal status.

The KDE method can be used to mine the distribution information from a large amount of data. In the field of statistics, the KDE method is used to estimate the probability density function of random variables. In the field of early fault detection and prognostics, the KDE method has been widely used to determine the control limit. Univariate kernel estimation is defined as follows

$$f(y) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{y-y(i)}{h}\right), \tag{17}$$

where $f(y)$ is the estimated probability density function, h is the smoothing parameter, N is the number of samples, and K is the kernel function. The kernel function can transform the inner product operation of the high-dimensional space into the kernel function calculation of the low-dimensional input space, effectively solving the problem of ‘‘dimensional disaster’’ in the high-dimensional feature space. The most widely used kernel function is the Gaussian kernel function.

Based on the KDE, the control limit $y_{\lim}$ of the key parameter to be predicted can be determined. If the predicted value $\left[\hat{y}(T+1), \hat{y}(T+2), \cdots, \hat{y}(T+M)\right]$ is less than the control limit, the energy storage power station is considered to be in the normal status for the subsequent time period M. Otherwise, it is considered that the status of the energy storage power station would occur the fault, and early fault detection and prognostics should be made.

D. OFFLINE MODELING STAGE AND ONLINE FAULT PROGNOSTICS STAGE

Offline modeling stage:

1) Collect historical normal dataset $\boldsymbol{y}$ of the key parameter to be predicted and historical normal datasets $\boldsymbol{X}$ of the operation data.
2) Standardize $\boldsymbol{y}$ and $\boldsymbol{X}$ using the z-score method to make the mean is 0 and the standard deviation is 1.
3) Combine the quantitative and qualitative methods to determine $\boldsymbol{X}_{Rr}$ which would affect $\boldsymbol{y}$.
4) Construct multiple elastic networks prediction models with different time delays.
5) Determine the control limit $y_{\lim}$ of $\boldsymbol{y}$ using the KDE method.

Online fault prognostics stage:

1) Obtain the operation data $[x_1(T)x_2(T)\cdots x_m(T)]$ of at the current moment T.
2) Standardize $[x_1(T)x_2(T)\cdots x_m(T)]$ with the mean and the standard deviation of $\boldsymbol{X}$.
3) Select the current factor $\left[\tilde{x}_1(T)\tilde{x}_2(T)\cdots\tilde{x}_p(T)\right]$ corresponding to $\boldsymbol{y}$ from the standardized operating data.
4) Calculate the predicted value $[\hat{y}(T+1), \hat{y}(T+2), \cdots, \hat{y}(T+M)]$ in the period M after the current time T.
5) By comparing the predicted value and the control limit, early fault detection and prognostics can be made.

IV. EXAMPLES AND APPLICATIONS

By analyzing the historical alarm information collected from the energy storage power station in Zhenjiang of China, the battery cell voltage is an important parameter, and almost all alarms are battery cell voltage alarms. Therefore, the battery cell voltage is selected as the energy storage power station key parameter to be predicted. The historical dataset of the energy storage power station is divided into two categories: battery cell data and battery non-cell data. The battery cell data include the voltage of 228 battery cells, the temperature of 228 battery cells, and the SOC of 228 battery cells. The monitored variables of battery non-cell data are listed in TABLE 1.

According to the historical normal dataset, the quantitative calculation based on the mutual information and the qualitative analysis based on the system mechanism of the energy storage power station, the series and parallel structure of the battery and the experience of experts are used to determine the key factors affecting the battery cell voltage. The data of the energy storage power station is sampled in minutes. The energy storage power station data within 1000 minutes before the current time T, that is, the data from the time T-999 to T, are collected as the training dataset to establish the early fault detection and prognostics model. A total of 440 data after the current time T, that is, the data from the time T + 1 to T + 440, are collected as the testing dataset. According to practical needs, in order to test the prediction ability for a period of time, three cases (5min, 10min, and 20min according to actual system needs) are set to compare the prediction performance. To show the advantage of the proposed

TABLE 1. Monitored variables of non-cell data.

number	battery non-cell data name
1	stack voltage
2	stack current
3	stack SOC
4	maximum stack voltage
5	minimum stack voltage
6	maximum stack temperature
7	minimum stack temperature
8	stack charge
9	stack discharge
10	total accumulated charge capacity of the stack
11	total accumulated discharge capacity of the stack
12	total cluster voltage
13	cluster precharge voltage
14	cluster current
15	insulation resistance+
16	insulation resistance-
17	cluster SOC remaining capacity
18	cluster average voltage
19	cluster maximum voltage
20	cluster minimum voltage
21	cluster average temperature
22	cluster maximum temperature
23	cluster minimum temperature
24	cluster cumulative charging capacity
25	cluster cumulative discharging capacity

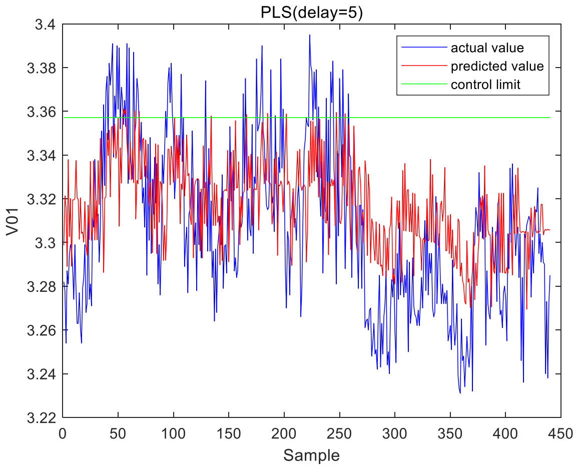

FIGURE 1. The first battery cell voltage prediction result of the PLS method with delay = 5 for 5min prediction.

MEN-ED method, the widely used PLS method with fixed delay, the EN method with fixed delay, and the MEN-ED method without key factors selection are compared.

For 5min prediction, Fig. 1, Fig. 2, Fig. 3, and Fig. 4 give the first battery cell voltage prediction result of the PLS method with delay = 5, the EN method with delay = 5, the MEN-TD method without key factors selection and the MEN-TD method. For 10min prediction, Fig. 5, Fig. 6, Fig. 7, and Fig. 8 give the first battery cell voltage prediction result of the PLS method with delay = 10, the EN method with delay = 10, the MEN-TD method without key factors selection and the MEN-TD method. For 20min prediction, Fig. 9, Fig. 10, Fig. 11, and Fig. 12 give the first battery cell voltage prediction result of the PLS method with delay = 20, the EN method with delay = 20, the MEN-TD method without key factors selection and the MEN-TD method. In these figures,

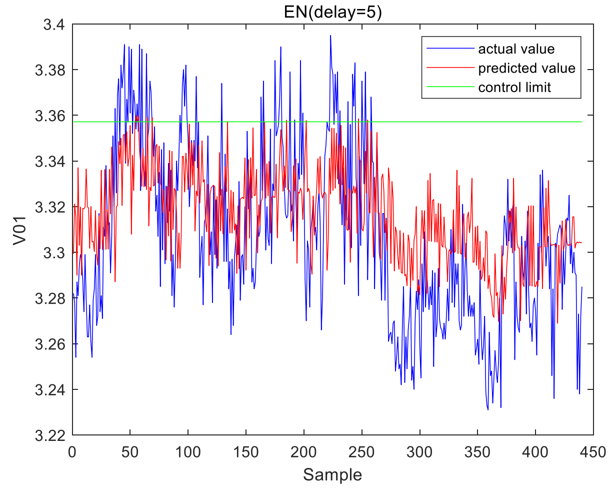

FIGURE 2. The first battery cell voltage prediction result of the EN method with delay = 5 for 5min prediction.

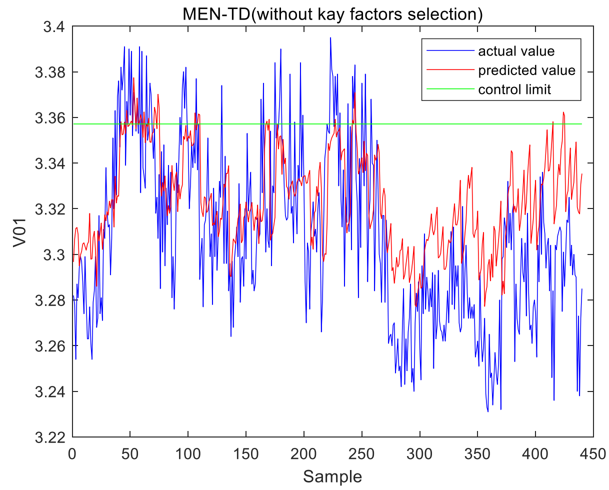

FIGURE 3. The first battery cell voltage prediction result of the MEN-TD method without key factors selection for 5min prediction.

the blue line represents the actual value of the first battery cell voltage, the red line represents the predicted value of the first battery cell voltage, and the green line represents the control limit of the first battery cell voltage. In addition, the root mean square errors of three methods are listed in TABLE 2, where the obtained minimum value is in bold.

As listed in TABLE 2, Fig. 4, Fig. 8, Fig. 12, for three cases (5min, 10min, 20min), the proposed MEN-TD can obtain the best root mean square error, where the root mean square errors for 5min, 10min, 20min are 0.03149, 0.3257, 0.03722. In contrast to the PLS method with single fixed delay, the EN method with single fixed delay can obtain better root mean square errors (0.03413 > 0.03379, 0.03725 > 0.03681, 0.04011 > 0.03960) for three cases, which can show the superiority of the EN method. Compared with the EN method with single fixed delay, the better prediction result (0.03379 > 0.03149, 0.03681 > 0.03275, 0.03960 > 0.03722) obtained by the MEN-TD method can show the advantage of constructing multiple EN models with

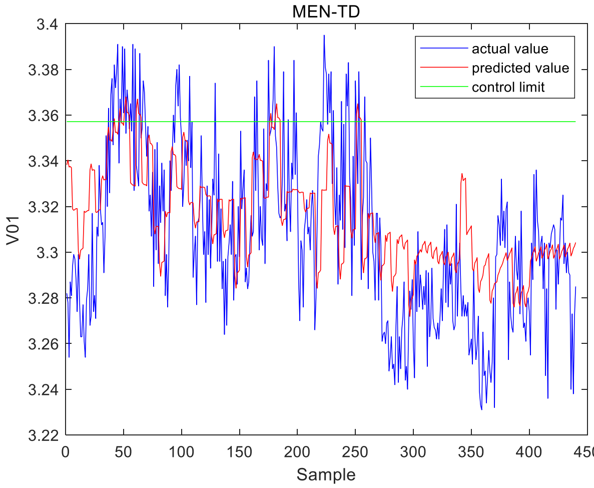

FIGURE 4. The first battery cell voltage prediction result of the MEN-TD method for 5min prediction.

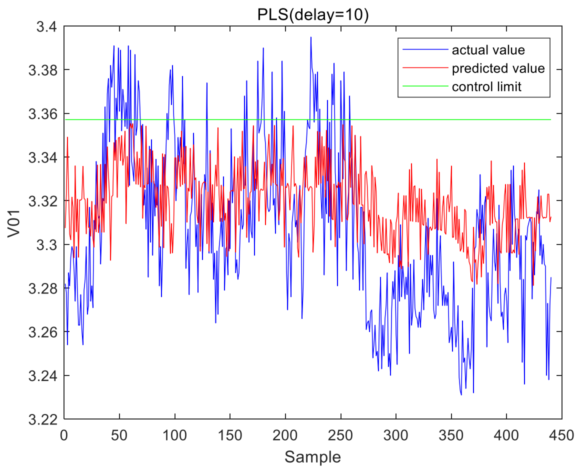

FIGURE 5. The first battery cell voltage prediction result of the PLS method with delay = 10 for 10min prediction.

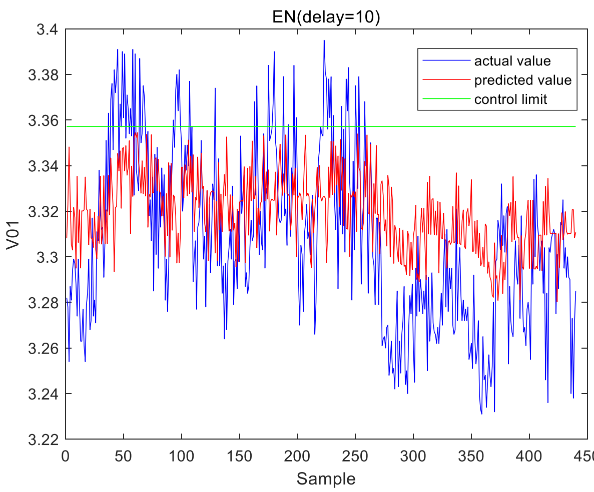

FIGURE 6. The first battery cell voltage prediction result of the EN method with delay = 10 for 10min prediction.

different time delays. In contrast to the MEN-TD without key factors selection method, the better prediction result

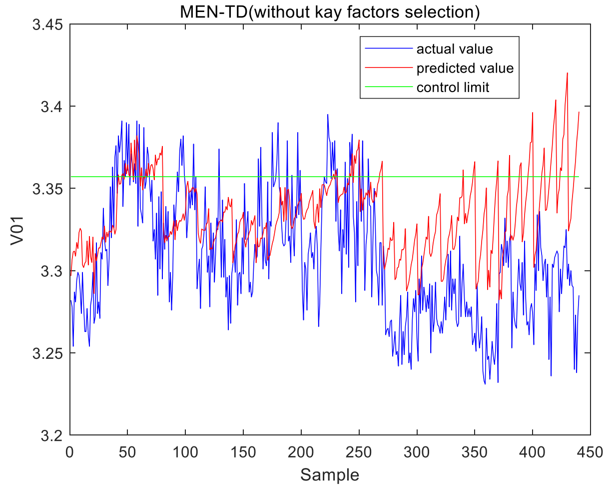

FIGURE 7. The first battery cell voltage prediction result of the MEN-TD method without key factor selection for 10min prediction.

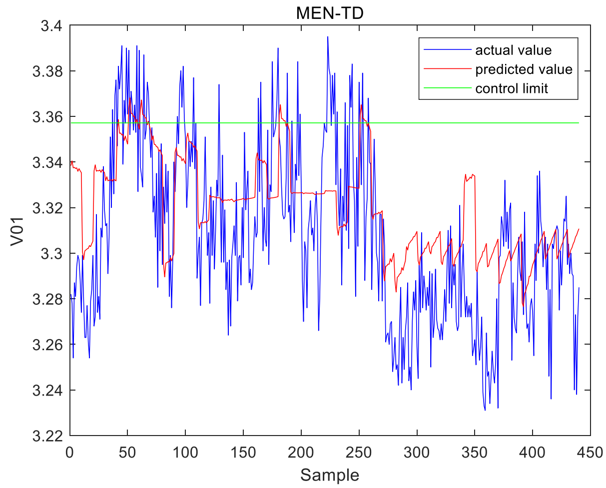

FIGURE 8. The first battery cell voltage prediction result of the MEN-TD method for 10min prediction.

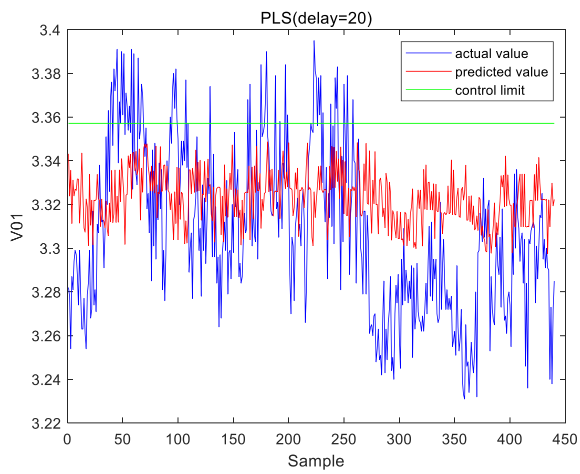

FIGURE 9. The first battery cell voltage prediction result of the PLS method with delay = 20 for 20min prediction.

(0.03396 > 0.03149, 0.04568 > 0.03275, 0.06360 > 0.03722) obtained by the MEN-TD method can prove the advantage of the key factors selection.

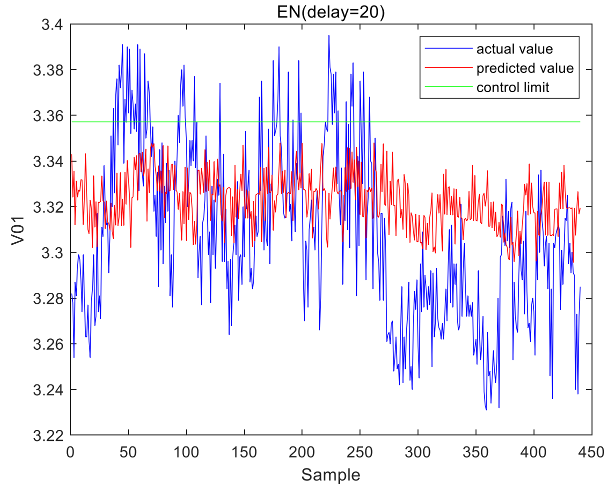

FIGURE 10. The first battery cell voltage prediction result of the EN method with delay = 20 for 20min prediction.

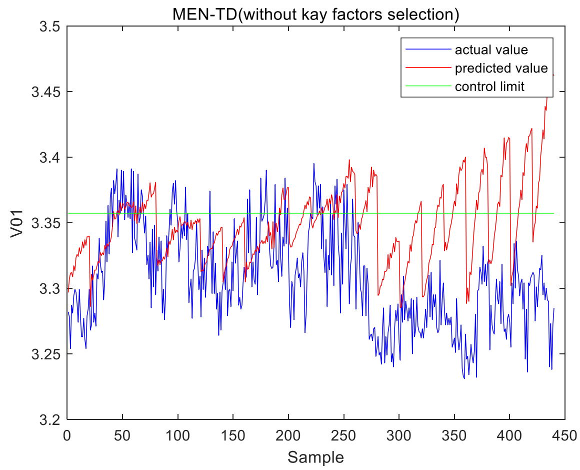

FIGURE 11. The first battery cell voltage prediction result of the MEN-TD method without key factor selection for 20min prediction.

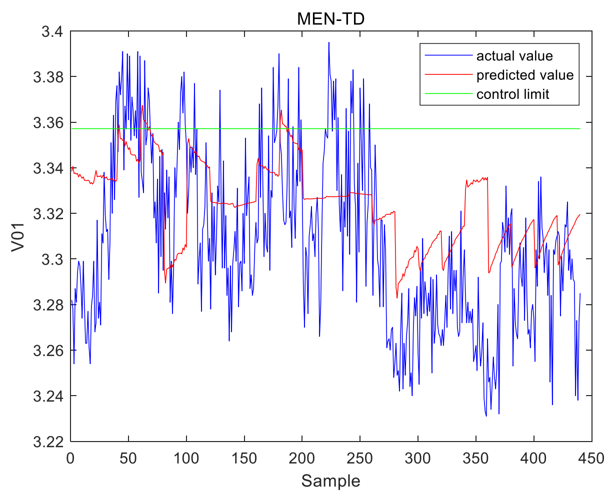

FIGURE 12. The first battery cell voltage prediction result of the MEN-TD method for 20min prediction.

As plotted in Fig.1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Fig. 10, Fig. 11, and Fig. 12, the trend of the blue line representing the actual value and the red line representing the predicted value can coincide. From Fig. 1, Fig. 5, Fig. 9, for different time periods need to be predicted at the same time, the prediction error is related to the delay value (0.03413 < 0.03725 < 0.04011, 0.03379 < 0.03681 < 0.03960, 0.03396 < 0.04568 < 0.06360, 0.03149 < 0.03257 < 0.03722). The smaller the delay value, the more accurate the prediction result. Although the delay value will affect the prediction accuracy, as long as the delay value is controlled within an appropriate range, the predicted trend can match the actual trend.

TABLE 2. Root mean square error.

Time period	PLS(with single fixed delay)	EN(with single fixed delay)	MEN-TD(without key factors selection)	MEN-TD
5min	0.03413	0.03379	0.03396	**0.03149**
10min	0.03725	0.03681	0.04568	**0.03257**
20min	0.04011	0.03960	0.06360	**0.03722**

By comparing with the control limit, it can realize early fault detection and prognostics. For the same time period, the proposed MEN-TD method has the best prediction accuracy. Moreover, the control limit of the first battery cell voltage is the same. Therefore, the high prediction accuracy means that the key parameter can be more accurately alarmed. In other words, the accuracy of the proposed MEN-TD method is highest among three methods.

V. CONCLUSION

This paper proposes the multiple elastic networks with time delays method for early fault detection and prognostics of the energy storage power station. By comparing with the control limit under the normal condition, it is expected to be able to give fault prognostics and ensure the healthy operation of the energy storage power station. First, the quantitative calculation method and the qualitative analysis are combined to obtain key factors that affect the energy storage power station key parameter. Then, multiple elastic networks with different time delays are constructed to predict the change trend of the key parameter over a period of time simultaneously. Based on a large amount of historical normal data, the kernel density estimation is used to determine the control limit of the energy storage power station parameter for early fault detection and prognostics. Finally, the data collected on the Zhenjiang energy storage power station was tested. The results show that the predicted value can match the actual value to a certain extent, which shows the effectiveness of the proposed MEN-TD method in this paper.

REFERENCES

[1] L. Wang and H. Shi, "Multivariate statistical process monitoring using an improved independent component analysis," *Chem. Eng. Res. Des.*, vol. 88, no. 4, pp. 403–414, Apr. 2010.
[2] B. Song and H. Shi, "Fault detection and classification using quality-supervised double-layer method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 10, pp. 8163–8172, Oct. 2018.
[3] C. Bhargava, P. K. Sharma, M. Senthilkumar, S. Padmanaban, V. K. Ramachandaramurthy, Z. Leonowicz, F. Blaabjerg, and M. Mitolo, "Review of health prognostics and condition monitoring of electronic components," *IEEE Access*, vol. 8, pp. 75163–75183, 2020.

[4] B. Song and H. B. Shi, "Temporal spatial global locality projections for multimode process monitoring," *IEEE Access*, vol. 6, pp. 9740–9749, 2018.

[5] L. Zhou, G. Li, Z. Song, and S. J. Qin, "Autoregressive dynamic latent variable models for process monitoring," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 1, pp. 366–373, Jan. 2017.

[6] B. Song, X. Zhou, S. Tan, H. Shi, B. Zhao, and M. Wang, "Process monitoring via key principal components and local information based weights," *IEEE Access*, vol. 7, pp. 15357–15366, 2019.

[7] Z. Lv, X. Yan, Q. Jiang, and N. Li, "Just-in-time learning-multiple subspace support vector data description used for non-Gaussian dynamic batch process monitoring," *J. Chemometrics*, vol. 33, no. 6, pp. 1–13, Jun. 2019.

[8] B. Cai, X. Shao, Y. Liu, X. Kong, H. Wang, H. Xu, and W. Ge, "Remaining useful life estimation of structure systems under the influence of multiple causes: Subsea pipelines as a case study," *IEEE Trans. Ind. Electron.*, vol. 67, no. 7, pp. 5737–5747, Jul. 2020.

[9] S. Eickmeier, W. Lemke, and M. Marcellino, "Classical time varying factor-augmented vector auto-regressive models-estimation, forecasting and structural analysis," *J. Roy. Stat. Soc. A, Statist. Soc.*, vol. 178, no. 3, pp. 493–533, Jun. 2015.

[10] Y. Zhu and G. Zhou, "Technical analysis: An asset allocation perspective on the use of moving averages," *J. Financial Econ.*, vol. 92, no. 3, pp. 519–544, Jun. 2009.

[11] G. Chen, X. Duan, J. H. Zhang, Q. He, and X. Wu, "A new approach for online identification of low frequency oscillation modes based on autoregressive moving-average model," *Power Syst. Technol.*, vol. 34, no. 11, pp. 48–54, Nov. 2010.

[12] M. Khashei, M. Bijari, and G. A. Raissi Ardali, "Improvement of autoregressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs)," *Neurocomputing*, vol. 72, nos. 4–6, pp. 956–967, Jan. 2009.

[13] H. Moeeni and H. Bonakdari, "Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model," *Stochastic Environ. Res. Risk Assessment*, vol. 31, no. 8, pp. 1997–2010, Oct. 2017.

[14] T. Singh, "On the GARCH estimates of exchange rate volatility in India," *Appl. Econ. Lett.*, vol. 9, no. 6, pp. 391–395, May 2002.

[15] B. Song, X. Zhou, H. Shi, and Y. Tao, "Performance-indicator-oriented concurrent subspace process monitoring method," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5535–5545, Jul. 2019.

[16] S. Zhang and C. Zhao, "Slow-feature-analysis-based batch process monitoring with comprehensive interpretation of operation condition deviation and dynamic anomaly," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3773–3783, May 2019.

[17] T. Mehmood, K. H. Liland, L. Snipen, and S. Sæbø, "A review of variable selection methods in partial least squares regression," *Chemometric Intell. Lab. Syst.*, vol. 118, pp. 62–69, Aug. 2012.

[18] T. Lee, "On algorithms for ordinary least squares regression spline fitting: A comparative study," *J. Stat. Comput. Simul.*, vol. 72, no. 8, pp. 647–663, Jan. 2002.

[19] S. Yin, X. Zhu, and O. Kaynak, "Improved PLS focused on key-performance-indicator-related fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1651–1658, Mar. 2015.

[20] L. Wang and H. Shi, "Improved kernel PLS-based fault detection approach for nonlinear chemical processes," *Chin. J. Chem. Eng.*, vol. 22, no. 6, pp. 657–663, Jun. 2014.

[21] T. Van Gestel, J. A. K. Suykens, D.-E. Baestaens, A. Lambrechts, G. Lanckriet, B. Vandaele, B. De Moor, and J. Vandewalle, "Financial time series prediction using least squares support vector machines within the evidence framework," *IEEE Trans. Neural Netw.*, vol. 12, no. 4, pp. 809–821, Jul. 2001.

[22] A. G. José, J. L. Mateo, and M. P. José, "Learning Bayesian networks by hill climbing: Efficient methods based on progressive restriction of the neighborhood," *Data Mining Knowl. Discovery*, vol. 22, nos. 1–2, pp. 106–148, Jan. 2011.

[23] L. Chen, Y. Liu, J. Zhao, W. Wang, and Q. Liu, "Prediction intervals for industrial data with incomplete input using kernel-based dynamic Bayesian networks," *Artif. Intell. Rev.*, vol. 46, no. 3, pp. 307–326, Oct. 2016.

[24] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mech. Syst. Signal Process.*, vol. 80, pp. 31–44, Dec. 2016.

[25] B. Cai, Y. Zhao, H. Liu, and M. Xie, "A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems," *IEEE Trans. Power Electron.*, vol. 32, no. 7, pp. 5590–5600, Jul. 2017.

[26] B. Cai, X. Kong, Y. Liu, J. Lin, X. Yuan, H. Xu, and R. Ji, "Application of Bayesian networks in reliability evaluation," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2146–2157, Apr. 2019.

[27] C. Hamzacebi, "Improving artificial neural networks' performance in seasonal time series forecasting," *Inf. Sci.*, vol. 178, no. 23, pp. 4550–4559, Dec. 2008.

[28] A. Khosravi, L. Machado, and R. O. Nunes, "Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil," *Appl. Energy*, vol. 224, pp. 550–566, Aug. 2018.

[29] M. Ghiassi and S. Nangoy, "A dynamic artificial neural network model for forecasting nonlinear processes," *Comput. Ind. Eng.*, vol. 57, no. 1, pp. 287–297, Aug. 2009.

[30] L. Saad Saoud, F. Rahmoune, V. Tourtchine, and K. Baddari, "Fully complex valued wavelet network for forecasting the global solar irradiation," *Neural Process. Lett.*, vol. 45, no. 2, pp. 475–505, Apr. 2017.

[31] M. Sheikhan and N. Mohammadi, "Time series prediction using PSO-optimized neural network and hybrid feature selection algorithm for IEEE load data," *Neural Comput. Appl.*, vol. 23, nos. 3–4, pp. 1185–1194, Sep. 2013.

[32] T. Kuremoto, S. Kimura, K. Kobayashi, and M. Obayashi, "Time series forecasting using a deep belief network with restricted Boltzmann machines," *Neurocomputing*, vol. 137, pp. 47–56, Aug. 2014.

[33] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3235–3243, Jul. 2018.

[34] J. T. Turner, "Time series analysis using deep feed forward neural networks," *Diss. Theses Gradworks*, vol. 105, pp. 564–580, Jun. 2014.

[35] G. Liu, H. Gu, X. Shen, and D. You, "Bayesian long short-term memory model for fault early warning of nuclear power turbine," *IEEE Access*, vol. 8, pp. 50801–50813, 2020.

[36] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[37] B. Song, H. Yan, H. Shi, and S. Tan, "Multisubspace elastic network for multimode quality-related process monitoring," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5874–5883, Sep. 2020.

DONGLIANG GUO received the Ph.D. degree. He is currently a Senior Engineer at the State Grid Jiangsu Electric Power Company Ltd. Research Institute, Nanjing, China. His research interests include high voltage technology, condition monitoring, and power grid equipment operation and maintenance technology.

WEN YANG received the Ph.D. degree in control theory and control engineering from Shanghai Jiao Tong University. She is currently a Professor at the Department of Automation, East China University of Science and Technology. Her research interests include information security, feature extraction, and fault detection.

FENGBO TAO received the Ph.D. degree. He is currently a Senior Engineer at the State Grid Jiangsu Electric Power Company Ltd. Research Institute, Nanjing, China. His research interests include data driven fault detection, high voltage technology, and power grid equipment operation and maintenance technology.

BING SONG received the B.E. degree in automation and the Ph.D. degree in control theory and control engineering from the East China University of Science and Technology, Shanghai, China, in 2012 and 2017, respectively. He is currently an Associate Professor at the Department of Automation, East China University of Science and Technology. His research interests include feature extraction, fault detection, fault diagnosis, and multimode process monitoring.

HUI LIU received the Ph.D. degree. He is currently a Senior Engineer at Shanghai Power & Energy Storage Battery System Engineering Technology Company Ltd., Shanghai, China. His research interests include data analysis, process monitoring, and energy storage battery system design and development.

LEI SUN received the master's degree. He is currently a Senior Engineer at the State Grid Jiangsu Electric Power Company Ltd. Research Institute, Nanjing, China. His research interests include fault diagnosis, high voltage technology, and power grid equipment operation and maintenance technology.

JIALE WANG is currently pursuing the master's degree with the East China University of Science and Technology, Shanghai, China. His research interests include data analysis, feature extraction, and fault detection.