

Research Article

Novel Condition Monitoring Method for Wind Turbines Based on the Adaptive Multivariate Control Charts and SCADA Data

Qinkai Han,¹ Zhentang Wang,² and Tao Hu ²

¹The State Key Laboratory of Tribology, Tsinghua University, Beijing 100084, China

²School of Mathematical Sciences, Capital Normal University, Beijing 100048, China

Correspondence should be addressed to Tao Hu; hutao@cnu.edu.cn

Received 12 July 2020; Revised 12 August 2020; Accepted 31 August 2020; Published 15 September 2020

Academic Editor: Changqing Shen

Copyright © 2020 Qinkai Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel condition monitoring method based on the adaptive multivariate control charts and the supervisory control and data acquisition (SCADA) system is developed. Two types of control charts are adopted: one is the adaptive exponential weighted moving average (AEWMA) control chart for abnormal state detection, and the other is the multivariate exponential weighted moving average (MEWMA) control chart for anomaly location determination. Optimization procedures for these control charts are implemented to achieve minimum out-of-control average running length. Multivariate regression analysis is utilized to obtain the normal condition prediction model of wind turbine with fault-free SCADA data. After comparing the regression accuracy of several popular algorithms in the MRA, the random forest is adopted for feature selection and regression prediction. Various tests on the wind turbine with normal and abnormal states are conducted. The performance and robustness of various control charts are compared comprehensively. Compared with conventional control charts, the AEWMA control chart is more sensitive to the abnormal state and thus has a more effective anomaly identification ability and better robustness. It is shown that the MEWMA control chart combined with the out-of-limit number index can effectively locate and identify the abnormal component.

1. Introduction

With the increasing sustainable energy and environmental demands, wind energy has become one of the world's fastest growing renewable and green energy sources. Due to unstable and unpredictable wind speed characteristics and energy potentials, which are very sensitive to variations in topography and weather patterns, the cost ratios of the operation and maintenance (O&M) costs over the total energy costs per unit output electrical energy from wind turbine systems are considerably high, which is up to 20%~25% [1]. Such high ratios of O&M costs may impede the applications of wind turbine systems compared to other renewable energy systems, such as solar photovoltaic or hot water systems. Consequently, effective condition monitoring (CM) methods for wind turbines are essential for maintenance decisions, which aim to reduce O&M cost [2]. Various signals, such as vibration [3], acoustic emission [4], and motor current [5, 6], have been utilized for wind turbine CM

systems. However, these approaches require the installation of additional sensors and data acquisition devices, which increase the capital cost and wiring complexity of wind turbine systems. Supervisory control and data acquisition (SCADA) systems have been installed in most modern wind turbines to monitor operational performances.

Currently, the SCADA signal has received a lot of attention owing to its application in wind speed-power forecasting [7–9], wind power assessment [10, 11], and wind farm performance analysis [12]. A typical SCADA system records comprehensive wind turbine condition parameters, including temperatures (e.g., bearing temperature and oil temperature), wind parameters (e.g., wind speed and wind direction), and energy conversion parameters (e.g., output power, pitch angle, and rotor speed), which would be fault informative. Since no additional sensors or data acquisition devices are needed, the wind turbine CM method based on SCADA data is a cost-effective approach to improve the reliability of wind turbines [13].

Building a model to predict the normal behavior of SCADA parameters is the first issue of the wind turbine CM system. By using advanced SCADA data mining methods, various normal condition prediction models (NCPMs) have been developed to detect the significant changes in wind turbine behavior prior to anomaly occurrences. Kusiak et al. [14–16] first employed various data mining algorithms to construct NCPMs for wind turbine anomalies. After detailed comparisons based on the SCADA data collected at a large wind farm, they found that the random forest (RF) algorithm models provided the best accuracy [14]. Gill et al. [17] developed a probabilistic model of a power curve for CM purposes based on copula statistics. Its practical use was demonstrated on the SCADA data taken from a fleet of operational wind turbines. The adaptive neurofuzzy interference systems [18, 19] and neural networks [20] have also been adopted to develop various NCPMs. Wang et al. [21] proposed a new NCPM based on heterogeneous signals and information collected from the SCADA system. A linear mixture self-organizing map classifier was applied to differentiate abnormal types. After simulations were carried out on the actual data from a wind farm in north China, the proposed technique was shown to be effective for abnormality detection and prediction. Recently, the Bayesian framework [22], spatiotemporal pattern network [23], and mathematical optimization models [24] were introduced for the early and unsupervised fault diagnosis of wind turbines using SCADA data.

For a given NCPM, the relationship between the input and output SCADA state variables of the wind turbine could be learned. Subsequently, the departure of the current turbine state from the NCPM could be measured online and yield a time series of residuals. The control chart from statistical process control is a time-honored tool to monitor the residuals [25]. If the residuals are statistically different from a normal (or fault-free) reference, the process is considered out of control, and an alarm would be raised accordingly. In recent years, the NCPM combined with control charts has been increasingly used in wind turbine CM systems. Most studies [26–30] used the Shewhart-type control charts, which have been proven to be very effective for detecting greater shifts [25]. However, they are slow in reacting to small and moderate shifts in the mean process. In that regard, the exponential weighted moving average (EWMA) control chart was developed to provide more sensitivity to small mean shifts [25]. Cambron et al. [31–33] first applied the EWMA control chart for the CM of wind turbines. Using several applications on the actual SCADA data, the results showed that a shift of 3.4% in annual energy production over a period of 5 years could be detected in time to plan proper maintenance. Helbing and Ritter [34] explained a straightforward method to incorporate non-constant variance to construct a flexible EWMA control chart. Simulations showed that the FEWMA has lower false alarm rate than the EWMA. Wang et al. [35] deployed the EWMA control chart to derive the criteria for detecting the oil temperature shifts of wind turbine gearboxes. Yang et al. [36] proposed an approach combining data mining and control charts for fault detection in actual wind turbines. Both EWMA and multivariate EWMA (MEWMA) control charts were constructed for comparisons. Their observations

showed that the MEWMA is more suitable for early detection and avoidance of errors.

Although the EWMA control chart can provide greater sensitivity to small shifts, it is not as effective as the Shewhart chart, where the shifts in the process mean level are relatively large due to the inertia problem [37]. In actual applications, such as monitoring of wind turbines, the shift of the residuals from the NCPM is unknown, which might cause the insufficiency of the EWMA control chart if the larger shift appears. To overcome the inertia problem, Capizzi and Masarotto [38] first presented an adaptive EWMA (AEWMA) by adaptively adjusting the weight on past observations according to a function of the prediction error. Later, Shu [39] extended the idea of the AEWMA chart on monitoring process locations to the case of monitoring process dispersion. The AEWMA chart is a smooth combination of the Shewhart and EWMA charts; thus, it can reduce the inertia problem. Using the examples on capsule weights and simulated data, both Capizzi and Masarotto [38] and Shu [39] showed that the AEWMA control chart is able to offer an overall good detection performance against shifts of different sizes. However, in the CM of wind turbines, the residual data would be more complicated, and the possibility of the AEWMA control chart holding a better performance than the EWMA chart is still unknown. To the authors' knowledge, the AEWMA control chart has not been used in the CM of wind turbines in the open literature.

In actual engineering, it is not only expected to alarm an abnormal state as early as possible, but determining the cause and location of the abnormal state is also expected. Since the SCADA system records condition parameters of the main components of wind turbines (e.g., the blade, gearbox, main bearing, and generator), the components with the abnormal state might be identified by modeling the control charts of these multivariate conditional parameters. Lately, Yang et al. [36] used the MEWMA to determine which components are likely to contribute to the fault. Their results showed that the MEWMA has a good potential in locating anomaly. The limitation of Yang et al.'s study [36] is that only specified values of MEWMA parameters were tested, indicating that the presented MEWMA might not be the optimal control chart. The optimal design of MEWMA should be conducted to fully realize the potential of the MEWMA in the CM of wind turbines.

A literature review indicated that only a few studies have used the multivariate control charts for the CM of wind turbines; this is particularly true for the abnormal state alarm of wind turbine using adaptive control charts. Moreover, there have been few attempts to comprehensively compare the performance and robustness of both EWMA and AEWMA control charts in monitoring the residuals from the NCPM of wind turbine SCADA data. Therefore, the novelty and contributions of this study can be summarized as follows:

- (i) The framework for the CM of wind turbines is introduced based on the adaptive multivariate control charts (AMCCs). Two AMCCs (AEWMA and MEWMA) are introduced for abnormal state alarm and anomaly location of wind turbines, respectively. An optimal design is conducted to ensure that the obtained control charts are in the optimal state.

- (ii) Multivariate regression analysis (MRA) is adopted to obtain the NCPM of wind turbine with fault-free SCADA data. Several popular algorithms in MRA, including the RF, least absolute shrinkage and selection operator (LASSO), and recursive feature elimination (RFE), are used for feature selection and regression prediction.
- (iii) Various tests on a wind turbine with normal and abnormal states are conducted. The exact anomaly time and type are known from the alarm log; thus, the performance and robustness of various control charts could be compared comprehensively.

The remainder of this paper is organized as follows. Section 2 introduces the proposed control charts. Section 3 provides the optimal design procedures. Section 4 describes feature selection and regression prediction on the SCADA data acquired from an operating wind turbine. Section 5 presents the flowchart of the AMCC-based CM method. Section 6 provides several CM examples and discusses the results. Finally, Section 7 lists the conclusions of the study.

2. AMCCs

Two AMCCs (AEWMA and MEWMA) are introduced for abnormal state alarm and anomaly location of wind turbines, respectively. The structures and procedures for these two control charts are derived in this section.

2.1. Abnormal State Alarm. Monitoring data that obey the same distribution are represented by $\{X_{i,1}, X_{i,1}, \dots, X_{i,1}\}$, where $i = 1, 2, \dots$ is the sampling time and n is the size of each sample. The mean and variance of the data are denoted by μ_0 and σ_0^2 , respectively. When the process is out of control, the mean of the data becomes $\mu_1 = \mu_0 + \delta\sigma_0$, in which δ is the shift parameter. We define the mean of the sample data as $X_i = (1/n) \sum_{j=1}^n X_{ij}$, and thus, the EWMA statistics for monitoring mean shift of the sample data could be written as follows:

$$Y_i = \lambda X_i + (1 - \lambda)Y_{i-1}, \quad (1)$$

where λ is the smoothing parameter, and $0 < \lambda < 1$. Without loss of generality, we can let $Y_{0=0}$. Lucas and Saccucci [40] pointed out that for smaller value of λ , the EWMA statics can detect a smaller mean shift faster. When takes λ a greater value, the EWMA statistic would have an accurate sensitivity to the larger mean shift. Theoretically, the EWMA control charts can be customized to detect specific shifts in the process.

However, for the actual wind turbine monitoring data, the mean shift is usually fluctuated in a certain range. The designed value of λ makes it difficult to adapt to the change in the actual mean shift. To overcome this inertia problem, the AEWMA statistic is proposed by [38]

$$Y_i = Y_{i-1} + \phi(e_i), \quad (2)$$

where $e_i = X_i - Y_i$ is the error term and $\phi(\cdot)$ represents the score function. Note that for $e_i \neq 0$, and the AEWMA statistic can be rewritten as

$$Y_i = w(e_i)X_i + (1 - w(e_i))Y_{i-1}, \quad (3)$$

where $w(e_i) = ((\phi(e_i))/e_i)$ is the equivalent smoothing parameter. Evidently, the AEWMA statistic can adaptively adjust the weight of the estimate value at the previous time (Y_{i-1}) according to the prediction error at the current time. Thus, it can balance the requirements of various mean shifts to the smoothing parameters. Yashchin [37] suggested the Huber function as the score function, and its expression is given by

$$\phi(e) = \begin{cases} e + (1 - \lambda)\gamma, & e < -\gamma; \\ \lambda e, & |e| \leq \gamma; \\ e - (1 - \lambda)\gamma, & e > \gamma; \end{cases} \quad (4)$$

where γ is the error limit. The static Y_i also obeys the same distribution with X_i and has the same mean value with X_i . When the sampling size n is large enough, the variance of Y_i can be expressed as $\sigma_Y^2 = \sigma_X^2$, where σ_X^2 is the variance of X_i , leaving us with $\sigma_X^2 = (\sigma_0^2/n)$. Therefore, the upper control limit (UCL) and lower control limit (LCL) of the AEWMA control chart could be expressed as follows:

$$\begin{aligned} \text{UCL}_{\text{AEWMA}} &= \mu_0 + k\sigma_X, \\ \text{UCL}_{\text{AEWMA}} &= \mu_0 - k\sigma_X, \end{aligned} \quad (5)$$

where k is the control limit parameter. From equations (3)–(5), it can be discovered that three parameters: λ , γ , and k should be determined to obtain the control limits of AEWMA control charts. The determination of these parameters will be discussed in the following section. It is observed that for $\gamma \rightarrow \infty$, we have $\phi(e_i) = \lambda e_i$ and $w(e_i) = \lambda$. In this case, the AEWMA statistic degenerates into the EWMA statistic, and its control limits can be expressed as

$$\begin{aligned} \text{UCL}_{\text{EWMA}} &= \mu_0 + k\sigma_X \sqrt{\frac{\lambda}{(2 - \lambda)}}, \\ \text{UCL}_{\text{EWMA}} &= \mu_0 - k\sigma_X \sqrt{\frac{\lambda}{(2 - \lambda)}}. \end{aligned} \quad (6)$$

2.2. Anomaly Location. In the abnormal state alarm of wind turbines, the data monitored by the AEWMA control chart are univariate, i.e., the output power data of the wind turbine. In addition to the early warning of an abnormal state, we also expect this method to identify the cause and location of the anomaly state. Fortunately, the SCADA system records condition parameters of the main components of wind turbines (e.g., the blade, gearbox, main bearing, and generator). Thus, we introduce the MEWMA control charts to monitor these multivariate conditional parameters, and then the components with anomaly state might be identified.

From the univariate EWMA control chart, Lowry et al. [41] proposed the MEWMA control chart, and its statistic can be expressed as

$$\mathbf{Y}_i = r\mathbf{X}_i + (1 - r)\mathbf{Y}_{i-1}, \quad (7)$$

where \mathbf{Y}_i and \mathbf{X}_i are the p dimensional multivariate data vectors. We assume that $\mathbf{Y}_0 = 0$. r denotes the smoothing parameter, leaving us with $0 < r < 1$. The MEWMA control chart will sound an alarm if the following conditions are satisfied:

$$Q_i = (\mathbf{Y}_i)^T \left(\sum \mathbf{Y}_i \right)^{-1} \mathbf{Y}_i > H, \quad (8)$$

in which H is the given control limit and $\sum \mathbf{Y}_i$ denotes the covariance matrix of the \mathbf{Y}_i . Hence, we have

$$\sum \mathbf{Y}_i = \left(\frac{r}{2-r} \right) \sum \mathbf{X}_i, \quad (9)$$

where $\sum \mathbf{X}_i$ denotes the covariance matrix of the \mathbf{X}_i . To evaluate the contribution of different dimensional data to the MEWMA statistic, following variables are defined as

$$Q_{ij} = Q_i \Big|_{\text{the } j^{\text{th}} \text{ variable is excluded}} \quad (10)$$

in which $j = 1, 2, \dots, p$. The larger variation of Q_{ij} indicates that the contribution of the j^{th} dimensional variable to the MEWMA statistic is significant. The component corresponding to this variable is more likely to be in the abnormal state. Therefore, the MEWMA control chart is adopted to monitor the multidimensional SCADA data and also to identify and locate the anomaly component of the wind turbine by analyzing the variation of Q_{ij} .

3. Optimal Design of Control Charts

The average run length (ARL), which refers to the average number of extracted samples from the beginning of the control to the emission of alarm, is used to measure the performance of various control charts. Here, the ARL_0 is used as the in-control ARL and the ARL_1 as the out-of-control ARL. Typically, ARL_0 is desired to be as large as possible and ARL_1 to be as small as possible. Hence, the designed control chart can raise an alarm on the existence of abnormal deviations as soon as possible on the premise of a lower false alarm rate. By satisfying the goal of minimizing the ARL_1 under certain ARL_0 , the optimal design procedures for the AEWMA and MEWMA control charts are presented, respectively, in the following subsections.

3.1. AEWMA. Three parameters, including λ , γ , and k , should be determined for the AEWMA control charts. Clearly, the selection of λ and γ plays a key role in the performance of

AEWMA control charts. Generally, the lower value of λ or greater value of γ should be selected for small mean shift, while the greater value of λ or lower value of γ would be favorable for detecting large mean shift. Therefore, the design of AEWMA control charts is a multiobjective optimization problem. Capizzi and Masarotto [38] utilized the simulated annealing algorithm (SAA) for the parameter optimization of AEWMA control charts. However, the requirement for the initial value of SAA is relatively high. Once the initial value significantly deviates from the optimal value, it is difficult to converge to the optimal value. To improve the convergence speed of SAA, Shu [39] proposed a “two-step method.” First, the AEWMA control chart is treated as a conventional EWMA control chart, and the optimal value of λ is obtained under certain ARL_0 using SAA. Then, on the premise of given value of λ , the value of γ is optimized. Figure 1 shows the flowchart for the optimal design of the AEWMA control chart. Detailed procedures are described as follows:

- (1) Sample size n and in-control ARL ARL_0 are selected. Two mean shift values δ_1 and δ_2 are given to ensure that $\delta_1 < \delta_2$.
- (2) Typically, the range of parameter optimization is selected as $\lambda < 0.8$, $3 < \lambda < 7$, and $k < 10$.
- (3) By setting $\gamma \rightarrow \infty$ (e.g., $\gamma = 1000$), the AEWMA control chart is degenerated to the EWMA control chart. The parameter of the EWMA under the shift δ_1 is denoted by $\theta_1 = (\lambda, k)$. The optimal θ_1 should satisfy the following optimization problem:

$$\begin{cases} \min & ARL_0(\delta_1, \theta_1) \\ \text{s.t.} & ARL_1(0, \theta_1) = ARL_0. \end{cases} \quad (11)$$

- (4) A small positive number α (taken as $\alpha = 0.05$ in this study) is selected to ensure that the control chart will not lose too much accuracy after the introduction of γ . Based on the optimal parameter λ obtained in step (3), the optimal parameter $\theta_2 = (\gamma, k)$ of the AEWMA control charts with mean shift δ_2 could be obtained by solving the following optimization problem:

$$\begin{cases} \min & ARL_1(\delta_2, \theta_2) \\ \text{s.t.} & ARL_1(0, \theta_2) = ARL_0, ARL_1(\delta_2, \theta_1) \leq (1 + \alpha)ARL_1(\delta_1, \theta_1). \end{cases} \quad (12)$$

In the above steps, the calculation of ARL can be obtained using the Monte Carlo sampling method.

3.2. MEWMA. The MEWMA control chart has two parameters: r and H . Similar to the AEWMA control chart, small values of the smoothing parameter r should be selected

for small mean shifts, while large r values have advantages in detecting large mean shift. Runger and Prabhu [42] proposed a Markov chain algorithm (MCA) for designing a MEWMA control chart. For $i, j = 0, 1, 2, \dots, m_1$, the transition probability from state i to state j is denoted by $p(i, j)$, and its definition is given by

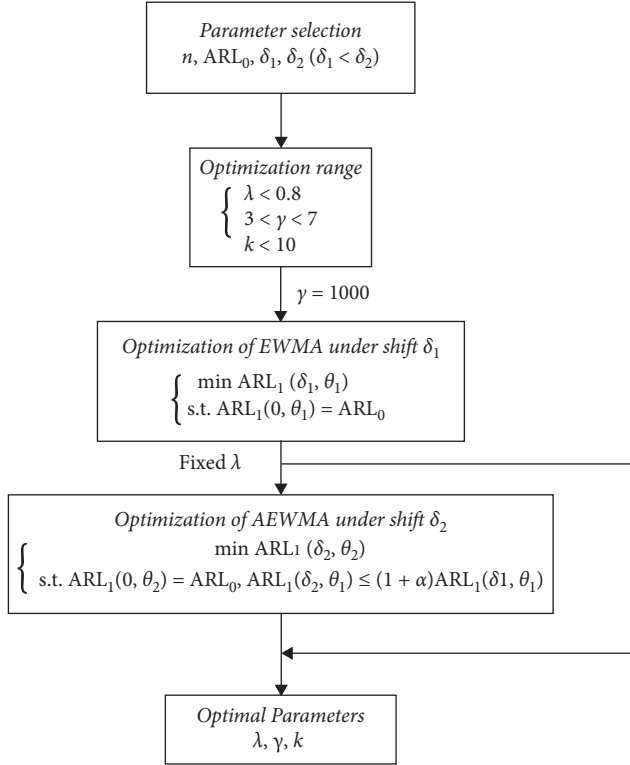


FIGURE 1: Flowchart for the optimal design of AEWMA control charts.

$$p(i, j) = \begin{cases} P\left\{\frac{(j-0.5)^2 g_1^2}{r^2} < \chi^2(p, c) < \frac{(j+0.5)^2 g_1^2}{r^2}\right\}, & \text{if } j \neq 0; \\ P\left\{\chi^2(p, c) < \frac{(0.5)^2 g_1^2}{r^2}\right\}, & \text{if } j = 0; \end{cases} \quad (13)$$

where $\chi^2(p, c)$ is the noncentral chi-square random variable, p is the number of dimensions, $c = (((1-r)ig_1)/r)^2$ is noncentral parameter, and $g_1 = (2/(2m_1 + 1))\sqrt{(r/(r-2))H}$. Based on the transition probability, the $(m_1 + 1)$ dimension transition matrix \mathbf{P}_0 could be constructed. Thus, the ARL_0 of MEWMA control charts could be calculated by

$$ARL_0 = \lim_{m_1 \rightarrow \infty} \mathbf{s}^T (\mathbf{I} - \mathbf{P})^{-1} \mathbf{1}, \quad (14)$$

where \mathbf{s} is the initial probability vector, \mathbf{I} is the unit vector, and $\mathbf{1}$ is a vector with all of its elements equal to 1. Similarly, the ARL_1 of the MEWMA control chart could also be obtained.

Based on the MCA [42], we use the partition method to obtain the optimal parameters of MEWMA control charts. The partition method generates a combination of a smoothing parameter r and a control limit H , satisfying a

given ARL_0 , and finding the optimal smoothing parameter. Figure 2 presents the flowchart for the optimal design of MEWMA control charts. Detailed procedures are described as follows:

- (1) For a fixed smoothing parameter r , the method inspects the middle point of a lower control limit H_{low} and a upper control limit H_{up} such that $ARL_{H_{low}} \leq ARL_0$ and $ARL_{H_{up}} \geq ARL_0$.
- (2) Once H_{mid} , the middle point of two control limits is obtained, and ARL can be calculated by using the MCA. If the difference between ARL_0 and the newly computed ARL is less than a small number (i.e., $\epsilon < 10^{-3}$), the smoothing parameter r and the control limit H_{mid} is a pair that can satisfy the given ARL_0 . Otherwise, keep following the previous procedures until a sought pair is found.
- (3) If this task is carried out until the method covers a whole range of smoothing parameter ($0 < r \leq 1$), a number of combinations of r and H can be obtained. With the combinations obtained, ARL_1 values can be calculated for a given shift δ . Then, the smoothing parameter r for which ARL_1 is the smallest can be identified.

4. MRA on Fault-Free SCADA Data

In previous sections, both control charts of AEWMA and MEWMA have been introduced for the abnormal state alarm and anomaly location of wind turbines. The optimal design procedures for these control charts have been presented. The residuals monitored by these control charts are yielded by the departure of real-time SCADA data from the predictions of NCPM. In this section, we utilize the MRA to construct the NCPM of wind turbines with fault-free SCADA data. Several popular algorithms in MRA, including the RF, LASSO, and RFE, are used for feature selection and regression prediction.

4.1. Data Descriptions. This study aims to monitor and diagnose doubly fed wind turbines with rated power of 2 MW. Typically, the SCADA data of the unit include output power, speed, torque, temperature, and pitch angle. The data record interval is 10 min. To correctly establish the NCPM of wind turbines, the anomaly data should be avoided as much as possible. By reading the record table of the SCADA system, it was found that no anomaly was reported in the time period from 12/26/2013 to 2/12/2014. The wind turbine unit was built and connected to the grid in early 2012. In this time period, the unit has passed the initial running stage and is in the stage of normal power generation. Therefore, the data segment is ideal for MRA to construct the NCPM of wind turbines. There are 45 variables recorded by the SCADA system. After excluding the lost data points and data

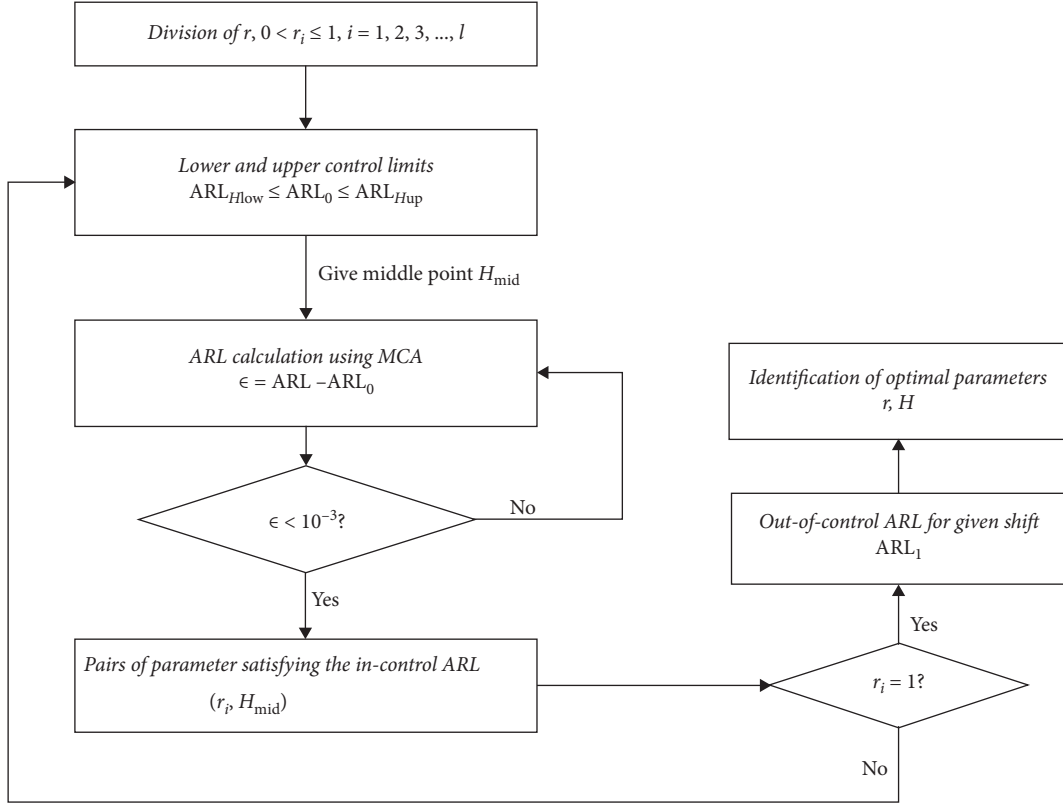


FIGURE 2: Flowchart for the optimal design of MEWMA control charts.

points during the maintenance downtime, the total amount of data is 6135 points.

4.2. Feature Selection and Regression Prediction. At the beginning of MRA on the fault-free SCADA data, to minimize the problem of model deviation due to the lack of important variables, we usually select as many argument variables as possible. In this study, we select the output power as the response variable and the remaining 44 variables as argument variables. However, in the process of actual modeling, it is necessary to select a variable subset (feature selection) which has the best ability to explain the response variable to improve the regression and prediction accuracy of the NCPM ([43, 44]). Before feature selection, the raw SCADA data should be standardized as follows:

$$Z_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}, \quad (15)$$

where x_{ij} is the i^{th} sample point of the j^{th} variable. $\max(x_j)$ and $\min(x_j)$ denote the maximum and minimum values of the j^{th} variable, respectively. Three metrics, including root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE), are defined to measure the goodness of fit of NCPM using MRA. They are expressed by

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \\ \text{MAPE} &= \left(\frac{1}{n} \right) \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \\ \text{MAE} &= \left(\frac{1}{n} \right) \sum_{i=1}^n |y_i - \hat{y}_i|, \end{aligned} \quad (16)$$

where n is the sample size and \hat{y}_i and y_i are predicted and actual values of output power, respectively.

The RF, LASSO, and RFE, which are popular algorithms in MRA, are utilized for feature selection and regression prediction. Basic ideas and characteristics for these algorithms are introduced.

The RF is an integrated machine learning method [45]. It employs random resampling technology bootstrap and node random splitting technology to construct multiple decision trees, with the final classification results are obtained by voting. The RF has the ability to analyze the classification characteristics of complex interactions. It has a fine robustness for noise data and a faster learning speed. Its variable importance measure can be used as a feature selection tool for high dimensional data. The core algorithm uses the RF package in R software, in which the parameter takes the value of \sqrt{n} recommended by Breiman [45] (n is the number of features of the training data set). The number of trees is set to be $n_{\text{tree}} = 500$.

LASSO [46] is a linear model for estimating sparse parameters, especially for reducing the number of parameters. This method uses the l_1 norm to compress the coefficient of the model and directly makes the values of insignificant model parameters smaller (including zero). This gives the LASSO the advantage of feature selection and ridge regression. Without changing the accuracy of the model test set, the dimension of the feature could be effectively reduced by using the LASSO regression model. The core algorithm adopts the LARS package in R software, and the cross validation is utilized to select the penalty parameter that controls the sparse parameter estimation.

The main idea of RFE [47] is to build the model iteratively and then select the best (or worst) feature (which can be selected according to the value of coefficients). The iteration process on the remaining feature will be conducted until all the features have been traversed. The stability of RFE mostly depends on the type of iteration model.

In this study, the core algorithm is implemented through the CARET package in R software. After a series of tests, the decision tree model (treebagFuncs) is selected as the iteration model.

For the fault-free wind power SCADA data, the above three algorithms are used for regression prediction and feature selection. Table 1 displays the regression accuracy in the metrics of RMSE, MAPE, and MAE. After comparisons have been conducted, it can be observed that the RF has the best accuracy in regression. Thus, the feature selection based on the RF is carried out, and the top 15 features are shown in Table 2. According to the common sense, the parameters closely related to the output power of the wind turbine are wind speed, generator speed, torque, rotor speed, etc. These features have been reflected in the feature ranking of RF. In addition, the rankings of generator phase current and phase voltage, as well as the parameters of several temperature measuring points (including the gearbox, bearing, generator, and even nacelle), were relatively higher. These parameters are not easy to judge and select directly through common sense. Figure 3 also presents the comparison between the source SCADA data and regression prediction results. For the sake of simplicity, only four argument variables (i.e., the rotation torque, generator current, average wind speed in 10 min, and generator speed) on the response variable (output power of the wind turbine unit) are illustrated in the figure.

5. Wind Turbine CM System Based on AMCCs

Some key contents, including the structures of AMCCs, optimal design procedures of these control charts, and construction of NCPM with fault-free SCADA data, have been introduced in the previous sections, respectively.

How to implement these core algorithms needs to be explained for engineering applications. Figure 4 presents the flowchart for the wind turbine CM system based on AMCCs. The entire process could be summarized as follows:

- (1) MRA is utilized to construct the NCPM of wind turbines with fault-free SCADA data. In this study,

TABLE 1: Comparisons of regression accuracy using various algorithms.

	RMSE	MAPE	MAE
RF	0.002208326	0.01083009	0.001168414
LASSO	0.006998046	0.05116372	0.005377479
RFE	0.005929714	0.03291574	0.004997808

TABLE 2: Feature selection result based on RF.

Feature no.	Feature ranking
1	Rotation torque
2	Generator phase A current
3	Average wind speed in 10 min
4	Generator speed
5	Rotor speed
6	Blade yaw angle
7	Generator temperature
8	Gearbox bearing temperature
9	Nacelle angle
10	Generator phase A voltage
11	Nacelle revolution
12	Gearbox temperature
13	Ambient temperature
14	Nacelle temperature
15	Bearing temperature

the RF shows better performance in feature selection and regression prediction.

- (2) Time-variable residuals of output power are produced by measuring the difference between the real-time SCADA data and the predictions of NCPM.
- (3) For the goal of minimum out-of-control ARL (see Figure 1), the optimal AEWMA control chart is constructed to monitor the output power residuals. Steps (2) and (3) will be continued until the abnormal state is alarmed.
- (4) The optimal MEWMA control chart is established (see Figure 2) to model conditional parameters of main components, which are acquired from real-time SCADA data. The component under an anomaly state could then be located.

In the following, the effectiveness of the proposed CM method is shown by several examples. The performance and robustness of various control charts are compared in detail.

6. CM Examples

Based on the feature selection and regression prediction results, CM practice on the wind turbine unit is carried out. During the period from 12/1/2015 to 6/1/2016, there were three anomalies, namely, the generator brush worn, gearbox running hot in low generator stage, and shaft bearing overtemperature. The specific time of alarm log is shown in Table 3. For each anomaly, the number of monitored data points is 500, and the exact anomaly data point is also given in the table for comparisons.

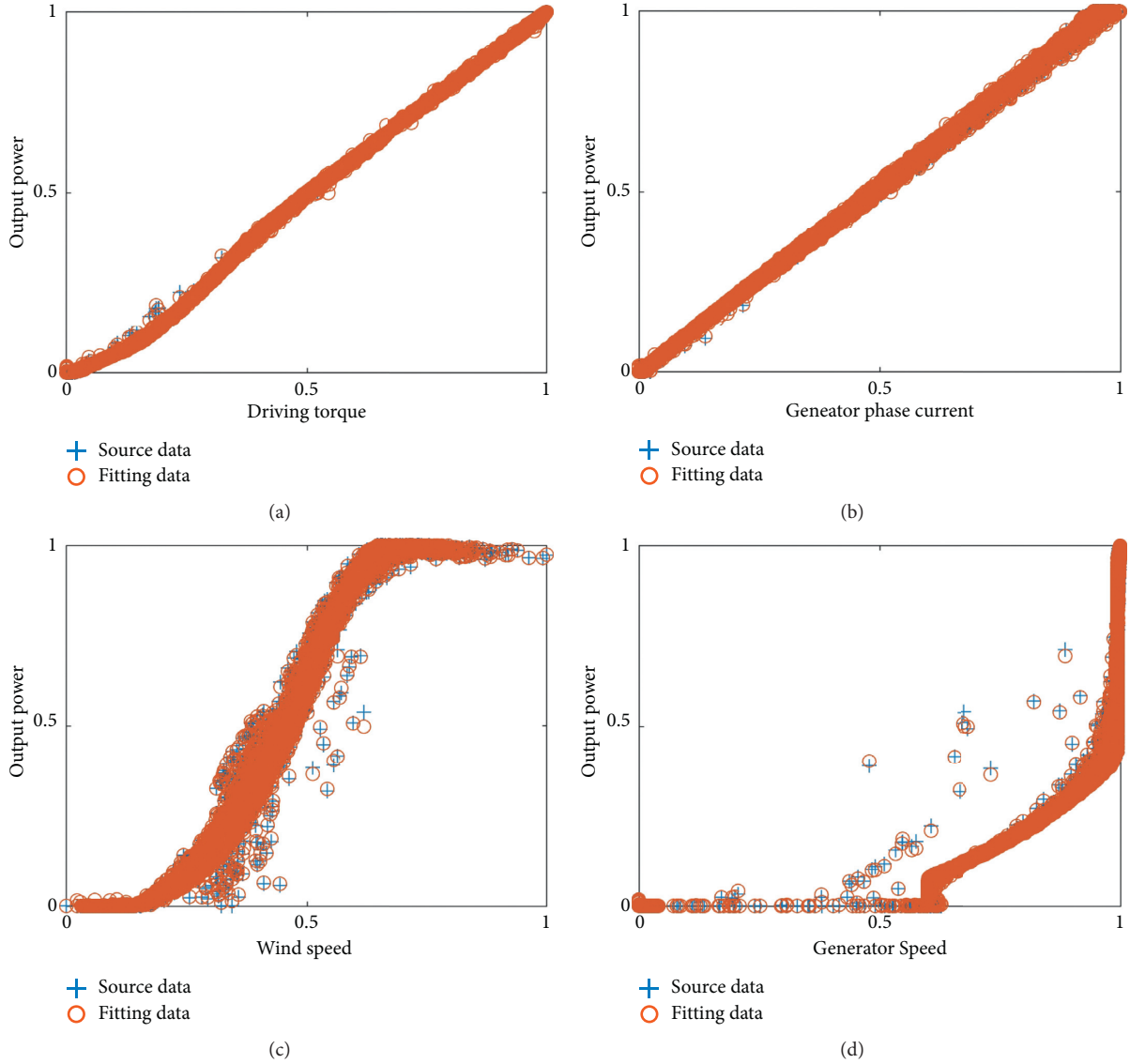


FIGURE 3: Comparison between the source SCADA data and regression prediction results: (a) rotation torque, (b) generator current, (c) average wind speed in 10 min, and (d) generator speed.

6.1. Abnormal State Alarm. By using the NCPM model obtained in the previous section, the output power of the unit before and after the fault (500 data points in Table 3) is predicted, and then the residual is obtained by measuring the difference in the actual output power. The mean and variance of the predicted residuals for three fault data are all less than 0.05 and 0.08. Given $ARL_0 = 500$ and shift range (0.4–4), the optimal parameters of AEWMA control chart are then obtained, as shown in Table 4. For comparison, the parameters of the optimal EWMA control charts corresponding to different shifts (0.4, 2, 4) are also given in the table.

As mentioned before, the out-of-control ARL is an important index to evaluate the performance of control charts. Figure 5 shows the variation of ARL_1 with mean shift in the range of 0–4 for the designed control charts in Table 4. Evidently, when the shift is zero, the out-of-control ARL is equal to the in-control ARL , i.e., $ARL_1 = ARL_0$. With an

increase in a shift, the value of ARL_1 gradually decreases. Under small shifts ($\delta < 1.5$), the value of ARL_1 for the AEWMA is lower than that of the EWMA control charts, especially for the EWMA control chart with larger smoothing parameters (EWMA-2 and EWMA-3). This means that the AEWMA behaves more sensitively and could give warnings of abnormal states earlier than the EWMA control charts.

When the shift becomes large enough ($\delta > 2$), the difference in ARL_1 between the AEWMA and EWMA control charts is not significant, indicating that under large shift, the AEWMA can still maintain a performance comparable to the EWMA control charts. This is consistent with the theoretical expectation of the AEWMA.

The AEWMA control charts are established for the output power residuals with anomaly A, B, and C, as shown in Figure 6. Figure 7 also presents the residuals monitored by various EWMA control charts for comparisons. It can be

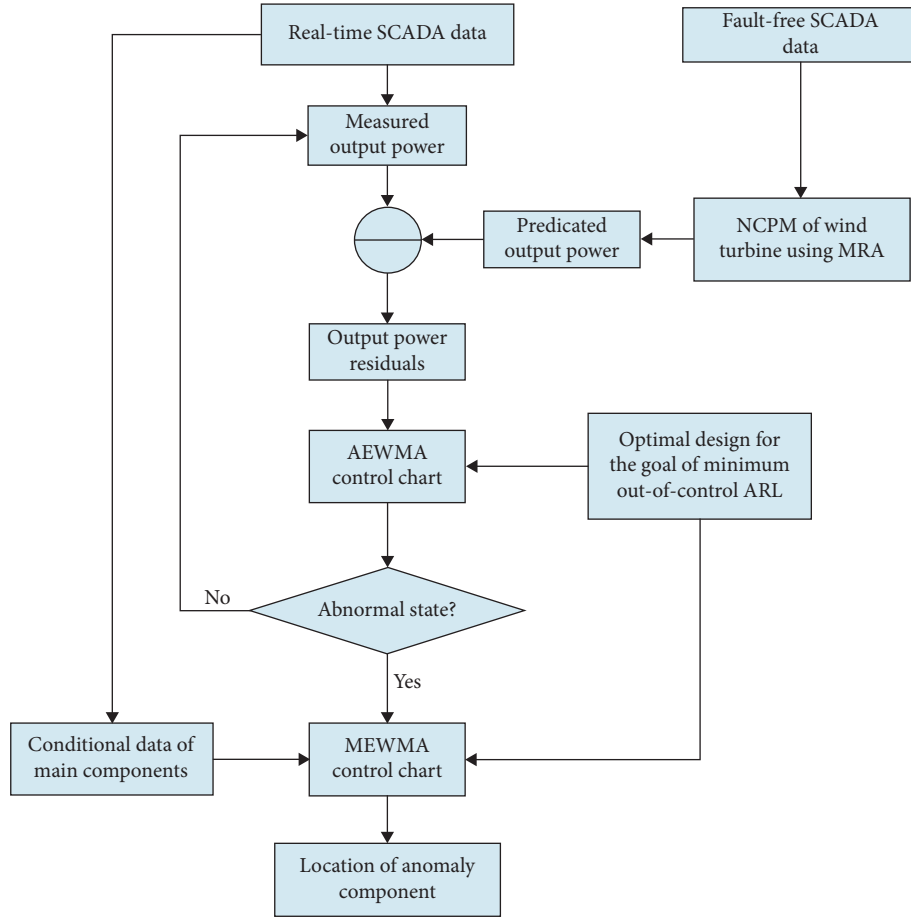


FIGURE 4: Flowchart for the wind turbine CM system based on AMCCs.

TABLE 3: Anomaly information for the wind turbine unit during the period from 12/1/2015 to 6/1/2016.

Anomaly no.	Anomaly description	Monitored period	Number of data points	Time of alarm log
A	Generator brush worn	12/3/2015 19:10–12/7/2015 6:30	500	12/6/2015 20:30 (alarm log: point no. 440)
B	Gearbox running hot in low generator stage	2/1/2016 00:00–2/4/2016 11:10	500	2/3/2016 22:10 (alarm log: point no. 421)
C	Shaft bearing overtemperature	5/28/2016 15:50–6/1/2016 3:10	500	5/31/2016 15:20 (alarm log: point no. 429)

TABLE 4: Optimization results of AEWMA and EWMA control charts for abnormal state alarm.

	ARL ₀	Mean shifts	λ	γ	k
AEWMA	500	$\delta_1 = 0.4, \delta_2 = 4$	0.1	5.4750	1.3531
EWMA-1	500	$\delta = 0.4$	0.1	—	6.8110
EWMA-2	500	$\delta = 2$	0.2	—	5.9430
EWMA-3	500	$\delta = 4$	0.4	—	6.9531

observed from the figures that the AEWMA control chart can effectively identify the abnormal state caused by the anomaly. Compared with the alarm log of the SCADA system, the AEWMA control chart can send the alarm in time. For anomaly A (see Figure 6(a)), one can see that

the AEWMA alarm time is $(440 - 407) * 10 = 330$ min (about 5.5 h) ahead. For anomaly B and C (see Figures 6(b) and 6(c)), the time of advance is about $(421 - 398) * 10 = 230$ min (about 3.8 h) and $(429 - 407) * 10 = 220$ min (about 3.7 h), respectively. Thus, one can say that the alarm time of AEWMA control charts could be several hours ahead of the SCADA system, and in this study, the maximum promptness appears in anomaly A (about 5.5 h).

Compared with the AEWMA control chart, the EWMA control charts behave less sensitively to fault and have poor robustness. For the EWMA-3 of anomaly A (see Figure 7(a)), and EWMA-1, EWMA-2, and EWMA-3 of anomaly B (see Figure 7(b)), the abnormal state is not identified during the monitoring period. Although the faults

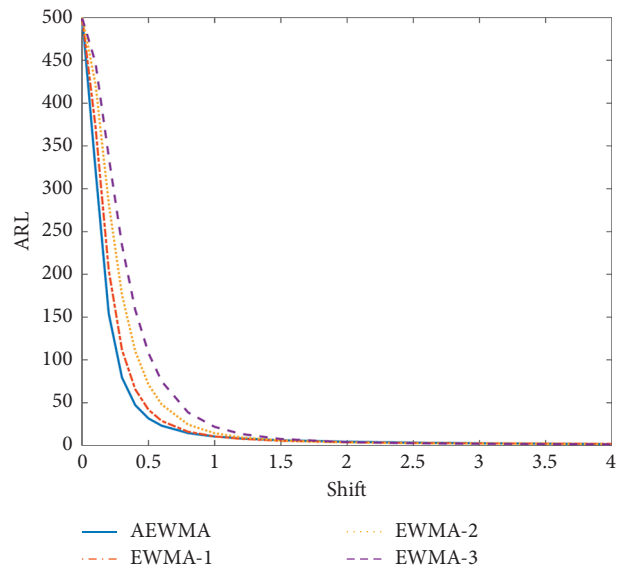
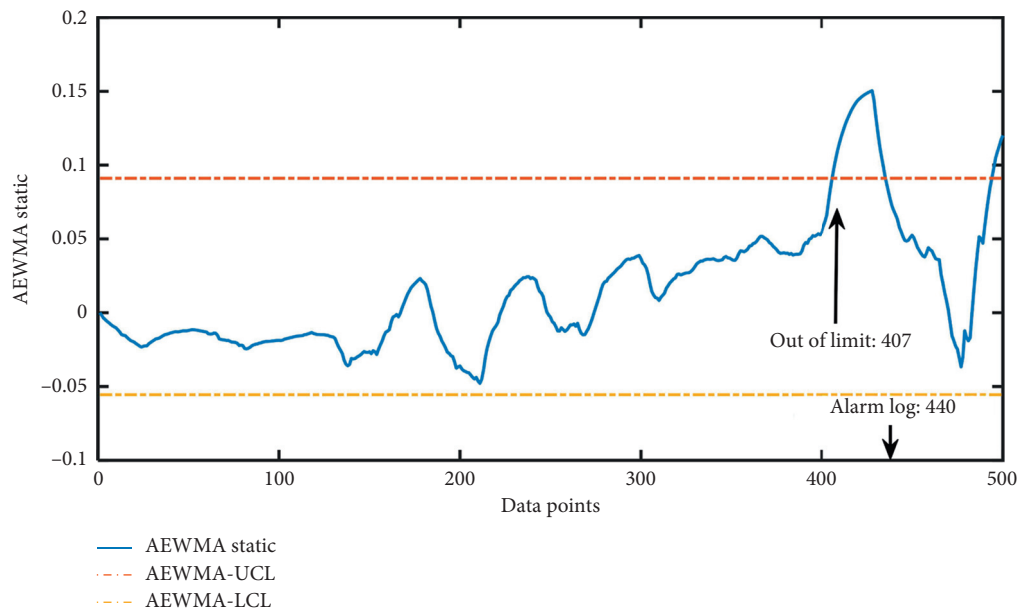


FIGURE 5: Variation of ARL_1 with mean shift in the range 0–4 for the designed control charts.



(a)
FIGURE 6: Continued.

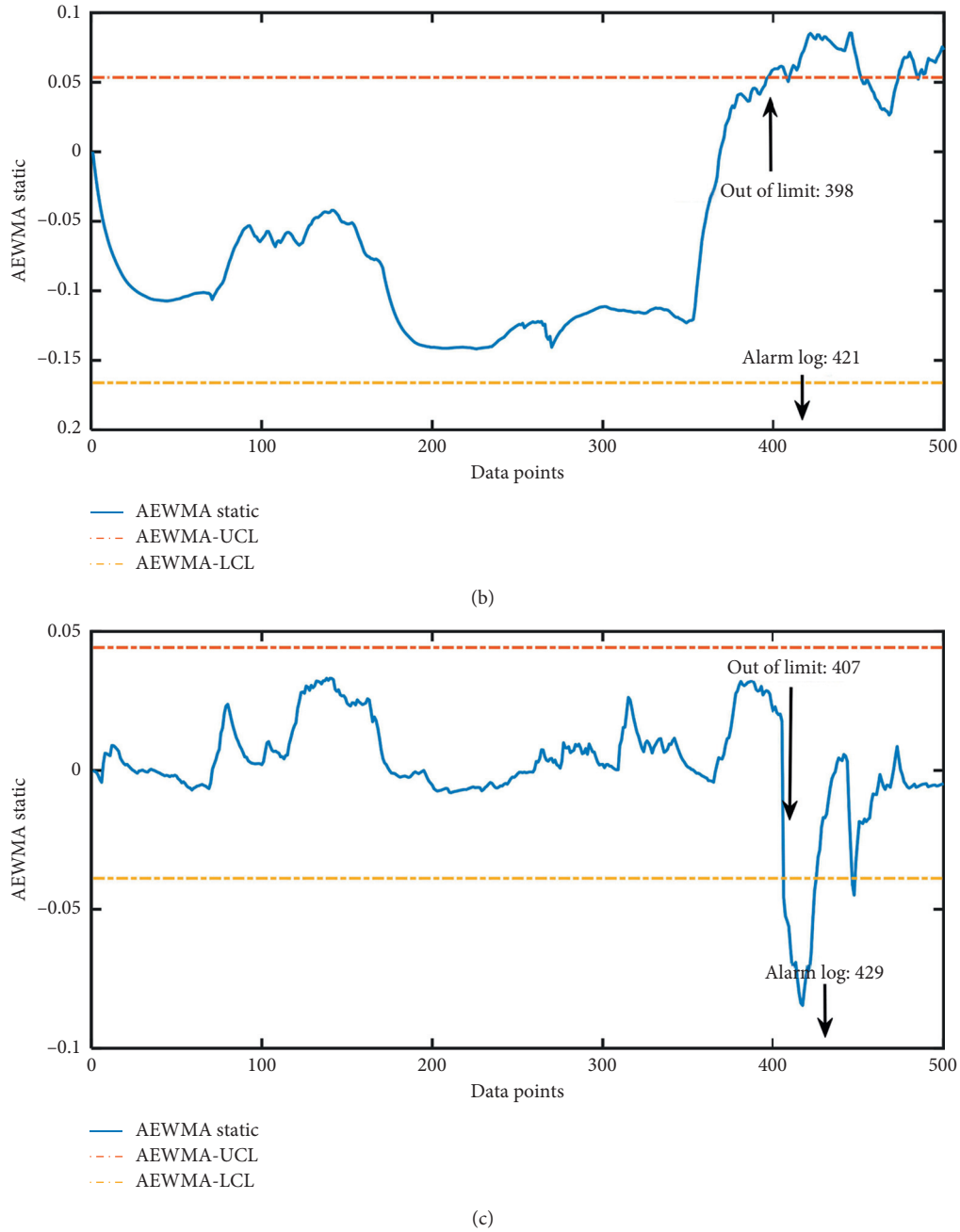


FIGURE 6: Residuals monitored by AEWMA control chart with (a) anomaly A, (b) anomaly B, and (c) anomaly C.

are signaled earlier than the SCADA system, the alarm time still lags behind the AEWMA control chart among the rest of EWMA control charts. For anomaly A, the EWMA-2 control chart (see Figure 7(a)) sends the earliest alarm, about $(440 - 414) \times 10 = 206$ min (about 4.3 h), which still lags behind the AEWMA control chart for $(5.5 - 4.3) = 1.2$ h. For anomaly C, the EWMA-2 control chart (see Figure 7(c)) has the best performance and its promptness is $(429 - 410) \times 10 = 190$ min (about 3.2 h), which still lags behind the AEWMA control chart for $(3.7 - 3.2) = 0.5$ h.

From the above CM examples, one can say that compared with the EWMA control charts, the AEWMA control

chart behaves more sensitively to the abnormal state. Thus, it can effectively identify the abnormal state and has better robustness. This is of great application value to the CM of practical wind turbine units.

6.2. Anomaly Location. In the previous section, it is demonstrated that the AEWMA control chart can effectively identify the abnormal state. However, for complex electromechanical systems (i.e., the wind turbine), in addition to the early warning of abnormal state, it is also expected to identify the anomaly component, which is called the

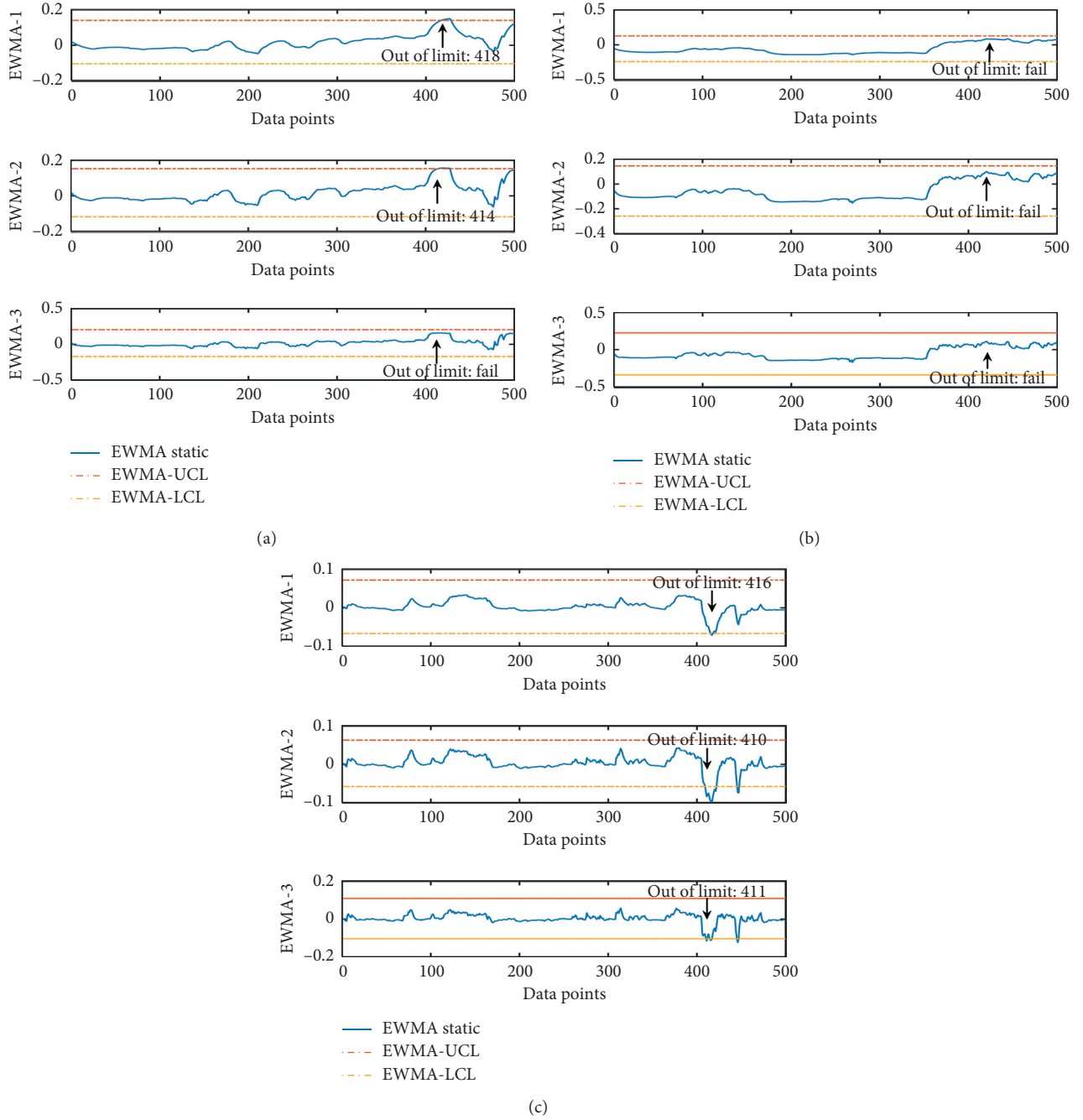


FIGURE 7: Residuals monitored by various EWMA control charts with (a) anomaly A, (b) anomaly B, and (c) anomaly C.

anomaly location. From the important features in Table 2, in addition to environmental features (such as the average wind speed and ambient temperature), there are also features characterizing the working conditions of main components, including the generator speed, gearbox temperature, and blade yaw angle. This section uses the MEWMA control chart to model multidimensional data, studies the influence of various features on the monitoring statistics, and realizes the effective location and identification of the anomaly.

The input parameter for MEWMA control chart should be determined by $\delta = \mu^T \sum_{i=1}^p \mu_i$, where μ represents the mean

vector of multidimensional data. After calculation, it is discovered that the values of δ for three samples are all lower than 4.5. Thus, by setting $\delta = 4.5$, data dimension $p = 15$, and $ARL_0 = 500$, the optimal parameters of MEWMA control chart can be obtained as $r = 0.48$ and $H = 29.65$. For the 15-dimensional monitoring data containing the anomaly A, B, and C, the MEWMA statistic could be evaluated. In Section 2.2, we have defined Q_{ij} (see equation (10)) to evaluate the contribution of various dimensional data to the MEWMA statistic.

When the different dimension variables are excluded, the changes in Q_{ij} are shown in Figure 8, respectively. As

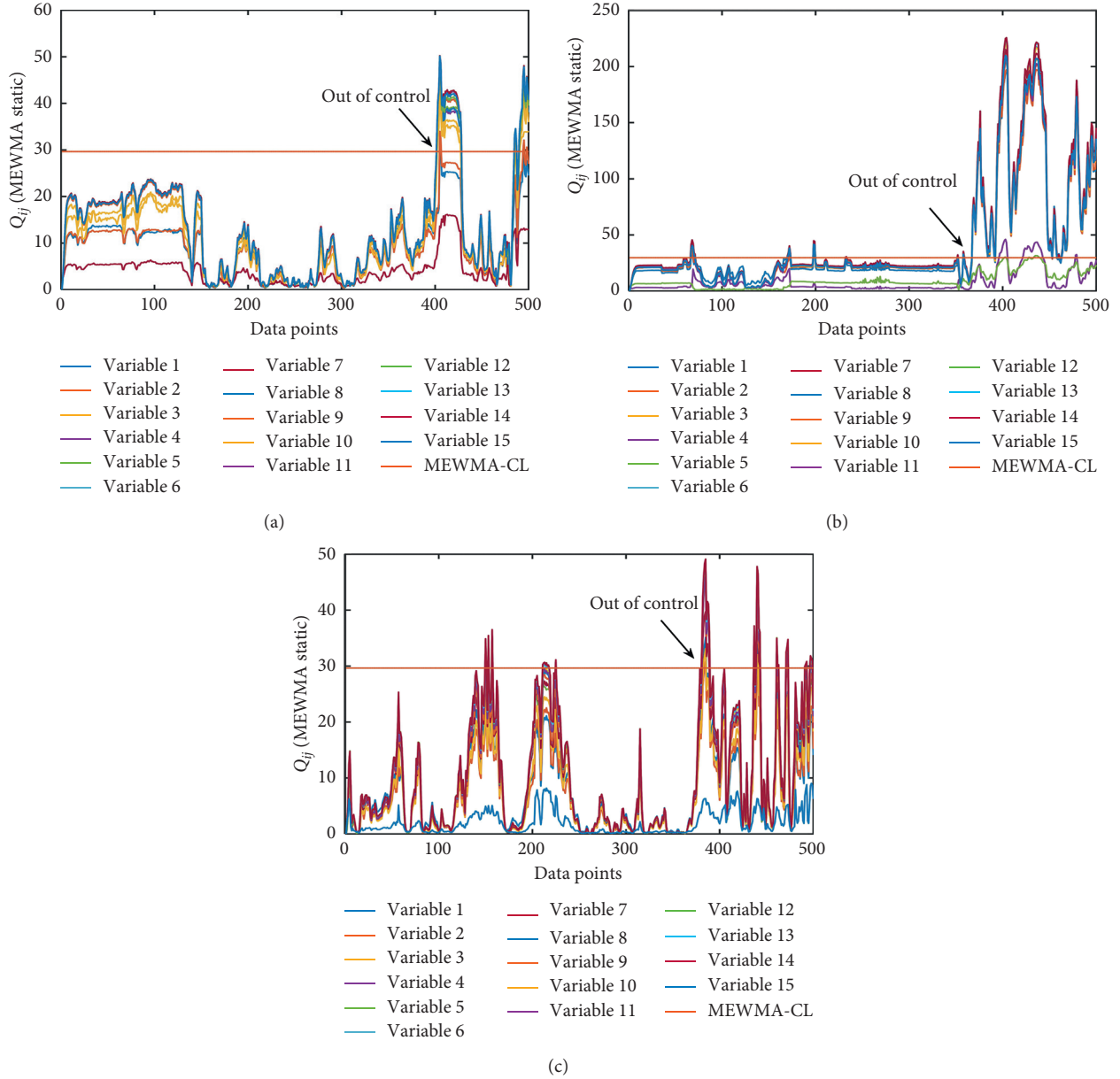


FIGURE 8: Multidimensional SCADA data monitored by the MEWMA control chart with (a) anomaly A, (b) anomaly B, and (c) anomaly C.

can be identified from the figures, some specific dimension variables contribute greatly to the MEWMA statistic, while others seem to have little influence on the statistic. Nevertheless, from Figure 8, it is not easy to directly identify which dimension variables have remarkable contribution on Q_{ij} . We can define the number (or frequencies) of MEWMA statistic beyond the control limit as an metric. Then, the out-of-limit number (OLN) variation of various dimensional data is shown in Figure 9, respectively. Some observations could be summarized as follows:

- (1) For anomaly A (see Figure 9(a)), the variations of OLN after the removal of variable 7, 1, and 2 are greater than the removal of other variables. From Table 2, one can see that the variable 7 represents

“generator temperature,” the variable 1 “rotation torque,” and variable 2 “generator phase A current.” Consequently, it is estimated that the generator is more likely to have an anomaly. The alarm log of the SCADA system confirms that anomaly A does appear in the generator, which is described as the generator brush worn in Table 3.

- (2) For anomaly B, as shown in Figure 9(b), the maximum OLN variation occurs at variable 12 (“gearbox temperature”), indicating that the gearbox might be in an abnormal state. This is also consistent with the anomaly description “gearbox running hot in low generator stage” (see Table 3).
- (3) When anomaly C is considered (see Figure 9(c)), one can find that the maximum OLN variation appears at

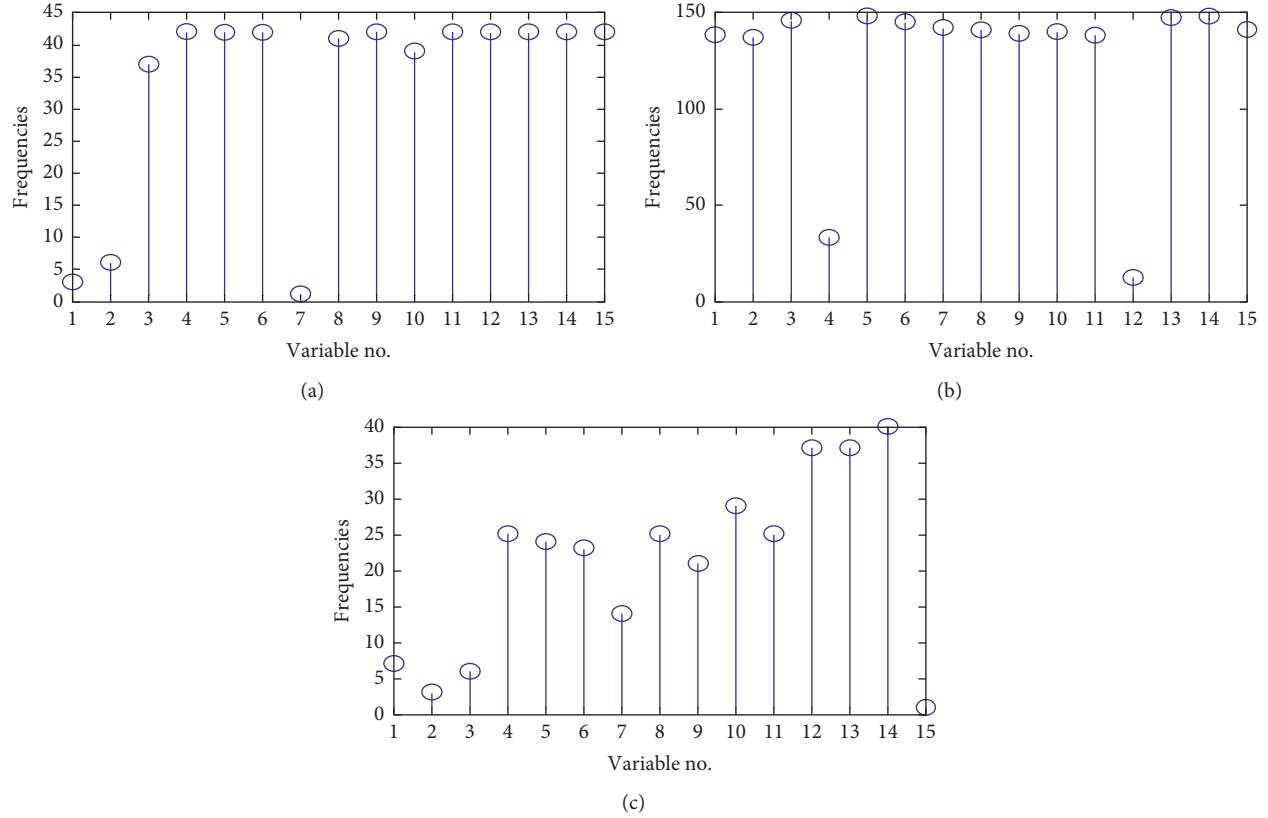


FIGURE 9: Out-of-limit number (OLN) variation of various dimensional data with (a) anomaly A, (b) anomaly B, and (c) anomaly C.

variable 15 (“bearing temperature”), and the main bearing is more likely to be in an abnormal state. This agrees well with the description of anomaly C “shaft bearing overtemperature” (see Table 3).

Through the accurate location of the three different anomalies, one can see that the MEWMA control chart combined with the OLN index can effectively locate and identify the abnormal component.

7. Conclusions

A novel CM method of wind turbines is introduced based on AMCCs and SCADA data. Two AMCCs (AEWMA and MEWMA) are proposed for abnormal state alarm and anomaly location of wind turbines, respectively. Optimization procedures for these control charts are implemented with the goal of minimum out-of-control ARL. MRA is utilized to obtain the NCPM of wind turbine with fault-free SCADA data. After conducting comparisons of the regression accuracy of several popular algorithms in the MRA, the RF is used for feature selection and regression prediction. Various tests on a wind turbine with normal and abnormal states are conducted. The performance and robustness of various control charts are compared comprehensively. Compared with the EWMA control charts, the AEWMA control chart behaves more sensitively to the abnormal state and thus has a more effective anomaly identification ability and better robustness. By accurately locating three different

anomalies, it is demonstrated that the MEWMA control chart combined with the OLN index can effectively locate and identify the abnormal component.

Data Availability

The wind turbine data used to support the findings of this study were supplied by a wind power plant under license and so cannot be made freely available. Requests for access to these data should be made to (Qinkai Han, Email: han-qinkai@hotmail.com).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Science Foundation of China under grant no. 11872222 and the State Key Laboratory of Tribology under grant no. SKLT2019B09. Tao Hu’s work was partly supported by the Beijing Talent Foundation Outstanding Young Individual Project, the Support Project of High-Level Teachers in Beijing Municipal Universities in the Period of 13th Five-Year Plan (grant CIT & TCD 201804078), Academy for Multidisciplinary Studies of Capital Normal University.

References

- [1] F. P. García Márquez, A. M. Tobias, J. M. Pinar Pérez, and M. Papaelias, "Condition monitoring of wind turbines: techniques and methods," *Renewable Energy*, vol. 46, pp. 169–178, 2012.
- [2] W. Y. Liu, B. P. Tang, J. G. Han, X. N. Lu, N. N. Hu, and Z. Z. He, "The structure healthy condition monitoring and fault diagnosis methods in wind turbines: a review," *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 466–472, 2015.
- [3] T. Wang, Q. Han, F. Chu, and Z. Feng, "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: a review," *Mechanical Systems and Signal Processing*, vol. 126, pp. 662–685, 2019.
- [4] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis-part II: signals and signal processing methods," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6546–6557, 2015.
- [5] Q. Han, Z. Ding, X. Xu, T. Wang, and F. Chu, "Stator current model for detecting rolling bearing faults in induction motors using magnetic equivalent circuits," *Mechanical Systems and Signal Processing*, vol. 131, Article ID 554575, 2019.
- [6] Q. Han, T. Wang, Z. Ding, X. Xu, and F. Chu, "Magnetic equivalent modeling of stator currents for localized fault detection of planetary gearboxes coupled to electric motors," *IEEE Transactions on Industrial Electronics*, 2020.
- [7] J. Wang, T. Niu, H. Lu, Z. Guo, W. Yang, and P. Du, "An analysis-forecast system for uncertainty modeling of wind speed: a case study of large-scale wind farms," *Applied Energy*, vol. 211, pp. 492–512, 2018.
- [8] H. Z. Wang, G. Q. Li, G. B. Wang, J. C. Peng, H. Jiang, and Y. T. Liu, "Deep learning based ensemble approach for probabilistic wind power forecasting," *Applied Energy*, vol. 188, pp. 56–70, 2017.
- [9] Q. Han, Z. Hao, T. Hu, and F. Chu, "Non-parametric models for joint probabilistic distributions of wind speed and direction data," *Renewable Energy*, vol. 126, pp. 1032–1042, 2018.
- [10] F. Castellani, D. Astolfi, P. Sdringola, S. Proietti, and L. Terzi, "Analyzing wind turbine directional behavior: SCADA data mining techniques for efficiency and power assessment," *Applied Energy*, vol. 185, pp. 1076–1086, 2017.
- [11] Q. Han, S. Ma, T. Wang, and F. Chu, "Kernel density estimation model for wind speed probability distribution with applicability to wind energy assessment in China," *Renewable and Sustainable Energy Reviews*, vol. 115, Article ID 109387, 2019.
- [12] D. Astolfi, F. Castellani, A. Garinei, and L. Terzi, "Data mining techniques for performance analysis of onshore wind farms," *Applied Energy*, vol. 148, pp. 220–233, 2015.
- [13] Q. Han, Z. Ding, Z. Qin, T. Wang, X. Xu, and F. Chu, "A triboelectric rolling ball bearing with self-powering and self-sensing capabilities," *Nano Energy*, vol. 67, Article ID 104277, 2020.
- [14] A. Kusiak and A. Verma, "A data-mining approach to monitoring wind turbines," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 150–157, 2012.
- [15] A. Kusiak and Z. Zhang, "Analysis of wind turbine vibrations based on SCADA data," *Journal of Solar Energy Engineering*, vol. 132, pp. 031008-1–031008-12, 2010.
- [16] A. Verma and A. Kusiak, "Fault monitoring of wind turbine generator brushes: a data-mining approach," *Journal of Solar Energy Engineering*, vol. 134, pp. 021001-1–021001-9, 2012.
- [17] S. Gill, B. Stephen, and S. Galloway, "Wind turbine condition assessment through power curve copula modeling," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 94–101, 2012.
- [18] M. Schlechtingen, I. F. Santos, and S. Achiche, "Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: system description," *Applied Soft Computing*, vol. 13, no. 1, pp. 259–270, 2013.
- [19] M. Schlechtingen and I. F. Santos, "Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 2: application examples," *Applied Soft Computing*, vol. 14, pp. 447–460, 2014.
- [20] P. Sun, J. Li, C. Wang, and X. Lei, "A generalized model for wind turbine anomaly identification based on SCADA data," *Applied Energy*, vol. 168, pp. 550–567, 2016.
- [21] S. Wang, Y. Huang, L. Li, and C. Liu, "Wind turbines abnormality detection through analysis of wind farm power curves," *Measurement*, vol. 93, pp. 178–188, 2016.
- [22] Z. Song, Z. Zhang, Y. Jiang, and J. Zhu, "Wind turbine health state monitoring based on a Bayesian data-driven approach," *Renewable Energy*, vol. 125, pp. 172–181, 2018.
- [23] W. Yang, C. Liu, and D. Jiang, "An unsupervised spatio-temporal graphical modeling approach for wind turbine condition monitoring," *Renewable Energy*, vol. 127, pp. 230–241, 2018.
- [24] P. Bangalore and M. Patriksson, "Analysis of SCADA data for early fault detection, with application to the maintenance management of wind turbines," *Renewable Energy*, vol. 115, pp. 521–532, 2018.
- [25] D. Montgomery, *Introduction to Statistical Quality Control*, John Wiley & Sons, Hoboken, NJ, USA, 2007.
- [26] A. Marvuglia and A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques," *Applied Energy*, vol. 98, pp. 574–583, 2012.
- [27] N. Yampikulsakul, E. Byon, S. Huang, S. Sheng, and M. You, "Condition monitoring of wind power system with non-parametric regression analysis," *IEEE Transactions on Energy Conversion*, vol. 29, no. 2, pp. 288–299, 2014.
- [28] E. Taslimi-Renani, M. Modiri-Delshad, M. F. M. Elias, and N. A. Rahim, "Development of an enhanced parametric model for wind turbine power curve," *Applied Energy*, vol. 177, pp. 544–552, 2016.
- [29] H.-H. Yang, M.-L. Huang, and S.-W. Yang, "Integrating auto-associative neural networks with hotelling T2 control charts for wind turbine fault detection," *Energies*, vol. 8, no. 10, pp. 12100–12115, 2015.
- [30] P. B. Dao, W. J. Staszewski, T. Barszcz, and T. Uhl, "Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data," *Renewable Energy*, vol. 116, pp. 107–122, 2018.
- [31] P. Cambron, R. Lepvrier, C. Masson, A. Tahan, and F. Pelletier, "Power curve monitoring using weighted moving average control charts," *Renewable Energy*, vol. 94, pp. 126–135, 2016.
- [32] P. Cambron, A. Tahan, C. Masson, and F. Pelletier, "Bearing temperature monitoring of a wind turbine using physics-based model," *Journal of Quality in Maintenance Engineering*, vol. 23, no. 4, pp. 479–488, 2017.
- [33] P. Cambron, C. Masson, A. Tahan, and F. Pelletier, "Control chart monitoring of wind turbine generators using the statistical inertia of a wind farm average," *Renewable Energy*, vol. 116, pp. 88–98, 2018.
- [34] G. Helbing and M. Ritter, "Power curve monitoring with flexible EWMA control charts," in *Proceedings of the 2017*

- International Conference on Promising Electronic Technologies*, Deir El-Balah, State of Palestine, October 2017.
- [35] L. Wang, Z. Zhang, H. Long, J. Xu, and R. Liu, "Wind turbine gearbox failure identification with deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1360–1368, 2017.
 - [36] H.-H. Yang, M.-L. Huang, C.-M. Lai, and J.-R. Jin, "An approach combining data mining and control charts-based model for fault detection in wind turbines," *Renewable Energy*, vol. 115, pp. 808–816, 2018.
 - [37] E. Yashchin, "Estimating the current mean of a process subject to abrupt changes," *Technometrics*, vol. 37, no. 3, pp. 311–323, 1995.
 - [38] G. Capizzi and G. Masarotto, "An adaptive exponentially weighted moving average control chart," *Technometrics*, vol. 45, no. 3, pp. 199–207, 2003.
 - [39] L. Shu, "An adaptive exponentially weighted moving average control chart for monitoring process variances," *Journal of Statistical Computation and Simulation*, vol. 78, no. 4, pp. 367–384, 2008.
 - [40] J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics*, vol. 32, no. 1, pp. 1–12, 1990.
 - [41] C. A. Lowry, W. H. Woodall, C. W. Champ, and S. E. Rigdon, "A multivariate exponentially weighted moving average control chart," *Technometrics*, vol. 34, no. 1, pp. 46–53, 1992.
 - [42] G. C. Runger and S. S. Prabhu, "A Markov chain model for the multivariate exponentially weighted moving averages control chart," *Journal of the American Statistical Association*, vol. 91, no. 436, pp. 1701–1706, 1996.
 - [43] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.
 - [44] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, pp. 36–55, 2018.
 - [45] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
 - [46] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, pp. 407–499, 2004.
 - [47] Classification and Regression Training, 2018, <https://github.com/topepo/caret/>.