Data-Driven Wind Turbine Power Generation Performance Monitoring

Huan Long, Long Wang, Zijun Zhang, *Member, IEEE*, Zhe Song, and Jia Xu

***Abstract*—This paper investigates the wind turbine power generation performance monitoring based on supervisory control and data acquisition (SCADA) data. The proposed approach identifies turbines with weakened power generation performance through assessing the wind power curve profiles. Profiles that statistically summarize the curvatures and shapes of a wind power curve over consecutive time intervals are constructed by fitting power curve models into SCADA data sets with a least square method. To monitor the variations of wind power curve profiles over time, multivariate and residual approaches are introduced and applied. Two blind industrial studies are conducted to validate the effectiveness of the proposed monitoring approach, and the results demonstrate high accuracy in detecting the abnormal power curve profiles of wind turbines and their associated time intervals.**

***Index Terms*—Multivariate approach, performance monitoring, power curve, residual analysis, wind energy.**

I. Introduction

THE technological maturity and foreseeable improvement of cost competitiveness made wind energy a widely accepted energy solution to sustainable living and the limited supply of fossil fuels. The wind power industry has rapidly expanded during the past decades, and the hot spot of wind energy technology development has gradually shifted from improving the wind power generation efficiency [1]–[3] to serving the wind farm operations and maintenance (O&M) at present [4], [5]. Advancing wind farm O&M technologies including condition monitoring is of great interest to wind farm operators because renewable energy subsidies are not sustainable.

In literature, studies of wind farm condition monitoring could be majorly categorized as frequency-domain and time-domain analyses. Frequency-domain analysis studies examined the power spectrum of the frequencies of vibration and acoustic emission signals, and they aimed to find the reason for the root causes of wind turbine subsystem abnormal conditions [6]–[9]. Wavelet transformation and its extended methods have been widely applied. Tang *et al.* [10] applied Morlet wavelet transformation to filter vibration signal noise and Wigner–Ville-distribution-based time-frequency analysis to identify faults. Tsai *et al.* [11] presented a continuous-wavelet-transform-based approach to identify the impaired conditions of wind turbine blades. Watson *et al.* [12] presented a wavelet-based monitoring approach to analyze high-frequency power output data. The frequency-domain analysis techniques are powerful in diagnosing the faults of individual wind turbine assemblies; however, they require extra instruments to collect high-frequency data that can induce additional cost and system complexity.

Time-domain analyses focused on detecting wind turbine abnormal statuses through tracking the trends of condition parameter values [13]–[15]. The power generation performance of wind turbines can be characterized by its power curves constructed from real-time supervisory control and data acquisition (SCADA) data. Thus, in time-domain analyses, previous wind turbine condition monitoring studies focused on analyzing the quality of produced power curves. Kusiak *et al.* [16] introduced nonparametric methods for modeling the wind power curve from industrial data and analyzed model fitting residuals to detect anomalies. The approach presented in [16] was then extended to realize online monitoring [17]. Kusiak and Li [18] investigated fault diagnosis through analyzing the patterns of power curve fitting residuals. The philosophy of the works in [16]–[18] is to develop statistical boundaries for detecting outliers based on a derived reference power curve. A similar approach was applied to monitor the vibration of a wind turbine gearbox [19]. In addition to residual analyses, Feng *et al.* [20] and Qiu *et al.* [21] presented physics-based data analysis methods for detecting the anomalies of wind turbine assemblies and performing diagnosis with SCADA data. Compared with the frequency-domain analysis, the time-domain analyses are directly conducted based on the collected SCADA data. Condition monitoring and fault diagnosis both have been studied with the time-domain analysis.

Based on the long-term observation of wind power curves, people recognized that the power curve curvature and shape of a healthy wind turbine are variable rather than fixed. Gill *et al.* [22] and Wang *et al.* [23] applied Copula-based approaches to conduct the initial analyses of the relationships of multiple wind turbine parameters and to detect their variations. The

Manuscript received September 28, 2014; revised March 28, 2015 and May 17, 2015; accepted May 21, 2015. Date of publication June 19, 2015; date of current version September 9, 2015. This work was supported in part by the Early Career Scheme Grant from the Research Grants Council of the Hong Kong Special Administrative Region under Project CityU 138313 and in part by the National Natural Science Foundation of China under Grant 71001050. *(Corresponding author: Zhe Song.)*

H. Long, L. Wang, and Z. Zhang are with the Department of Systems Engineering and Engineering Management, College of Science and Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: zijzhang@cityu.edu.hk).

Z. Song is with the School of Business, Nanjing University, Nanjing 210093, China (e-mail: zsong1@nju.edu.cn).

J. Xu is with the Centre of Wind Farm Data Analysis and Performance Optimization, China Longyuan Power Group Corporation Ltd., Beijing 100034, China (e-mail: xujia@clypg.com.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2015.2447508

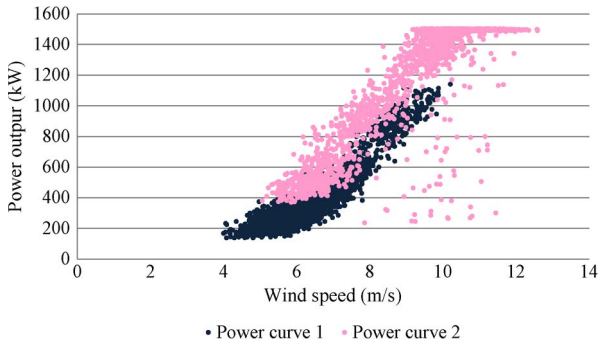


Fig. 1. Wind power curves of a turbine in two different time windows.

reported monitoring approaches [16]–[18] based on fixed reference power curves did not count the contribution of wind power curve variations and thus could produce false alarms. The wind power curve can shift due to factors rather than faults, such as the change of air density. Thus, it is challenging to determine the wind turbine conditions with analyzing individual data points.

In this paper, a monitoring approach is proposed to monitor the wind power generation performance through analyzing the variation of wind power curves rather than individual data points. The SCADA data of a wind turbine are partitioned into subdata sets based on consecutive equal time intervals, and each subdata set provides a power curve of the wind turbine over the time interval. Linear and Weibull cumulative distribution function (WCDF)-based power curve models are applied to generate power curve profiles by fitting into the subdata sets. Power curve profiles are composed of power curve model parameters that depict power curve characteristics. A multivariate approach is applied to monitor multiple parameters simultaneously, and a residual approach is utilized to analyze the residuals of fitting power curves. The effectiveness of the presented approach is demonstrated through two blind industrial studies.

II. Wind Power Curve Profile

In reality, the power curve of a wind turbine is dynamic due to the variations of factors, including the weather, the air density, system controls, etc. Let t denote M time windows with equal lengths, where $t = 1, 2, \ldots, M$, and the power curves of a wind turbine in different time windows deviate, as shown in Fig. 1. It is observable that the power curve shifts over time when the wind turbine is healthy. In addition, power curve 2 includes many outliers. A traditional monitoring approach for analyzing individual points will generate a set of alarms for power curve 2, which is not necessary. To generate more reliable monitoring results, a data-driven profile monitoring approach is proposed.

The target monitored in the proposed monitoring framework includes the wind power output, i.e., P, and the wind power curve profile described in Definition 1.

Definition 1: Given a data set, i.e., $\mathcal{S} = \{(v_1, P_1), (v_2, P_2), \ldots, (v_N, P_N)\}$, where v is the wind speed, P is the power output, and N is the number of data points, and a model, i.e., $f(\mathbf{x})$, for fitting $\mathcal{S}$ with a least square method, where $\mathbf{x}$ is a vector of parameters in $f(\cdot)$, the profile for depicting the characteristic of the wind power curve contained in $\mathcal{S}$ is $\mathbf{x}$ if and only if $\mathbf{x}$ satisfies

$$\mathbf{x} = \arg\min \left\{ \sum_{n=1}^{N} \left(P_n - f(v_n, \mathbf{x})\right)^2 \right\}. \tag{1}$$

Let t index $\mathcal{S}, t = 1, 2, \ldots, M$, which is a set of power curve profiles; $\mathbf{x}_t$ can be obtained through (1) for monitoring. Numerous power curve modeling approaches have been reported in [24]–[26]. In this paper, we select two types of profiles, i.e., the linear profile and the WCDF profile, to demonstrate the accuracy and effectiveness of the proposed monitoring framework.

A. Linear Profile

The fundamental approach for approximating a wind power curve is a stepwise linear model shown as follows:

$$P = \begin{cases} 0, & v_{\text{ci}} > v, v > v_{\text{co}} \\ av + b, & v_{\text{ci}} \le v < v_r \\ P_{\max}, & v_r \le v \le v_{\text{co}} \end{cases} \tag{2}$$

where a is the slope, b is the intercept, $P_{\max}$ is the maximal power capacity, v_{ci} is the cutin wind speed, v_r is the rate wind speed, and v_{co} is the cutout wind speed.

Based on (2), it is observable that the characteristics of wind power curves are depicted by both a and b. Through fitting model (2) to $\mathcal{S}_t$ by (1), the linear power curve profiles are $\mathbf{x}_t = (a_t, b_t)^T$, where a_t and b_t yield the least square error for fitting (2) to $\mathcal{S}_t$.

B. WCDF Profile

Since the shapes of the WCDF function and the wind power curve are highly similar and the variable in the Weibull distribution is naturally positive, wind power curves can be described by the following WCDF-based stepwise model:

$$P = \begin{cases} 0, & v_{\text{ci}} > v, v > v_{\text{co}} \\ P_{\max}\left(1 - e^{-(x/c)^k}\right), & v_{\text{ci}} \le v < v_r \\ P_{\max}, & v_r \le v \le v_{\text{co}} \end{cases} \tag{3}$$

where k and c are the shape and scale parameters of the Weibull distribution, respectively. Other notations have been presented in Section II-A. Similar as Section II-A, the WCDF profiles, i.e., $\mathbf{x}_t = (k_t, c_t)^T$, are generated through fitting model (3) to $\mathcal{S}_t$.

III. Monitoring Approaches

The wind power curve monitoring can be decomposed to three scenarios according to v_{ci}, v_r, and v_{co}, respectively. When $v < v_{\text{ci}}$ and $v > v_{\text{co}}$, wind turbines do not produce power, and the power curve monitoring is not needed in this scenario. When $v_r \le v \le v_{\text{co}}$, the wind power is rated at $P_{\max}$. To monitor turbine conditions, we can simply apply line $P = P_{\max}$ to

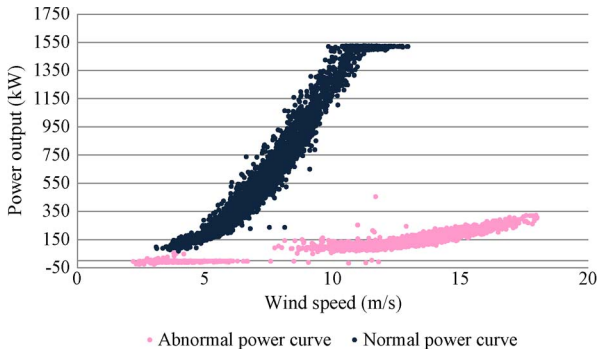


Fig. 2. Power curves with abnormal and normal curvatures.

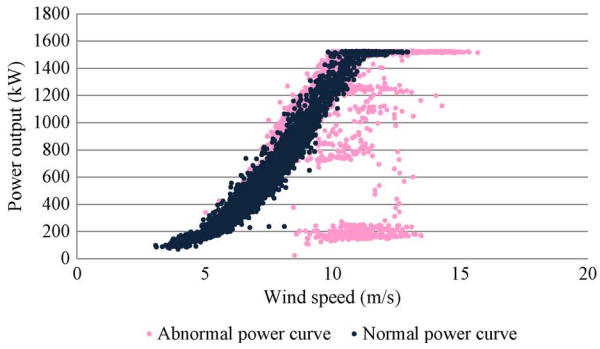


Fig. 3. Power curves with abnormal and normal shapes.

examine the deviation of the observed power outputs from $P_{\max}$. When $v_{\text{ci}} \leq v < v_r$, the shape and curvature of the power curve of a healthy wind turbine are variable over time (see Fig. 1). To monitor the quality of power curves in such a scenario, profiles are developed, as described in Section II, to conclude the characteristics of power curves based on sub-data sets. Impaired power curves can demonstrate abnormal curvature and shape caused by different reasons, as shown in Figs. 2 and 3. Fig. 2 demonstrates that the abnormal curvature of power curves might be caused by subsystem malfunctions, severe weather conditions, power curtailment commands, etc. Fig. 3 shows that the abnormal shape of power curves is possibly due to the system degradation and inappropriate control settings. To identify the power curve with abnormal curvature and shape, a multivariate approach for monitoring power curve profiles and a residual approach for monitoring power curve fitting errors are introduced in the next sections.

A. Multivariate Approach

As presented in (2) and (3), the linear and WCDF-based power curve profiles are the vectors of two parameters. A straightforward solution of monitoring profiles is to construct control charts to monitor the two parameters separately. However, such solution presumes the independence between two parameters and omits the correlation. It is obvious that the two parameters in (2) and (3) are dependent. To monitor the two parameters simultaneously, Hotelling's T^2 control chart [27], [28] is applied.

Given profiles $\mathbf{x}_t = (x_{0,t}, x_{1,t})^T$, $t = 1, 2, \ldots, M$, where $x_{0,t}$ is a_t in (2) and k_t in (3), and $x_{1,t}$ is b_t in (2) and c_t in (3), the expected value $\boldsymbol{\mu}$ and variance–covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{x}_t$ are defined as

$$\boldsymbol{\mu} = (\mu_0, \mu_1)^T \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_0^2 & \sigma_{01}^2 \\ \sigma_{01}^2 & \sigma_1^2 \end{pmatrix} \tag{4}$$

where μ_0 and μ_1 are the means of $x_{0,t}$ and $x_{1,t}$ $\forall_t$, respectively, σ_0^2 and σ_1^2 are the corresponding variances, and σ_{01}^2 is the covariance.

Based on (4), the sample statistic in the T^2 control chart is computed as

$$T_t^2 = (\mathbf{x}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_t - \boldsymbol{\mu}). \tag{5}$$

The T^2 control chart only has an upper control limit (UCL), and it is $\chi_{2,\alpha}^2$ when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known. However, in reality, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are typically unavailable. Thus, $\mathbf{X}$ and $\mathbf{S}$, which are the unbiased estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, are defined in the following and need to be computed:

$$\mathbf{X} = \left(\frac{\sum_{t=1}^{M} x_{0,t}}{M}, \frac{\sum_{t=1}^{M} x_{1,t}}{M} \right)^T$$
$$\mathbf{S} = \begin{pmatrix} \text{MSE}\left(\frac{1}{N} + \frac{\bar{v}^2}{S_{vv}}\right) & -\text{MSE}\left(\frac{\bar{v}^2}{S_{vv}}\right) \\ -\text{MSE}\left(\frac{\bar{v}^2}{S_{vv}}\right) & \frac{\text{MSE}}{S_{vv}} \end{pmatrix} \tag{6}$$

where the MSE and S_{vv} are calculated as

$$\text{MSE} = \frac{\sum_{t=1}^{M} (N-2)^{-1} \sum_{n=1}^{N} (P_n - f(v_n, \mathbf{x}))^2}{M} \tag{7}$$

$$S_{vv} = \sum_{n=1}^{N} (v_n - \bar{v})^2. \tag{8}$$

The sample statistic for the T^2 control chart is then described as follows:

$$T'^2 = \frac{M}{M-1} (\mathbf{x}_t - \mathbf{X})^T \mathbf{S}^{-1} (\mathbf{x}_t - \mathbf{X}). \tag{9}$$

The UCL of the T^2 control chart is then equal to $2F_{2,(N-2)M,\alpha}$ [27]. In addition to the T^2 control chart, the generalized variance (GV) chart with the UCL, the central line (CL), and the lower control limit (LCL) expressed in the following is also conducted to examine the shift of Σ:

$$\begin{aligned} \text{UCL} &= \frac{|\mathbf{S}|}{B_1} \left(B_1 - 3B_2^{0.5}\right) \\ \text{CL} &= B_1 |\mathbf{S}| \\ \text{LCL} &= \frac{|\mathbf{S}|}{B_1} \left(B_1 - 3B_2^{0.5}\right) \end{aligned} \tag{10}$$

where B_1 and B_2 are computed as

$$B_1 = \frac{1}{(N-1)^q}\prod_{i=1}^{q}(N-i) \tag{11}$$

$$B_2 = \frac{1}{(N-1)^{2q}}\prod_{i=1}^{q}(N-i)\left[\prod_{j=1}^{q}(N-j+2)-\prod_{j=1}^{q}(N-j)\right] \tag{12}$$

given the number of parameters q.

B. Residual Approach

The multivariate approach is effective in detecting power curves with an irregular curvature but might not be sensitive to the variation of power curve shapes. For example, as shown in Fig. 3, it is obvious that a similar pair of $\mathbf{x}$ will be generated if we fit either model (2) or model (3) to the data. The residuals of model fitting can be analyzed to detect power curves with impaired shapes. The individual–moving range (I-MR) chart [28] composed of two charts, i.e., the individual chart and the moving range (MR) chart, is applied to perform monitoring. The integration of individual and MR charts offers two dimensions of residual monitoring information, i.e., the trend of model fitting residuals and the variation between model fitting residuals.

Let e_t denote the mean absolute percentage error (MAPE) of fitting model $f(\cdot)$ to the data sets, i.e., $\mathcal{S}_t$, $t = 1, 2, \ldots, M$, which are expressed in (13); the averages of e_t, i.e., $\bar{e}$, for $t = 1, 2, \ldots, M$ can be obtained through

$$e_t = \left(\frac{1}{N}\right)\sum_{n=1}^{N}|P_n - f(v_n)/P_n| \tag{13}$$

$$\bar{e} = \frac{\sum_{t=1}^{M} e_t}{M}. \tag{14}$$

To construct the I-MR chart for e_t, MR_t needs to be computed based on

$$\mathrm{MR}_t = |e_t - e_{t-1}|. \tag{15}$$

Let $\overline{\mathrm{MR}}$ be the average of MR_t; the UCL, the CL, and the LCL of the individual chart are presented as

$$\mathrm{UCL} = \bar{e} + 3\frac{\overline{\mathrm{MR}}}{d_2} \qquad \mathrm{CL} = \bar{e} \qquad \mathrm{LCL} = \bar{e} - 3\frac{\overline{\mathrm{MR}}}{d_2} \tag{16}$$

where d_2 is 1.128 because we consider the moving range of two consecutive observations [21].

To construct the MR chart, its UCL, CL, and LCL are set up as

$$\mathrm{UCL} = D_4\overline{\mathrm{MR}} \qquad \mathrm{CL} = \overline{\mathrm{MR}} \qquad \mathrm{LCL} = D_3\overline{\mathrm{MR}} \tag{17}$$

where D_4 and D_3 are constants depending on n. Since $n = 2$ in this paper, $D_4 = 3.267$ and $D_3 = 0$ according to the work in [28].

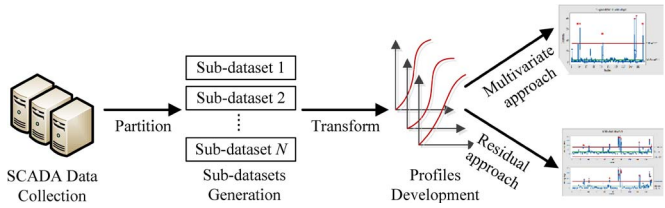


Fig. 4. Online monitoring procedure of the proposed model.

C. Online Monitoring Procedure

The online monitoring procedure of the proposed approach shown in Fig. 4 is composed of four steps as follows.

Step 1) Partition the SCADA data set to subdata sets according to the fixed time intervals.
Step 2) Generate the power curve profiles and the power curve fitting errors based on the subdata sets.
Step 3) Apply the multivariate approach and the residual approach to develop control charts for monitoring.
Step 4) Continuously collect SCADA data, produce new power curve profiles, and apply the developed control charts.

IV. Blind Industrial Studies

This section presents two blind industrial studies based on the proposed approach. The first study is conducted based on the 10-s SCADA data of wind turbines collected from a 100-MW class wind farm in the USA. The second study investigates the 10-min wind turbine SCADA data collected from a wind farm owned by China Long Yuan Power Group Corporation Ltd. In the two blind studies, the proposed approach was first applied without the prior knowledge of the considered wind turbines to generate the monitoring results. Next, the results of the first blind study are validated with the fault logs, whereas the results of the second study are examined by the domain experts in the China Long Yuan Power Group Corporation.

A. Study 1

In study 1, the v_{ci}, v_r, and v_{co} values of the considered wind turbine are 3.5, 11.5, and 20 m/s, respectively. The monitoring results are compared with the fault logs to show the detection accuracy.

1) Data Description: In this paper, the 10-s data collected for 2 months from January 1, 2011 to February 28, 2011 from a wind turbine in a U.S. wind farm were applied to the power curve profile construction and the condition monitoring. The collected data contained the values of the wind turbine performance parameters, the wind conditions, and the fault logs.

The SCADA data are partitioned into 177 subfiles based on the fixed time interval, and each subfile included 3600 data points that recorded the 10-h power generation performance of the wind turbine.

2) Power Curve Profile Computation: The least square method is applied to fitting models (2) and (3) based on the data of 177 subfiles to generate two types of power curve profiles.

TABLE I
STATISTICAL SUMMARY OF THE LINEAR PROFILES

Linear profiles	Min	Max	Median	Mean	Standard deviation
a	-52.46	260.53	177.93	158.69	66.07
b	-1331.50	1422.80	-688.22	-596.55	450.04

TABLE II
STATISTICAL SUMMARY OF THE WCDF-BASED PROFILES

Weibull CDF based profiles	Min	Max	Median	Mean	Standard deviation
c	1.00	100.00	8.99	12.49	16.06
k	1.00	5.00	3.62	3.63	0.84

TABLE III
FITTING RESULTS OF THE LINEAR PROFILES

Linear Profiles	Min	Max	Median	Mean	Standard deviation
RMSE (W)	2.44	548.93	84.95	90.14	64.61
MAPE (%)	1.58	4757.30	19.87	115.31	492.94

TABLE IV
FITTING RESULTS OF THE WCDF-BASED PROFILES

Weibull CDF based profile	Min	Max	Median	Mean	Standard deviation
RMSE (W)	4.98	608.29	67.20	72.59	61.21
MAPE (%)	1.11	4812.62	14.93	116.96	488.21

The ranges of c and k for fitting model (3) are set to $1 \leq c \leq 100$ and $1 \leq k \leq 5$, respectively. The range of c is determined artificially. The minimum and maximum of k are set to 1 and 5, respectively, because $k > 5$ will lead to $P = P_{\max}$ when $v < v_r$ and $k < 1$ will lead to a concave curve for the WCDF rather than an S-shape curve. The generated power curve profiles are summarized in Tables I and II.

In Table I, it is observable that the values of a and b in the linear profiles have significant variations. However, the median and the mean indicate that a and b are more likely to be positive and negative among all linear profiles. Table II summarizes the values of c and k in the WCDF profiles statistically. It is observable that the values of c and k have the possibility of reaching the minimal and maximal boundaries. The median and mean of c and k reflect that they normally should be around 10 and 3.6, respectively.

The fitting residuals of models (2) and (3) are statistically summarized in Tables III and IV, respectively. The RMS error (RMSE) is the square root of $(N-2)^{-1}\sum_{n=1}^{N}(P_n - f(v_n, \mathbf{x}))^2$, and the MAPE is computed based on (13).

Based on the RMSEs in Tables III and IV, we observe that the minimal and maximal RMSE values provided by the linear models are less than those provided by the WCDF-based models in model fittings. However, the median, mean, and standard deviation of the RMSE offered by the WCDF-based models are significantly lower than those of the linear models, which indicate that the general performance of fitting wind power curves based on the WCDF-based models is slightly better than that based on linear models. It is reasonable because the curvature of the WCDF model is closer to that of the reference wind power curve.

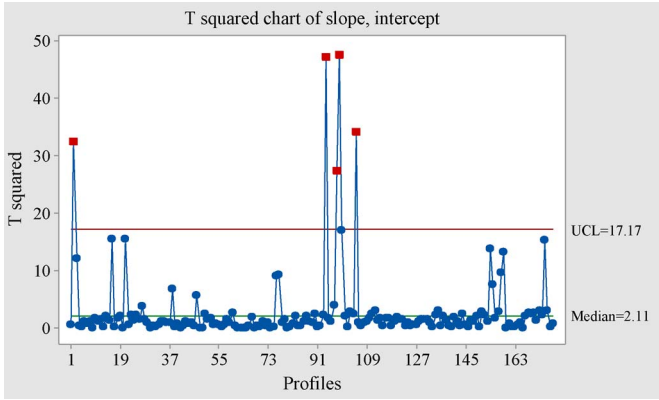


Fig. 5. Phase 1 monitoring of the linear profiles with the T^2 chart.

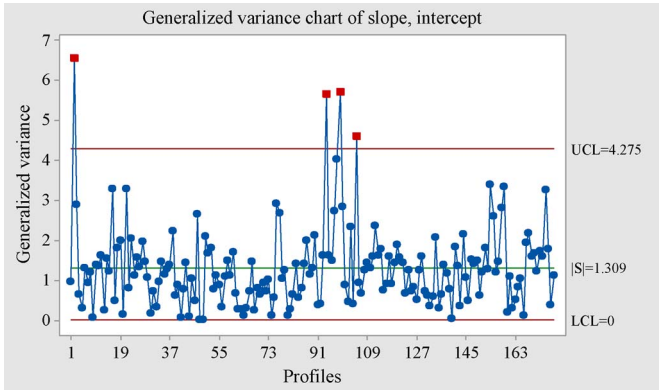


Fig. 6. Phase 1 monitoring of the linear profiles with the GV chart.

3) Monitoring With Multivariate Approach: The monitoring process is split into two phases, i.e., Phases 1 and 2, because the prior information of the wind turbine faults is unknown. In Phase 1, the control limits are developed to filter outliers. In Phase 2, the control limits are recomputed after removing the outliers to perform monitoring. Next, the abnormal profiles detected by the multivariate and residual approaches are analyzed by checking with the fault logs.

The multivariate approach is first applied to monitor the linear profiles generated in Section IV. In the Phase 1 analysis, the T^2 and GV charts are developed in Figs. 5 and 6, respectively, to filter the outliers.

It is observable that profiles 2, 94, 98, 99, and 105 are considered outliers by the T^2 chart, whereas the GV chart provides a similar result. Thus, the information of profiles 2, 94, 98, 99, and 105 is eliminated, and the control limits of the T^2 and GV charts are recalculated. The monitoring results in Phase 2 are demonstrated in Figs. 7 and 8.

In the Phase 2 analysis, the wind turbine condition in profile 100 is considered abnormal because the sample statistic exceeds the limit of the T^2 chart. In conclusion, the wind turbine power curves in a total of 6 linear profiles (2, 94, 98, 99, 100, and 105, respectively) are considered abnormal.

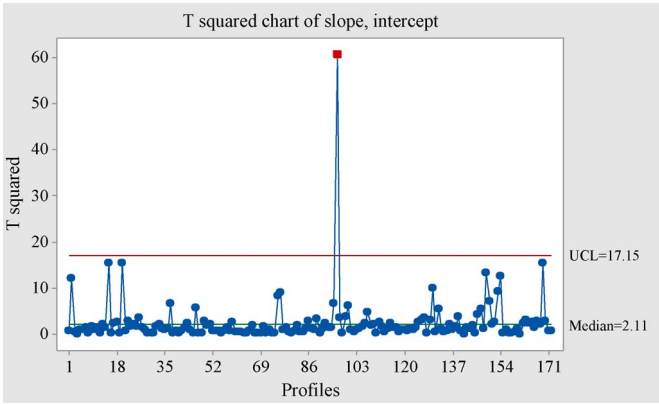


Fig. 7. Phase 2 monitoring of the linear profiles with the T^2 chart.

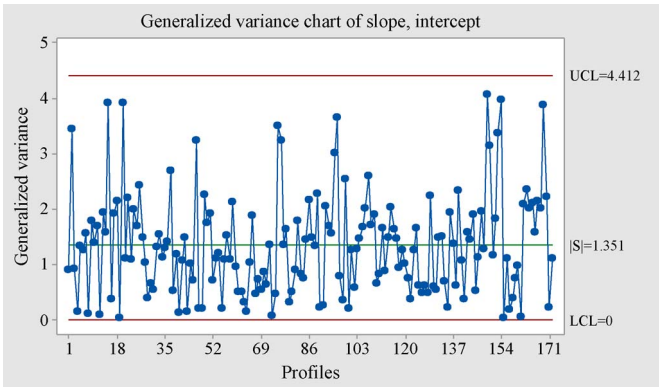


Fig. 8. Phase 2 monitoring of the linear profiles with the GV chart.

TABLE V
MONITORING RESULTS OF THE WCDF PROFILES WITH THE MULTIVARIATE APPROACH

	Phase 1	Phase 2
T^2 chart	16,21,76,154,149,174	3,77,99
GV chart	2,16,21,47,76,154,155,158,159,174	3,77,94,98

A similar monitoring procedure is applied to examine the WCDF profiles. The indexes of the abnormal power curve profiles detected in the analyses of Phases 1 and 2 are provided in Table V. Through monitoring the WCDF profiles, the T^2 and GV charts detect that the wind turbine power curves are abnormal when $t = 2, 3, 16, 21, 47, 76, 77, 94, 98, 99, 154, 155, 158, 159,$ and 174.

4) Monitoring With Residual Approach: The abnormal profiles detected by the multivariate approach were excluded before applying the residual approach. The condition monitoring with the residual approach also includes two phases. Phase 1 is applied to filter the outliers, and Phase 2 aims to detect the anomalies of power curves.

The Phase 1 individual and MR charts based on the linear profiles are demonstrated in Fig. 9. The top chart in Fig. 9 is the individual chart, whereas the bottom chart is the MR chart. Four profiles at $t = 34, 82, 87,$ and 137, respectively, with the MAPE higher than 1000%, are considered as outliers and are filtered. After filtering the outliers in Phase 1, the control

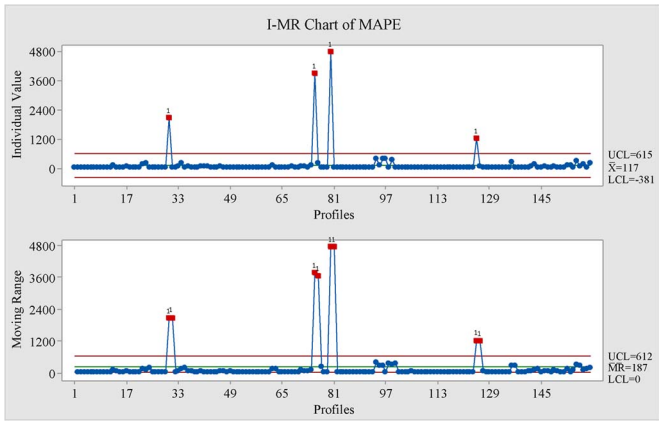


Fig. 9. Phase 1 monitoring of the linear profiles with the residual approach.

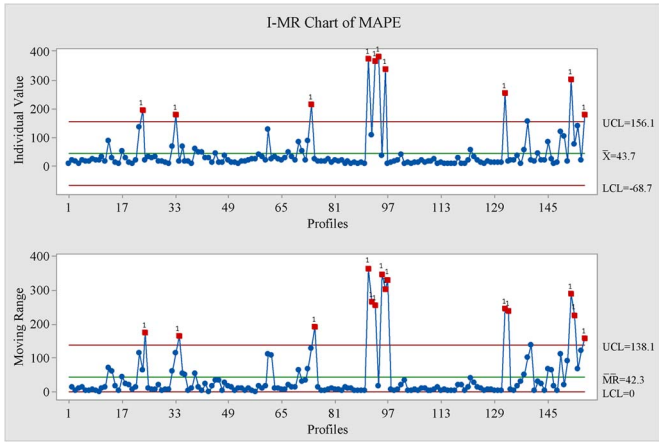


Fig. 10. Phase 2 monitoring of the linear profiles with the residual approach.

TABLE VI
MONITORING RESULTS OF THE WCDF PROFILES WITH THE RESIDUAL APPROACH

	Phase 1	Phase 2
I chart	34,82,87,137	27,38,83,106,108,109,111,148,157,172
MR chart	34,82,87,137	27,38,83,106,108,109,111,148,157,172

limits of the individual and MR charts are recomputed to conduct the Phase 2 analysis based on the linear profiles presented in Fig. 10. Table VI presents the profiles with irregular fitting errors, with the residual approach in Phases 1 and 2 based on the WCDF profiles. All the detected abnormal residuals of the profiles include the profiles at $t = 27, 34, 38, 82, 83, 106, 108, 109, 111, 137, 148, 157, 172,$ and 177.

5) Analysis of Detected Faults: Through the eye observation, we discover that power curves based on 27 subdata sets display abnormal curvatures. The multivariate approach successfully identified 17 of the 27 profiles based on the linear and WCDF profiles. The patterns of 27 irregular profiles can be briefly categorized to three types, as shown in Figs. 11–13, respectively, based on the root causes.

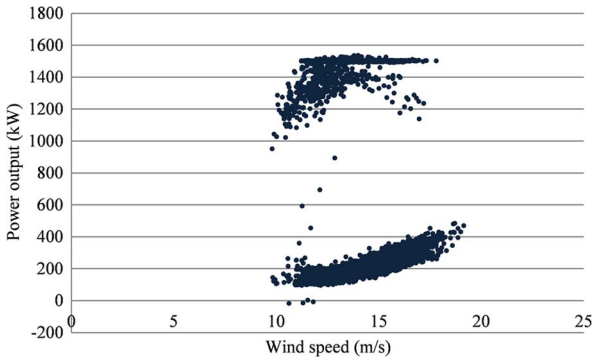


Fig. 11. Power curve of subdata set 2.

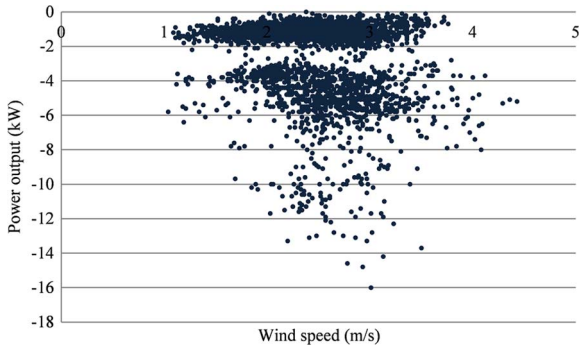


Fig. 12. Power curve of subdata set 16.

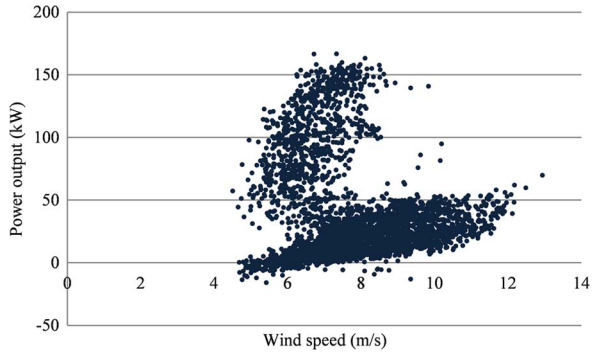


Fig. 13. Power curve of subdata set 154.

The anomaly of the power curve in Fig. 11 is majorly caused by the power curtailment command, although pitch faults, low gearbox oil pressure, and battery problems were reported in the log. In Fig. 12, the wind turbine does not generate power but draws power from the grid because the wind speed is lower than the cutin speed most of the time. The power curve in Fig. 13 is abnormal because, during the period, the wind turbine suffered a series of faults, including the yaw problem, diverter malfunction, an incorrect blade angle, generator brushes worn, cable twisting, and periodical restartup.

The unidentified profiles through the multivariate approach generally displayed a pattern similar to that in Fig. 14. In subdata set 27, the wind speed fluctuated around the cutin wind speed, and it is difficult to control the wind turbine to produce a stable power curve.

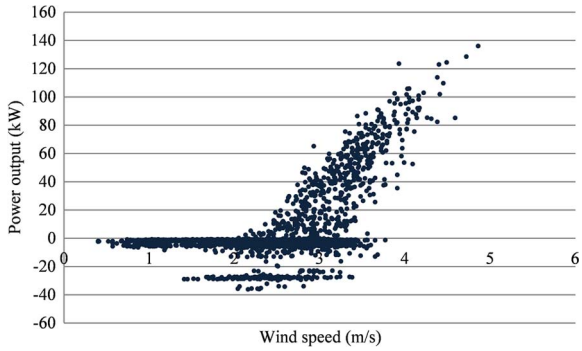


Fig. 14. Power curve of subdata set 27.

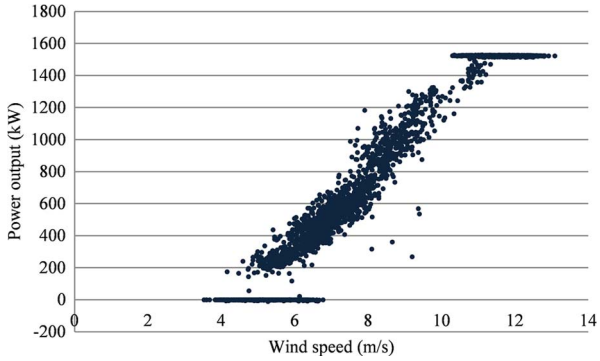


Fig. 15. Power curve with a normal curvature but an impaired shape.

The residual approach detects five of the remaining ten visually detected abnormal profiles because those profiles also lead to significant residuals of fitting models. The more significant contribution of applying the residual approach is to detect ten more profiles with the pattern of the power curve shown in Fig. 15. Although the power curve in Fig. 15 displays a normal curvature, the data points reflect that the wind turbine was temporarily shut down many times when $v > v_{\mathrm{ci}}$. The log recorded that the wind turbine experienced faults including emergency stop, brake malfunction, and generator brush worn, and it received a repair during the operating period.

B. Study 2

In this paper, the 10-min SCADA data of a wind farm including 32 1.6-MW class wind turbines in China collected from April 1, 2014 to July 30, 2014 are investigated. The v_{ci}, v_r, and v_{co} values of the considered wind turbines are 3, 10, and 21 m/s, respectively. The collected SCADA data are partitioned into subdata sets, with two-day intervals for generating the power curve profiles, and each subdata set includes 288 data points. The proposed monitoring approach has been applied to monitor the power generation performance of all 32 wind turbines. Since the fault logs are not available and the industrial partner only manually examined the power generation performance of two wind turbines, the proposed approach is validated through comparing the generated monitoring results and the industrial analysis report of the two selected wind turbines.

TABLE VII
MONITORING RESULTS OF THE SELECTED WIND TURBINE 1

Time period	April 1st – 2nd	April 3rd – 4th	April 5th – 6th	April 7th – 8th	April 9th – 10th
Results			**a, c**	**c**	**c**
Time period	April 11th – 12th	April 13th – 14th	April 15th – 16th	April 17th – 18th	April 19th – 20th
Results					**c**
Time period	April 21st – 22nd	April 23rd – 24th	April 25th – 26th	April 27th – 28th	April 29th – 30th
Results	**c, d**	**d**			**c, d**
Time period	May 1st – 2nd	May 3rd – 4th	May 5th – 6th	May 7th – 8th	May 9th – 10th
Results	**d**			**a, c**	**c**
Time period	May 11th – 12th	May 13th – 14th	May 15th – 16th	May 17th – 18th	May 19th – 20th
Results					
Time period	May 21st – 22nd	May 23rd – 24th	May 25th – 26th	May 27th – 28th	May 29th – 30th
Results					

TABLE VIII
MONITORING RESULTS OF THE SELECTED WIND TURBINE 2

Time period	April 1st – 2nd	April 3rd – 4th	April 5th – 6th	April 7th – 8th	April 9th – 10th
Results			**c**	**c**	
Time period	April 11th – 12th	April 13th – 14th	April 15th – 16th	April 17th – 18th	April 19th – 20th
Results					**d**
Time period	April 21st – 22nd	April 23rd – 24th	April 25th – 26th	April 27th – 28th	April 29th – 30th
Results	**c, d**	**c, d**	**d**		
Time period	May 1st – 2nd	May 3rd – 4th	May 5th – 6th	May 7th – 8th	May 9th – 10th
Results					**b**
Time period	May 11th – 12th	May 13th – 14th	May 15th – 16th	May 17th – 18th	May 19th – 20th
Results					
Time period	May 21st – 22nd	May 23rd – 24th	May 25th – 26th	May 27th – 28th	May 29th – 30th
Results					

TABLE IX
INDUSTRIAL ANALYSIS RESULTS OF THE TWO TURBINES

Time period	Power curve analysis results
April 1st – April 14th	Impaired
April 15th – April 19th	Improved but slightly scattered
April 20th – April 24th	Impaired
April 25th – April 30th	Improved but slightly scattered
May 1st – May 9th	Significantly degraded
May 10th – May 11th	Scattered
May 12th – May 30th	Improved
June 1st	Scattered

The same monitoring procedure in study 1 is applied here. The linear and WCDF-based models are applied to generate power curve profiles based on subdata sets covering the 10-min SCADA data of two days. The multivariate and residual approaches are employed to monitor the linear and WCDF-based power curve profiles and their fitting errors. The monitoring results of the selected wind turbines 1 and 2 are presented in Tables VII and VIII, respectively. The analysis results of domain experts are provided in Table IX.

The domain experts only examined two wind turbines for the period from April 1, 2014 to June 1, 2014. Thus, the monitoring results generated with the proposed approach over the same period are selected and reported in Tables VII and VIII. In Tables VII and VIII, symbols **a**, **b**, **c**, and **d** are applied to label the abnormal power curves detected by the multivariate approach based on the linear profiles, by the residual approach based on the linear profiles, by the multivariate approach based on the WCDF profiles, and by the residual approach based on the WCDF profiles, respectively. Table IX reports the brief summary of the manual examinations of the two selected wind turbines, which provided similar power generation performance. Through comparing Tables VII and VIII with Table IX, we could observe that the monitoring results generated by the proposed approach could basically match the summary of the industrial examination report. Since such monitoring study was conducted based on 10-min data, better results could be produced if higher frequency data, such as 10-s data, are available.

V. CONCLUSION

A data-driven approach for monitoring the variation of a wind turbine power curve was presented to identify impaired power generation performance. To depict the curvature and shape of a wind power curve, power curve profiles were developed. Two power curve models, i.e., the linear and WCDF-based models, were applied to generate the linear and WCDF power curve profiles based on a sequence of SCADA subdata sets. In the monitoring, a multivariate approach was applied to analyze power curve profiles, whereas a residual approach was utilized to analyze the errors produced by fitting power curve models.

Two blind industrial studies were conducted to prove the effectiveness and accuracy of the proposed monitoring approach. The results demonstrated that good monitoring accuracy was obtained and that few false alarms were generated. It was because the monitoring analysis of the proposed approach was conducted based on subdata sets rather than individual points.

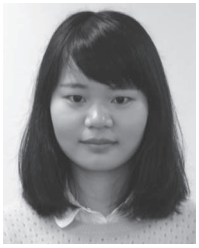
The proposed approach could be implemented for online monitoring. It could accurately detect faults but not as timely as point-based condition monitoring methods. This is simply because a certain amount of time was required for constructing new power curve profiles.

Future research will investigate a mechanism to partition data sets with the completion of the power curve information to enhance the monitoring power of the proposed approach. In addition, the relationship between the wind power and multiple parameters will be considered in the monitoring.

REFERENCES

[1] R. Li and D. Xu, "Parallel operation of full power converters in permanent-magnet direct-drive wind power generation system," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1619–1629, Apr. 2013.
[2] J. Chen, J. Chen, and C. Gong, "New overall power control strategy for variable-speed fixed-pitch wind turbines within the whole wind velocity range," *IEEE Trans. Ind. Electron.*, vol. 60, no. 7, pp. 2652–2660, Jul. 2013.
[3] A. Kusiak and Z. Zhang, "Control of wind turbine power and vibration with a data-driven approach," *Renew. Energy*, vol. 43, pp. 73–82, Jul. 2012.
[4] A. Kusiak, Z. Zhang, and G. Xu, "Minimization of wind farm operational cost based on data-driven models," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 756–764, Jul. 2013.
[5] S. Bernal-Perez, S. Ano-Villalba, R. Blasco-Gimenez, and J. Rodriguez-D'Derlee, "Efficiency and fault ride-through performance of a diode-rectifier- and VSC-inverter-based HVDC link for offshore wind farms," *IEEE Trans. Ind. Electron.*, vol. 60, no. 6, pp. 2401–2409, Jun. 2013.
[6] X. Gong and W. Qiao, "Current-based mechanical fault detection for direct-drive wind turbines via synchronous sampling and impulse detection," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1693–1702, Mar. 2015.
[7] W. Wang and P. McFadden, "Application of wavelets to gearbox vibration signals for fault detection," *J. Sound Vib.*, vol. 192, no. 5, pp. 927–939, May 1996.
[8] X. Gong and W. Qiao, "Bearing fault diagnosis for direct-drive wind turbines via current-demodulated signals," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3419–3428, Aug. 2013.

[9] K. Iyer, X. Lu, Y. Usama, V. Ramakrishnan, and N. Kar, "A twofold Daubechies-wavelet-based model for fault detection and voltage regulation in SEIGs for distributed wind power generation," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1638–1651, Apr. 2013.
[10] B. Tang, W. Liu, and T. Song, "Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner–Ville distribution," *Renew. Energy*, vol. 35, no. 12, pp. 2862–2866, Dec. 2010.
[11] C. Tsai, C. Hsieh, and S. Huang, "Enhancement of damage-detection of wind turbine blades via cwt-based approaches," *IEEE Trans. Energy Convers.*, vol. 21, no. 3, pp. 776–781, Sep. 2006.
[12] S. Watson, B. Xiang, W. Yang, P. Tavner, and C. Crabtree, "Condition monitoring of the power output of wind turbine generators using wavelet," *IEEE Trans. Energy Convers.*, vol. 25, no. 3, pp. 715–721, Sep. 2010.
[13] H. Sanchez, T. Escobet, V. Puig, and P. Odgaard, "Fault diagnosis of an advanced wind turbine benchmark using interval-based ARRs and observers," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3783–3793, Jun. 2015.
[14] J. Blesa, P. Jimenez, D. Rotondo, F. Nejjari, and V. Puig, "An interval NLPV parity equations approach for fault detection and isolation of a wind farm," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3794–3805, Jun. 2015.
[15] N. Freire, J. Estima, and A. Marques Cardoso, "Open-circuit fault diagnosis in PMSG drives for wind turbine applications," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 3957–3967, Sep. 2013.
[16] A. Kusiak, H. Zheng, and Z. Song, "Models for monitoring wind farm power," *Renew. Energy*, vol. 34, no. 3, pp. 583–590, Mar. 2009.
[17] A. Kusiak, H. Zheng, and Z. Song, "On-line monitoring of power curves," *Renew. Energy*, vol. 34, no. 6, pp. 1487–1493, Mar. 2009.
[18] A. Kusiak and W. Li, "The prediction and diagnosis of wind turbine faults," *Renew. Energy*, vol. 36, no. 1, pp. 16–23, Jan. 2011.
[19] Z. Zhang, A. Verma, and A. Kusiak, "Fault analysis and condition monitoring of the wind turbine gearbox," *IEEE Trans. Energy Convers.*, vol. 27, no. 2, pp. 526–535, Jun. 2012.
[20] Y. Feng, Y. Qiu, C. Crabtree, H. Long, and P. Tavner, "Monitoring wind turbine gearboxes," *Wind Energy*, vol. 16, no. 5, pp. 728–740, Jul. 2013.
[21] Y. Qiu, W. Zhang, M. Cao, Y. Feng, and D. Infield, "An electro-thermal analysis of a variable-speed doubly-fed induction generator in a wind turbine," *Energies*, vol. 8, no. 5, pp. 3386–3402, 2015.
[22] S. Gill, B. Stephen, and S. Galloway, "Wind turbine condition assessment through power curve copula modeling," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 94–101, Jan. 2012.
[23] Y. Wang, D. Infield, B. Stephen, and S. Galloway, "Copula-based model for wind turbine power curve outlier rejection," *Wind Energy*, vol. 17, no. 11, pp. 1677–1688, Nov. 2014.
[24] Z. Zhang, Q. Zhou, and A. Kusiak, "Optimization of wind power and its variability with a computational intelligence approach," *IEEE Trans. Sustain. Energy*, vol. 5, no. 1, pp. 228–236, Jan. 2014.
[25] S. Li, D. Wunsch, E. O'Hair, and M. Giesselmann, "Comparative analysis of regression and artificial neural network models for wind turbine power curve estimation," *ASME J. Solar Energy Eng.*, vol. 123, no. 4, pp. 327–332, Jul. 2001.
[26] T. Ustuntas and A. Sahin, "Wind turbine power curve estimation based on cluster center fuzzy logic modeling," *J. Wind Eng. Ind. Aerodyn.*, vol. 96, no. 5, pp. 611–620, 2008.
[27] L. Kang and S. Albin, "On-line monitoring when the process yields a linear profile," *J. Qual. Technol.*, vol. 32, no. 4, pp. 418–426, Oct. 2000.
[28] D. Montgomery, *Introduction to Statistical Quality Control*. New York, NY, USA: Wiley, 2008.



Huan Long received the B.E. degree in automation from Huazhong University of Science and Technology, Wuhan, China, in 2013. She is currently working toward the Ph.D. degree in the Department of Systems Engineering and Engineering Management, College of Science and Engineering, City University of Hong Kong, Kowloon, Hong Kong.

Her research interests include wind farm layout planning and analysis, wind turbine monitoring, and wind farm performance robust optimization.

Long Wang received the B.Eng. degree in irrigation and drainage engineering and the M.Eng. degree in hydraulic engineering from China Agricultural University, Beijing, China, in 2011 and 2013, respectively, and the M.S. degree (with distinction) in computer science from University College London, London, U.K., in 2014. He is currently working toward the Ph.D. degree in the Department of Systems Engineering and Engineering Management, College of Science and Engineering, City University of Hong Kong, Kowloon, Hong Kong.

Zijun Zhang (M'12) received the B.Eng. degree in systems engineering and engineering management from The Chinese University of Hong Kong, Shatin, Hong Kong, in 2008 and the M.S. and Ph.D. degrees in industrial engineering from the University of Iowa, Iowa City, IA, USA, in 2009 and 2012, respectively.

Currently, he is an Assistant Professor with the Department of Systems Engineering and Engineering Management, College of Science and Engineering, City University of Hong Kong, Kowloon, Hong Kong. His research focuses on data mining and computational intelligence, with applications in the modeling, monitoring, optimization, and operations of systems in wind energy; heating, ventilating, and air conditioning; and wastewater processing domains.

Zhe Song received the Ph.D. degree in industrial engineering from the University of Iowa, Iowa City, IA, USA, in 2008.

After graduation, he continued his research as a Postdoctoral Researcher with the University of Iowa. Since 2009, he has been with the School of Business, Nanjing University, Nanjing, China, as an Associate Professor. His research interests include operations research, data mining, control theory, computational intelligence, and their applications in business, energy, and manufacturing systems modeling and optimization, including mass customization, power plant performance optimization, wind power management, and optimal decision making in heating, ventilating, and air conditioning systems.

Jia Xu received the M.S. degree in instrument science and technology from Xi'an Jiaotong University, Xi'an, China, in 2010.

Currently, he is the Head of the Centre of Wind Farm Data Analysis and Performance Optimization, China Longyuan Power Group Corporation Ltd., Beijing, China.