

Monitoring Wind Turbine Vibration Based on SCADA Data

Zijun Zhang

Andrew Kusiak

e-mail: andrew-kusiak@uiowa.edu

Department of Mechanical and Industrial Engineering,
The University of Iowa,
3131 Seamans Center,
Iowa City,
IA 52242-1527

Three models for detecting abnormalities of wind turbine vibrations reflected in time domain are discussed. The models were derived from the supervisory control and data acquisition (SCADA) data collected at various wind turbines. The vibration of a wind turbine is characterized by two parameters, i.e., drivetrain and tower acceleration. An unsupervised data-mining algorithm, the k-means clustering algorithm, was applied to develop the first monitoring model. The other two monitoring models for detecting abnormal values of drivetrain and tower acceleration were developed by using the concept of a control chart. SCADA vibration data sampled at 10 s intervals reflects normal and faulty status of wind turbines. The performance of the three monitoring models for detecting abnormalities of wind turbines reflected in vibration data of time domain was validated with the SCADA industrial data. [DOI: 10.1115/1.4005753]

Keywords: turbine vibration, monitoring, control chart, k-means clustering, drivetrain acceleration, tower acceleration, data-mining, neural networks ensemble

1 Introduction

The management of wind farms is challenging because it involves several difficult tasks, such as wind forecasting and the operations and maintenance of turbines. The maintenance of wind turbines has received attention in recent years due to its impact on the cost of generating power from wind. Two types of maintenance strategies are considered, preventive maintenance, and corrective maintenance [1]. The goal of preventive maintenance is to perform maintenance operations for wind turbines to prevent their failure, and the goal of corrective maintenance is to perform maintenance after a failure has occurred. Preventive maintenance can be classified further as scheduled and condition-based maintenance [1]. Numerous studies on condition-based maintenance of wind turbines have been reported in the literature. Nilsson and Bertling [2] developed a condition monitoring system to improve the maintenance efficiency based on the lifecycle cost analysis of wind turbines. Andrawus et al. [3] presented an approach integrating reliability-centered maintenance and asset lifecycle analysis techniques for selecting suitable condition-based maintenance activities for wind turbines. Hameed et al. [4] evaluated the viability of a condition monitoring system for the maintenance and performance improvement of wind turbines.

Condition-based maintenance of wind turbines relies on the monitoring of turbine parameters. The existing approaches for monitoring wind turbines fall into two categories. The first category includes analytical models based on aerodynamics and physics. Caselitz and Giebhardt [5] presented an online monitoring approach to detect faults in the performance of rotors. Yang et al. [6] utilized the wavelet transform technique to monitor electrical faults in the generator and mechanical faults in the drivetrain. Wiggelinkhuizen et al. [7] assessed the added value of various monitoring techniques to optimize the maintenance procedure for off-shore wind turbines. The second category involves soft computing tools. Kusiak et al. [8] presented an on-line monitoring model for power curves. The analytical models are explicit. However, they usually focus on labeling (detecting faults that have occurred). Although the detection of faults is useful, a preferred solution is to determine when a fault will occur. The latter

solution may reduce or even prevent damage of the wind turbine systems. Data-mining is able to cope with this challenge.

The research reported in this paper is to develop monitoring models using the vibration data collected at the wind turbine drivetrain and tower. Clustering and control chart based monitoring models are introduced. The clustering model identifies abnormal vibration based on the trained clusters and the control chart monitors the trend of vibration acceleration. Control charts have been widely researched and evaluated in the statistical quality control literature [9]. Furthermore, control charts have been applied for monitoring tasks in manufacturing, sensor calibration, and logistics [10–12]. In this research, the control chart concept was used to establish the upper and lower boundaries to detect abnormalities in the vibration of wind turbines. These two monitoring models are capable of detecting abnormalities and determining the onset of the abnormal vibration. SCADA data sampled at 10 s intervals is used in this research. Although vibration data are normally sampled at higher frequency, normal and faulty statuses of wind turbines are reflected in the 10 s data.

2 Background and Data Description

The identification of occurrences of abnormal performance of wind turbines during their operation, as reported by the SCADA system, needs additional research attention. The SCADA system collects values of parameters measured by the sensors installed at the wind turbine. Vibration is recognized as an important reference for characterizing conditions of wind turbines. In this paper, data-mining algorithms are used to develop models for monitoring the vibration of wind turbines. The 1.5 MW variable speed, pitch controlled wind turbines are used in this research.

The data utilized in this research were collected by the SCADA system of a large wind farm. Vibrations of the drivetrain and the tower of wind turbines are measured by two accelerometers and stored in the SCADA system. Thus, to monitor conditions of wind turbines, two models are needed. One for monitoring drivetrain acceleration and the other for tower acceleration. In this paper, the drive train acceleration is measured by the accelerometer installed at the bottom back of the nacelle and attached with the drivetrain. The direction of drivetrain acceleration is measured transverse to the drivetrain. The tower acceleration is measured by the accelerometer mounted near the connection of the nacelle and the tower. The direction of tower acceleration is along the wind direction.

Contributed by the Solar Energy Division of ASME for publication in the JOURNAL OF SOLAR ENERGY ENGINEERING. Manuscript received January 24, 2011; final manuscript received December 6, 2011; published online February 27, 2012. Assoc. Editor: Spyros Voutsinas.

Table 1 Data set description

Dataset	Start time stamp	End time stamp	Number of observations	Sampling rate (Hz)
Training	Oct. 23, 2009 8:00:50 AM	Oct. 26, 2009 5:58:50 PM	29,800	0.1
Test	Nov. 7, 2009 6:12:00 AM	Nov. 7, 2009 8:58:40 AM	1000	0.1
Error	Nov. 5, 2009 11:00:40 PM.	Nov. 6, 2009 8:58:50 AM	3572	0.1

The SCADA data used in this research were collected at six wind turbines and the sampling interval is 10 s (the 10 s mean value of higher frequency data). SCADA system normally records data at 10 min averages. The 10 s (0.1 Hz) SCADA data have obtained by a data logger developed in the Intelligent Systems Laboratory. Wind turbine vibration analysis commonly calls for higher frequency data, 1 s or even 0.1 s; however, such data are not widely shared by the wind industry. The data collected include training, test, and error data. The training dataset contains 29, 800 data points collected between 8:00:50 AM on Oct. 23, 2009 and 5:58:50 PM on Oct. 26, 2009. The test dataset contained 1000 data points that were acquired between 6:12:00 AM on Nov. 7, 2009 and 8:58:40 AM on Nov. 7, 2009. During the two time periods, all six wind turbines (labeled as turbines 1–6) operated in normal conditions. The error dataset reflects the faulty status of wind turbine 1. For the first fault, several abnormal occurrences were reported, such as pitch malfunction, diverter malfunction, and pitch controller time out. For the second fault, the occurrence of tower vibration was reported in addition to the abnormal occurrences that were reported in the first period. The error dataset contained 3572 data points collected in the period between 11:00:40 PM on Nov. 5, 2009 and 8:58:50 AM on Nov. 6, 2009. Table 1 presents the description of the training, test, and the error datasets.

The accelerometers are sensitive to noise, so Daubechies' (Daub) [13] wavelet was applied to denoise the turbine acceleration data. The Fix_Hard [14,15] threshold scheme of the Daub 5 wavelet with five levels was used to denoise the drivetrain and tower acceleration.

In this study, the drivetrain and tower acceleration are analyzed independently. Although the drivetrain and tower acceleration are coupled, such relationship is not statistically significant based on the collected data. Sensitivity analysis [16] is performed with the drivetrain acceleration considered as an output first. The inputs are the tower acceleration and parameters impacting the drivetrain acceleration, such as generator torque, wind speed, and blade pitch angle. The coefficients assigned to the generator torque, wind speed, and blade pitch angle are much higher than the tower acceleration coefficient. Next, the tower acceleration is treated as output and drivetrain acceleration is treated as one of its inputs. The coefficients of all inputs are similar. This indicates that the generator torque, wind speed, and blade pitch angle are statistically more significant than the coupling relationship between the drivetrain and the tower acceleration.

3 Clustering-Based Wind Turbine Vibration Monitoring Model

3.1 The Modified k -Means Algorithm. In this section, a modified k -means clustering algorithm [17] is introduced. This algorithm was used to develop a model for monitoring the vibration of the drivetrain and tower acceleration. The k -means algorithm is an unsupervised learning algorithm that clusters data into groups by evaluating their similarity. (The Euclidean distance was used here.) To determine an appropriate value of k , clustering cost function (1) was used in a tenfold, cross-validation scheme [18,19].

$$d(k, \mathbf{x}, \mathbf{c}) = \frac{1}{n} \sum_{i=1}^k \left(\sum_{\mathbf{x}_j \in C_i} \|\mathbf{x}_j - \mathbf{c}_i\|^2 \right) \quad (1)$$

where d is the clustering cost, k is the number of clusters, n is the total number of data points contained in the dataset, \mathbf{x} is a set of observations $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_j\}$ where each observation \mathbf{x}_j is a vector of parameters, \mathbf{c} represents a set of centroids of each cluster, \mathbf{c}_i represents the centroid of cluster i , j is the index of each observation, and C_i represents cluster i .

In the k -means algorithm, the total number of data points, n , is computed by summing the number of data points in each cluster, as shown below

$$n = \sum_{i=1}^k m_i \quad (2)$$

where m is the number of data points in each cluster. Then, the clustering function can be rewritten as shown in Eq. (3), based on Eqs. (1) and (2)

$$d(k, \mathbf{x}, \mathbf{c}) = \frac{1}{\sum_{i=1}^k m_i} \sum_{i=1}^k \left(\sum_{\mathbf{x}_j \in C_i} \|\mathbf{x}_j - \mathbf{c}_i\|^2 \right) \quad (3)$$

Here, the maximum number of clusters is fixed at $k=25$, and the minimum number of clusters k is set at 2. The parameters in each observation, \mathbf{x}_j , are normalized to $[0, 1]$ before implementing the modified k -means algorithm. The k -means algorithm groups the data into clusters representing abnormal and normal patterns of turbine vibration. The steps of the modified k -means algorithm are listed next:

Repeat until the criterion $d(k, \mathbf{x}, \mathbf{c}) - d(k-1, \mathbf{x}, \mathbf{c}) \leq \xi$ is satisfied.

- (1) Increment the value of k by 1, then set the initial value of k to 2 and the maximum value of k to 25.
- (2) Decompose the dataset into ten subsets of equal size.
- (3) Repeat ten times.
 - (i) Randomly select nine subsets for training and use the tenth subset for testing.
 - (ii) Initialize k centroids.
 - (iii) Repeat the following two steps until the centroids do not change:
 - Assign data point to the closest cluster by $C_i^t = \{\mathbf{x}_j : \|\mathbf{x}_j - \mathbf{c}_i^t\| \leq \|\mathbf{x}_j - \mathbf{c}_{i^*}^t\|, i^* = 1, 2, \dots, k\}$.
 - Update the values of the centroids by $\mathbf{c}_i = \sum_{\mathbf{x}_j \in C_i} \mathbf{x}_j / n$
 - (iv) Compute the clustering cost, d .
- (4) Estimate the average of clustering cost d in tenfold cross-validation, where the threshold, ξ , is arbitrarily set to 0.001.

3.2 Monitoring Drivetrain Acceleration: A Case Study. The error dataset considered in this research contains data on both normal and abnormal occurrences for wind turbines. Error log data, snapshots of faults, is used here to validate whether the k -means clustering algorithm can identify the normal and abnormal occurrences effectively. The error logs report status codes, time, and severity of the recorded abnormal component behavior. A vector of two parameters, wind speed and drivetrain acceleration, is used in developing the k -means clustering model. The evaluation result of value of k is demonstrated by Fig. 1. As shown in Fig. 1, the cost of clustering starts to converge from

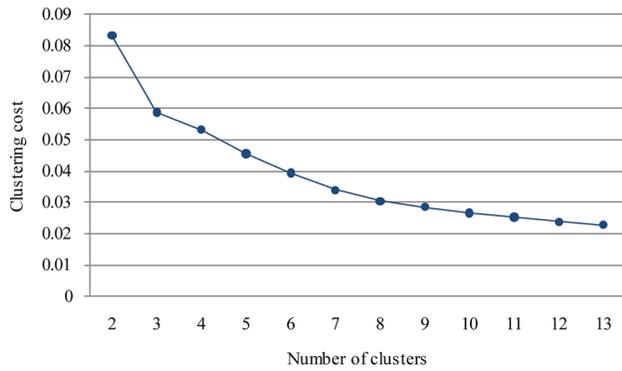


Fig. 1 Evaluation of the number of clusters k in monitoring drivetrain acceleration

$k = 12$, so $k = 12$ is selected as the number of clusters for the k -means algorithm.

Table 2 summarizes the clustering results produced by the k -means algorithm. Clusters are indexed from 1 to 12. The cluster number, centroid values of the clusters, number of points in each cluster, and the percentage of data points in each cluster are reported in Table 2.

As shown in Table 2, an abnormal status of drivetrain acceleration can be determined based on the corresponding wind speed. In Table 2, cluster 3 is associated with a fault status since the corresponding wind speed is about 10 m/s and the drivetrain acceleration is surprisingly high (around 234). Cluster 3 contains data points that present a period of malfunction of the blade pitch indicated in the status report from the wind farm. Clusters 4, 5, 7, 8, and 12 point to a period that wind turbine is shut down after the fault described in cluster 3 has occurred. In this period, the drivetrain acceleration is almost zero and the generator torque is mostly zero, while the wind speed is higher than 9 m/s. If the generated power is positive, clusters 8 and 12 might describe faulty data transmission of the accelerometer although this scenario does not appear in the dataset. The remaining six clusters, i.e., 1, 2, 6, 9, 10, and 11, represent the normal vibration status of the drivetrain system. Among the six clusters, cluster 6 represents a scenario that wind speed and vibration are both close to 0 as wind turbine is shut down.

Figure 2 shows the clustering that resulted from the use of the k -means algorithm. The horizontal axis presents the values of drivetrain acceleration, and the vertical axis shows the values of wind speed. The distributions of the 12 clusters are circled and tagged in Fig. 2. The scatter plot in Fig. 3 shows the relationship between the drivetrain acceleration and the wind speed of turbine

Table 2 Summary of clustering results from monitoring drivetrain acceleration

Cluster number	c_1 [Drivetrain acceleration (mm/s ²)]	c_2 [Wind speed (m/s)]	Generator torque (Nm)	Number of points	Percentage (%)
1	71.96	9.98	75.06	313	8.76
2	65.84	9.42	61.08	295	8.25
3	233.92	9.58	41.36	96	2.69
4	17.42	7.13	1.11	240	6.71
5	3.37	8.99	0	437	12.22
6	0.37	0.40	0	217	6.07
7	18.14	8.10	0	410	11.47
8	0.77	10.57	0	419	11.72
9	62.05	8.81	51.46	283	7.92
10	81.75	10.68	83.12	181	5.06
11	83.81	8.11	56.12	101	2.83
12	0.93	9.79	0	583	16.31

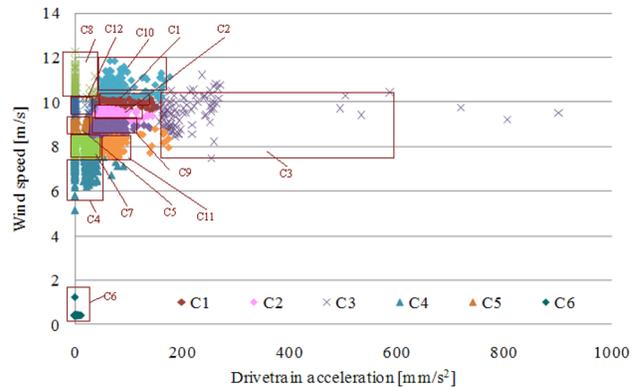


Fig. 2 Clustering results from monitoring drivetrain acceleration

2, which was considered to be operating under normal conditions in the same time period.

3.3 Case Study of Monitoring Tower Acceleration. The same error data described in Sec. 3.2 are used in this section. The tower acceleration and the wind speed are utilized here to develop the clustering model for monitoring tower acceleration. The same scheme used in Sec. 3.2 to evaluate the value of k is also applied here to determine the appropriate number of clusters. Figure 4 addresses the result of the evaluation of k . As presented in Fig. 4, the cost of clustering converges at $k = 10$ and thus the number of initial centroids for this k -means algorithm is 10.

The results produced by the k -means algorithm are summarized in Table 3. In Table 3, clusters 5, 6, and 9 reflect erroneous status

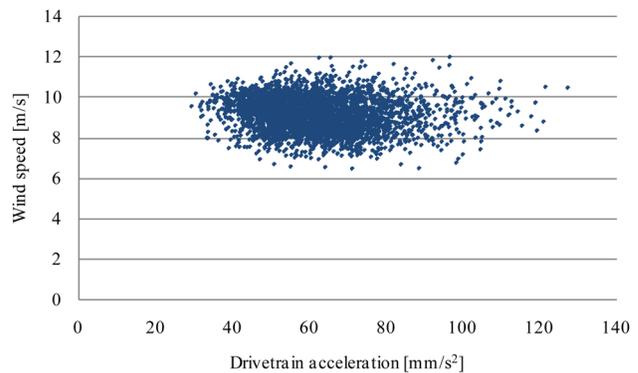


Fig. 3 Relationship between values of wind speed and drivetrain acceleration at normal conditions

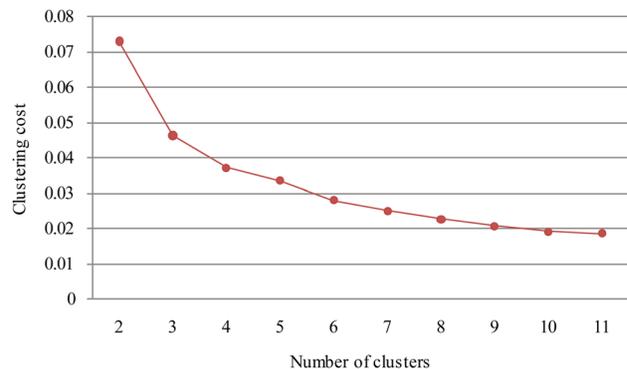


Fig. 4 Evaluation of k in monitoring tower acceleration

Table 3 Summary of clustering results in monitoring tower acceleration

Cluster number	c_1 [Tower acceleration (mm/s ²)]	c_2 [Wind speed(m/s)]	Generator torque (Nm)	Number of points	Percentage (%)
1	53.50	9.21	26.57	594	16.62
2	58.81	9.72	31.50	647	18.10
3	46.25	7.94	9.60	406	11.36
4	46.06	7.06	1.33	198	5.54
5	426.08	9.86	34.61	70	1.96
6	-407.00	0.40	0	217	6.07
7	58.08	8.64	22.40	520	14.55
8	62.39	10.82	31.38	314	8.78
9	1357.84	9.20	2.73	21	0.59
10	54.99	10.22	28.43	588	16.45

conditions of tower acceleration. The data points in clusters 5 and 9 reveal extremely high tower acceleration, and they match the record of a faulty period in the error report. Cluster 6 reflects a failure in data collection. In this cluster, all wind speed and tower acceleration values are constant for a period of time.

Clusters 3 and 4 in Table 3 raise a concern because both include faulty values. This can be clearly observed in Fig. 5. The corresponding clusters are labeled and framed in Fig. 5. Figure 6 presents a scatter plot that shows the relationship between tower acceleration and wind speed of turbine 2, which remains in normal status. The remaining clusters in Table 3 represent the normal status of tower acceleration.

3.4 Monitoring Procedure of Clustering-Based Model. The idea of implementing this clustering-based monitoring model is to identify the normal and abnormal status of wind turbine vibration. The procedure of implementing this model is addressed as the following steps:

- Step 1: Obtain new observed data point \mathbf{x} , $\mathbf{x} = [x_1, x_2]^T$, where x_1 is the value of wind speed and x_2 is the value of drivetrain acceleration or tower acceleration.
- Step 2: Calculate the Euclidean distance between the new data and the centroids of all clusters. Label the cluster index to these new data if its distance is the shortest, using $C_i = \{\mathbf{x} : \|\mathbf{x} - \mathbf{c}_i\| \leq \|\mathbf{x} - \mathbf{c}_{i'}\|, i^* = 1, 2, \dots, k\}$.
- Step 3: Make turbine maintenance decisions. If the data belong to a cluster representing the error status, a warning message is created and appropriate wind turbine maintenance actions are taken. If the data belong to a cluster indicating an undetermined status of the wind turbine, randomly choose some of the occurrences and perform the diagnosis analysis. If the data belong to a cluster

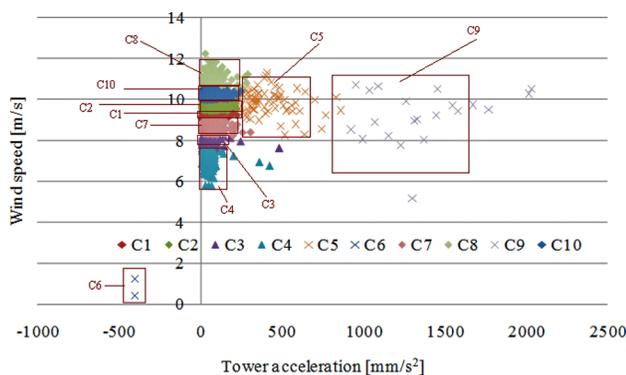


Fig. 5 Visualization of clustering result of monitoring tower acceleration

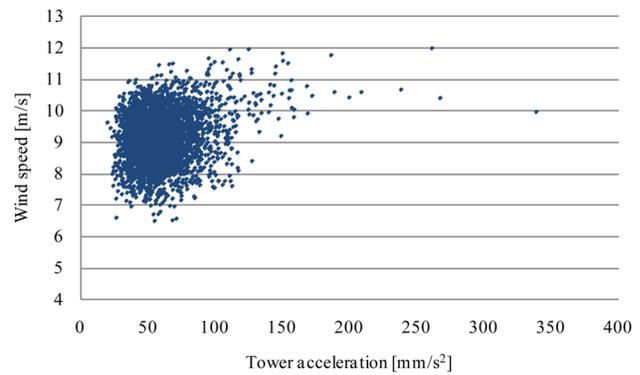


Fig. 6 Relationship between normal values of wind speed and tower acceleration

representing the normal status, it is not necessary to take any action.

Although clustering is able to identify normal and abnormal status of wind turbine vibration in general, integration with other models addressing the boundary of the clusters is needed. Moreover, since the time factor is not considered in clustering, the relationship between acceleration and time cannot be observed. Without observing the acceleration value in time (a trend), it is difficult to determine the future condition of a wind turbine. Therefore, a model that is capable to monitor the acceleration trend needs to be established.

4 Control-Chart-Based Wind Turbine Vibration Monitoring Model

The concept of a control chart from quality control [20] is applied to identify abnormal occurrences for wind turbines by monitoring the acceleration trend. The SCADA collected data are used to develop models for prediction of the drivetrain and tower acceleration. The control chart is constructed based on an accurate model for vibration prediction. The upper and lower bounds serve as the threshold for detecting abnormal vibration status of a wind turbine indicated by the points falling outside the two bounds.

4.1 Drivetrain Acceleration Baseline Model. The training and test dataset of wind turbine 1 described in Sec. 2 are utilized to develop the baseline model accurately predicting its drivetrain acceleration. The test data of wind turbine 1 are used for testing the prediction accuracy of the baseline model. The SCADA system collects data on more than 120 parameters. However, in this research, only meaningful parameters that are potentially related to drivetrain acceleration are selected based on domain knowledge and the literature. The parameter selection aims at reducing the dimensionality of the data and simplifying the model.

4.1.1 Parameter Selection. The SCADA parameters impacting the drivetrain acceleration, such as wind speed, blade pitch angle, generator torque, and wind deviation, are considered in prediction of the drivetrain acceleration. The past values of these parameters and the drivetrain acceleration impact the current drivetrain acceleration and they are considered in the development of the prediction model.

Definition 4.1. Assume that the current time is t and that the data sampling time is T ; then, the time index of parameters observed at the current time will be expressed as t , and the time index of parameters observed n steps backward can be expressed as $t - nT$. For example, one step backward will be $t - T$, and two steps backward will be $t - 2T$, and so on.

Usually, not all past states of parameters have significant impact on the current drivetrain acceleration. In addition, the

Table 4 Pool of parameters before and after selection for prediction of drivetrain acceleration

Parameter pool				Selected parameters			
$v[t]$	Wind speed at t	$\tau[t - 3T]$	Generator torque at time $t - 3T$	$v[t]$	Wind speed at t	$d[t - T]$	Wind deviation at time $t - T$
$v[t - T]$	Wind speed at $t - T$	$d[t]$	Wind deviation at time t	$v[t - 3T]$	Wind speed at $t - 3T$	$d[t - 3T]$	Wind deviation at time $t - 3T$
$v[t - 2T]$	Wind speed at $t - 2T$	$d[t - T]$	Wind deviation at time $t - T$	$\beta[t]$	Blade pitch angle at time t	$A_d[t - T]$	Drivetrain acceleration at time $t - T$
$v[t - 3T]$	Wind speed at $t - 3T$	$d[t - 2T]$	Wind deviation at time $t - 2T$	$\beta[t - T]$	Blade pitch angle at time $t - T$	$A_d[t - 2T]$	Drivetrain acceleration at time $t - 2T$
$\beta[t]$	Blade pitch angle at time t	$d[t - 3T]$	Wind deviation at time $t - 3T$	$\beta[t - 2T]$	Blade pitch angle at time $t - 2T$	$A_d[t - 3T]$	Drivetrain acceleration at time $t - 3T$
$\beta[t - T]$	Blade pitch angle at time $t - T$	$A_d[t - T]$	Drivetrain acceleration at time $t - T$	$\beta[t - 3T]$	Blade pitch angle at time $t - 3T$	—	—
$\beta[t - 2T]$	Blade pitch angle at time $t - 2T$	$A_d[t - 2T]$	Drivetrain acceleration at time $t - 2T$	$\tau[t]$	Generator torque at time t	—	—
$\beta[t - 3T]$	Blade pitch angle at time $t - 3T$	$A_d[t - 3T]$	Drivetrain acceleration at time $t - 3T$	$\tau[t - T]$	Generator torque at time $t - T$	—	—
$\tau[t]$	Generator torque at time t	—	—	$\tau[t - 2T]$	Generator torque at time $t - 2T$	—	—
$\tau[t - T]$	Generator torque at time $t - T$	—	—	$\tau[t - 3T]$	Generator torque at time $t - 3T$	—	—
$\tau[t - 2T]$	Generator torque at time $t - 2T$	—	—	$d[t]$	Wind deviation at time t	—	—

importance of the parameters may vary due to the data sampling interval T . Therefore, a parameter selection procedure is needed to identify most important parameters for developing data-driven models. In this section, a wrapper method [21,22] with a genetic search has been applied for parameter selection. Table 4 illustrates the parameter pool before the parameter selection and the selected parameters based on the wrapper approach.

Using the selected parameters, the drivetrain acceleration model is expressed in Eq. (4)

$$A_d[t] = f(v[t], v[t - 3T], \beta[t], \beta[t - T], \beta[t - 2T], \beta[t - 3T], \tau[t], \tau[t - T], \tau[t - 2T], \tau[t - 3T], d[t], d[t - T], d[t - 3T], A_d[t - T], A_d[t - 2T], A_d[t - 3T]) \quad (4)$$

The notation used here is described in Table 4.

4.1.2 Algorithm Comparison. To construct the model in Eq. (4), seven different data-mining algorithms have been used, namely, neural network ensemble (NNE) [23], neural network (NN) [24–26], boosting regression tree (BT) [27,28], support vector machine (SVM) [29,30], random forest with regression (RF) [31], standard classification and regression tree (CART) [32], and k nearest neighbor neural network (k NN) [33]. Four metrics, the mean absolute error (MAE), standard deviation of absolute error (SD of AE), mean square error (MSE), and the standard deviation of square error (SD of SE), are utilized to evaluate the performance of data-mining algorithms in model extraction. The four metrics are defined in Eqs. (5)–(8), below

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5)$$

$$SD \text{ of AE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(|\hat{y}_i - y_i| - \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \right)^2} \quad (6)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (7)$$

$$SD \text{ of SE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[(\hat{y}_i - y_i)^2 - \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \right]^2} \quad (8)$$

where n is the total number of data points in the dataset, \hat{y} is the predicted value, and y is the observed value.

Table 5 presents the test results for models developed by the seven data-mining algorithms. The NNE model provided the lowest values for all four metrics. Thus, the NNE model is recognized as the most suitable for determining drivetrain acceleration. To quantify the accuracy of the NNE model, the mean absolute percentage error (MAPE) metric in Eq. (9) is used. The MAPE of the NNE model is 0.09 which corresponds to the model accuracy of 91%.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left(\left| \frac{\hat{y}_i - y_i}{y_i} \right| \right) \times 100\% \quad (9)$$

Figure 7 illustrates the results of the first 100 points of the test dataset based on the NNE model. In general, the predicted values follow the observed ones. However, due to the measurement errors and information loss in data sampling the predicted values fluctuate around the observed data.

4.2 Baseline Model for Tower Acceleration. In order to establish a control chart for monitoring the vibration of a turbine wind tower, a baseline model is introduced in this section. A data-driven approach is used to extract this model based on the data collected for turbines presented in Sec. 4.1.

Table 5 Testing performance of data-mining algorithms

Algorithm	MAE	SD of AE	MSE	SD of SE
NNE	5.47	5.16	56.48	128.03
NN	5.51	5.44	59.94	152.61
BT	11.36	10.83	246.22	789.11
SVM	11.24	8.41	196.89	444.24
RF	12.43	17.39	456.50	2074.08
CART	17.08	15.34	526.79	1695.82
k NN	8.47	9.35	159.02	580.23

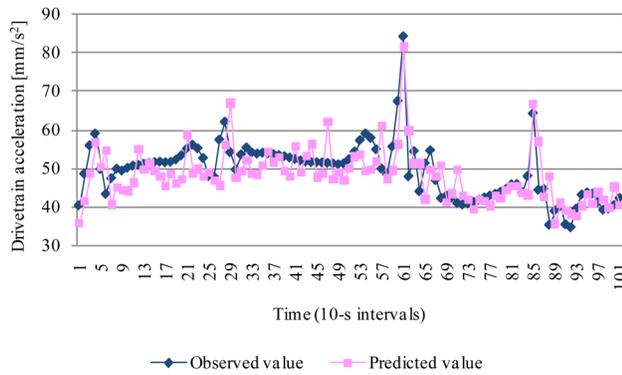


Fig. 7 Predicted and observed values of the drivetrain acceleration for the first 100 test points

4.2.1 Parameter Selection. Similar to Sec. 4.1.1, parameter selection is utilized to determine the parameters that impact tower acceleration. Table 6 presents a pool of parameters used in the selection process and the parameters selected with the wrapper approach introduced in Sec. 4.1.1.

The tower acceleration model is expressed in Eq. (10), below

$$A_t[t] = f(v[t], v[t - T], v[t - 2T], v[t - 3T], \beta[t], \beta[t - T], \beta[t - 3T], \tau[t], \tau[t - T], \tau[t - 2T], \tau[t - 3T], d[t - 3T], A_t[t - T]) \quad (10)$$

4.2.2 Comparison of Algorithms. The seven data-mining algorithms of Sec. 4.1.2 have been applied to develop a model for predicting tower acceleration. The test data of Sec. 2 are used to test the accuracy of the developed models. The four metrics introduced in Sec. 4.1.2 are used to evaluate the performance of these models. Table 7 summarizes the test results of the models learned by the seven data-mining algorithms, with the NNE model producing the smallest MAE, and the k NN model producing the smallest SD of MSE. However, in general, the NN model provides the best performance (with MAE values close to the MAE of NNE) and the values of the other metrics better than NNE. In contrast to the k NN model, the values of the remaining three metrics of NN outweigh those of the k NN model, although the SD of SE

Table 7 Test performance of models derived by seven data-mining algorithms

Algorithm	MAE	SD of AE	MSE	SD of SE
NNE	6.53	13.23	217.42	1663.38
NN	6.72	11.13	168.89	1070.69
BT	13.38	24.81	793.74	7824.97
SVM	59.10	19.30	3864.54	2223.74
RF	23.40	27.36	1295.39	8037.39
CART	15.92	26.68	964.73	8221.05
k NN	15.71	14.13	446.21	907.52

of NN is slightly lower than the MSE of k NN. Thus, the NN model is considered to be the most suitable algorithm for developing the model for predicting tower acceleration. The accuracy of the NN model is 90% (MAPE = 0.1).

Figure 8 demonstrates the results of the prediction of first 100 points in the test dataset based on the model trained by an NN algorithm. The predicted tower acceleration deviates slightly from the observed data. The prediction accuracy can be further improved by reducing the measurement errors and increasing the data sampling frequency.

4.3 Monitoring Wind Turbine Vibration Based on Control Charts. The baseline models discussed in Secs. 4.1 and 4.2 have been shown to be accurate enough to predict drivetrain and tower acceleration. Since the training dataset does not contain abnormal data points of wind turbine vibration, the prediction results offered by the baseline models are considered as trustworthy references of the normal status of vibration of the wind turbine. Although the baseline models can provide accurate prediction results, residuals between the predicted value and the observed values of vibration still exist. Thus, monitoring abnormal occurrences of vibration translates into monitoring abnormal values of residuals. The control-chart concept of quality control is useful in monitoring the residuals and the variations of residuals.

Definition 4.2. The residual between the value \hat{y} predicted by the baseline model and the observed value y is defined as $\varepsilon = \hat{y} - y$, where \hat{y} can be computed based on Eqs. (4) and (10).

Definition 4.3. To monitor the residuals, the center line of the control chart is defined as the mean of ε obtained from the

Table 6 Pool of parameters before and after selection for predicting tower acceleration

Parameter pool				Selected parameters			
$v[t]$	Wind speed at t	$\tau[t - 3T]$	Generator torque at time $t - 3T$	$v[t]$	Wind speed at t	$d[t - 3T]$	Wind deviation at time $t - 3T$
$v[t - T]$	Wind speed at $t - T$	$d[t]$	Wind deviation at time t	$v[t - T]$	Wind speed at $t - T$	$A_t[t - T]$	Tower acceleration at time $t - T$
$v[t - 2T]$	Wind speed at $t - 2T$	$d[t - T]$	Wind deviation at time $t - T$	$v[t - 2T]$	Wind speed at $t - 2T$	—	—
$v[t - 3T]$	Wind speed at $t - 3T$	$d[t - 2T]$	Wind deviation at time $t - 2T$	$v[t - 3T]$	Wind speed at $t - 3T$	—	—
$\beta[t]$	Blade pitch angle at time t	$d[t - 3T]$	Wind deviation at time $t - 3T$	$\beta[t]$	Blade pitch angle at time t	—	—
$\beta[t - T]$	Blade pitch angle at time $t - T$	$A_t[t - T]$	Tower acceleration at time $t - T$	$\beta[t - T]$	Blade pitch angle at time $t - T$	—	—
$\beta[t - 2T]$	Blade pitch angle at time $t - 2T$	$A_t[t - 2T]$	Tower acceleration at time $t - 2T$	$\beta[t - 3T]$	Blade pitch angle at time $t - 3T$	—	—
$\beta[t - 3T]$	Blade pitch angle at time $t - 3T$	$A_t[t - 3T]$	Tower acceleration at time $t - 3T$	$\tau[t]$	Generator torque at time t	—	—
$\tau[t]$	Generator torque at time t	—	—	$\tau[t - T]$	Generator torque at time $t - T$	—	—
$\tau[t - T]$	Generator torque at time $t - T$	—	—	$\tau[t - 2T]$	Generator torque at time $t - 2T$	—	—
$\tau[t - 2T]$	Generator torque at time $t - 2T$	—	—	$\tau[t - 3T]$	Generator torque at time $t - 3T$	—	—

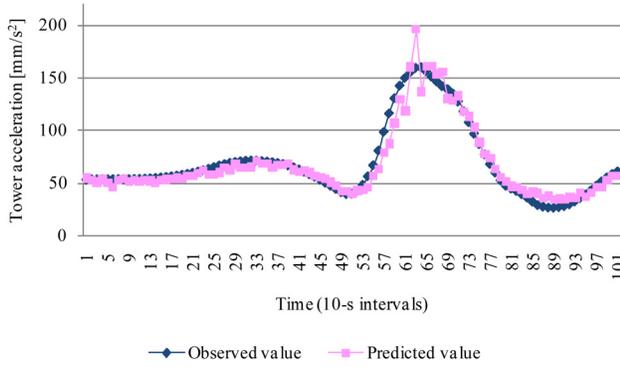


Fig. 8 Predicted and observed values of tower acceleration for the first 100 test points

training baseline model. The mean of ε is expressed as μ_t , and the calculation of this value can be formulated in Eq. (11). To monitor the variation of the residual, the center line of the control chart is defined as the variation of ε for the training of baseline model. This variation is represented by s_t expressed in Eq. (12).

$$\mu_t = \left[\sum_{i=1}^n (\hat{y}_i - y_i) \right] / n \quad (11)$$

$$s_t = \sqrt{\left[\sum_{i=1}^n ((\hat{y}_i - y_i) - \mu_t)^2 \right] / (n - 1)} \quad (12)$$

where n is the number of data points in the training dataset.

The upper and lower bounds of the control-chart monitoring residuals are constructed based on the center line, and this control chart is expressed as

$$UCL_r = \mu_t + c(s_t/\sqrt{m}) \quad (13)$$

$$CL_r = \mu_t \quad (14)$$

$$LCL_r = \mu_t - c(s_t/\sqrt{m}) \quad (15)$$

where UCL_r , CL_r , and LCL_r represent the upper control limit, center line, and the lower control limit of residuals, m is the number of data points sampled for monitoring, c is the parameter that controls the sensitivity of the control chart, and the remaining notation is the same as in Definition 4.3.

To monitor the variation of residuals, the upper and lower bounds and the center line are expressed as

$$UCL_v = \frac{s_t^2}{m-1} \chi_{\alpha/2, m-1}^2 \quad (16)$$

$$CL_v = s_t^2 \quad (17)$$

$$LCL_v = 0 \quad (18)$$

where $\chi_{\alpha/2, m-1}^2$ denotes the right $\alpha/2$ percentage of the chi-square distribution, and the remaining notation is identical to that of Eqs. (13)–(15).

The error dataset of wind turbine 1 introduced in Sec. 2 is used here to address the capability of the control-chart-based monitoring model in detecting vibration abnormalities in the wind turbine. The mean and variation of the residual for m data points from the error dataset are computed from the following:

$$\mu_e = \left[\sum_{i=1}^m (\hat{y}_i - y_i) \right] / m \quad (19)$$

$$s_e = \sqrt{\left[\sum_{i=1}^m ((\hat{y}_i - y_i) - \mu_e)^2 \right] / (m - 1)} \quad (20)$$

Definition 4.4. If the values of μ_e or s_e are greater than the upper control limit or lower than the lower control limit, the values of μ_e or s_e are considered to be abnormal, and this may indicate issues with vibration of the drivetrain or the tower. If neither of these is the case, the vibration of the drivetrain or tower is considered to be normal.

To demonstrate the monitoring result in a more straightforward manner, the control chart of monitoring residuals can be modified to directly monitor the values of future observed drivetrain acceleration or tower acceleration. The modified control chart can be written as

$$UCL_s = \hat{y}_e + \mu_t + c(s_t/\sqrt{m}) \quad (21)$$

$$CL_s = \hat{y}_e + \mu_t \quad (22)$$

$$LCL_s = \hat{y}_e + \mu_t - c(s_t/\sqrt{m}) \quad (23)$$

where \hat{y}_e presents the predicted value of drivetrain acceleration or tower acceleration for the error dataset, and the other notation is the same as in Eqs. (13)–(15).

Definition 4.5. Based on this modified control chart, abnormal values of drivetrain acceleration and tower acceleration will be detected if the observed values of drivetrain acceleration or tower acceleration fall outside the upper control limit with an increasing trend and the lower control limit with a decreasing trend.

4.3.1 Results From Monitoring Drivetrain Acceleration. In this section, the modified control chart stated in Eqs. (21)–(23) is applied to both the testing dataset and the error dataset of wind turbine 1 introduced in Sec. 2 to demonstrate the results of monitoring drivetrain acceleration. Figure 9 shows the monitoring results of the first 100 points of the test dataset. Since all the data points contained in the test dataset present a normal status of vibration in the drivetrain system, all of the observed values fall in the region between the upper bound and lower bound. Figure 10 shows the monitoring results of 100 points in the error dataset. As shown in Fig. 10, many of the points fall outside the upper and lower bound. The root causes of this abnormality are reported in the error log file. There are two occurrences of pitch and diverter malfunction in the period represented by these 100 points.

4.3.2 Results From Monitoring Tower Acceleration. The performance of the modified control chart in monitoring the tower acceleration is addressed in this section. Figure 11 shows the performance of the control chart in monitoring normal data, while Fig. 12 demonstrates the capability of the control chart in detecting abnormal tower acceleration. A fault, high tower vibration, is reported 10 s later than the pitch malfunction fault in the error log file. Therefore, it is possible that the abnormal vibration is caused by the pitch overrun.

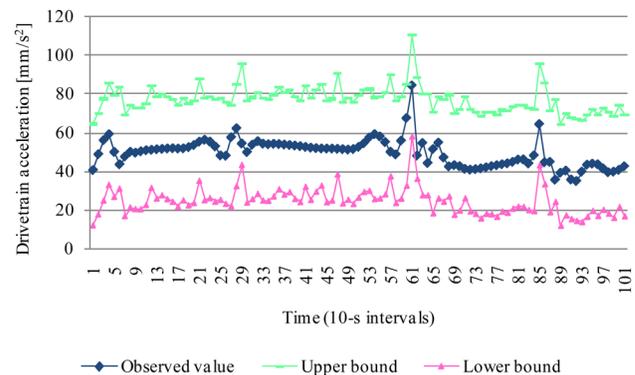


Fig. 9 The control chart for the dataset that contains normal drivetrain acceleration data

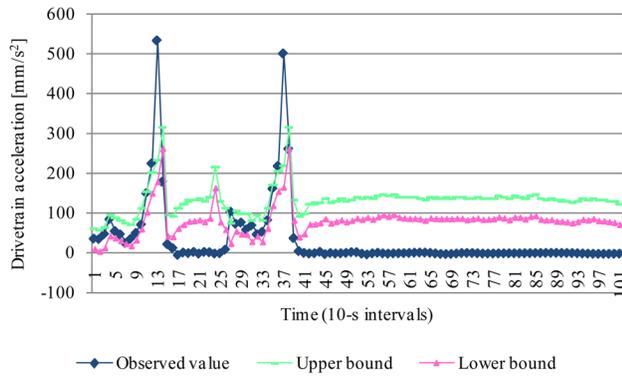


Fig. 10 The control chart for the dataset that contains some abnormal drivetrain acceleration data

5 Virtual Sensor for Monitoring the Vibration of a Wind Turbine

In Sec. 4, data-driven models for predicting drivetrain acceleration and tower acceleration were utilized as the baseline models for conducting the control chart for monitoring the vibration of the wind turbine. In order to establish the data-driven prediction models, in Sec. 4, parameters of wind turbine 1, such as wind speed and generator torque, are considered as inputs to the model that computes the output, which is drivetrain acceleration or tower acceleration of wind turbine 1. However, in the real case, it is still hazardous to establish a control chart to monitor turbine vibration just by relying on the baseline models introduced in Sec. 4. This is because there is always a certain level of risk that data transmission errors may occur in data collection due to a malfunction of the information system. In this scenario, developing the prediction model for drivetrain acceleration or tower acceleration for one wind turbine without using the parameters of the wind turbine itself becomes a challenging issue. To handle this circumstance, the concept of virtual sensor models is introduced to offer another option of predicting drivetrain acceleration and tower acceleration. The virtual sensor model has been reported in some of the previous studies [34] and [35].

The basic concept of the virtual sensor model is that the vibration of one wind turbine can be predicted based on information borrowed from other wind turbines rather than its own parameters. In this research, to demonstrate the virtual sensor concept in a simpler way, it was assumed that there were obstacles in acquiring information for the parameters of wind turbine 1. In order to predict the drivetrain acceleration and tower acceleration for this turbine, virtual sensor models were developed. Since the parameters of turbine 1 are unavailable under the assumption, the first challenging problem is to determine how meaningful information can

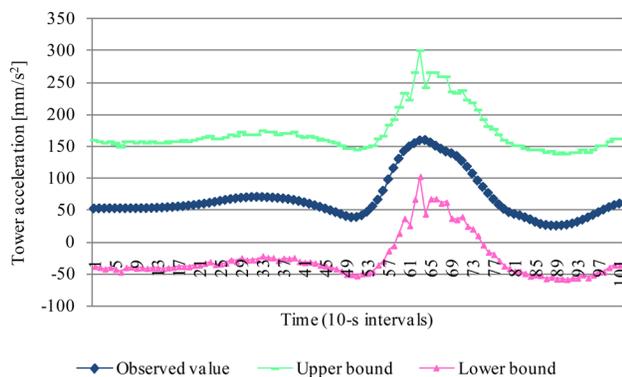


Fig. 11 The control chart for the dataset that contains normal tower acceleration data

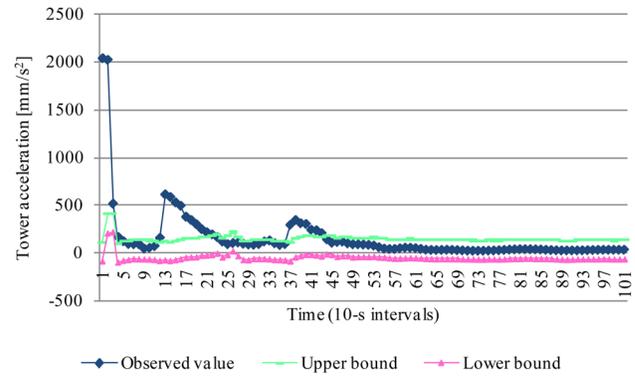


Fig. 12 The control chart for the dataset that contains some abnormal tower acceleration data

be selected from other wind turbines to assist the development of the virtual sensor models. Since wind speed is recognized as the most significant factor that contributes to the vibration of the wind turbine based on importance analysis by the boosting tree algorithm [27,28], information from wind turbines that have similar wind conditions to those of turbine 1 is considered more helpful than information from other wind turbines.

The training dataset and test dataset of all six wind turbines presented in Sec. 2 are used here to develop the virtual sensor models for turbine 1. In the six wind turbines, turbines 1, 2, and 3 were installed in the same sector, and turbines 4, 5, and 6 were installed in a different sector. The layout design of the six wind turbines is presented in Fig. 13.

In Fig. 13, the wind turbine in each sector that has the most similar wind speed conditions compared with those of turbine 1 will be selected to conduct the virtual sensor models for turbine 1.

Definition 5.1. Assume that n is the total number of data points in the training dataset, i is the index of the data points in this dataset, v is the wind speed, and j is the index of the wind turbine. Then, the difference of the wind speeds between turbine 1 and any other wind turbine, which is expressed as ϕ_j , is computed as follows:

$$\phi_j = \left(\sum_{i=1}^n \|v_1^i - v_j^i\| \right) / n, \quad \text{for } j = 2, 3, 4, \dots \quad (24)$$

The selection procedure is implemented based on the training dataset and can be described in the following two steps:

- (1) Calculate $\phi_j = \left(\sum_{i=1}^n \|v_1^i - v_j^i\| \right) / n$ for every wind turbine in each sector.

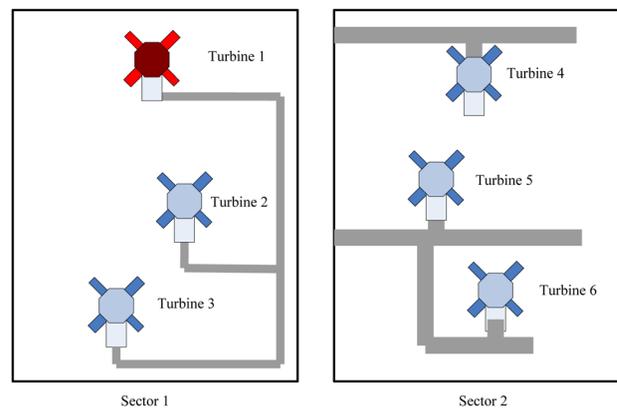


Fig. 13 Layout design of six wind turbines

Table 8 Difference of wind speed between turbine 1 and five other turbines

Sector 1		Sector 2		
ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
1.21	1.42	1.60	1.19	1.30

(2) Select the turbine with the smallest ϕ in each court to develop virtual sensor models.

Table 8 shows the results of the selection. The information of turbines 2 and 5 is considered in developing the virtual sensor models for turbine 1.

5.1 Virtual Sensor Model for Drivetrain Acceleration. To develop the virtual sensor model for estimating the drivetrain acceleration of turbine 1, the training dataset and test dataset of turbines 1, 2, and 5 are used in this section.

5.1.1 Parameter Selection. Similar to the parameter selection discussed in Sec. 4.1, various parameters, such as wind speed, wind deviation, drivetrain acceleration, and tower acceleration, for turbines 2 and 5 are considered as important factors for predicting the drivetrain acceleration of turbine 1. In addition to these parameters, the past three states of these parameters and the drivetrain acceleration of turbine 1 are also included in the prediction. The same approach for parameter selection that was discussed in

Sec. 4.1 is used here to select the most important parameters for use in developing a virtual sensor model that can predict drivetrain acceleration of turbine 1. Table 9 presents the results of the parameter selection process.

Based on the results of parameter selection, the virtual sensor model for predicting drivetrain acceleration can be expressed as

$$\begin{aligned}
 Da_1[t] = & f(Da_1[t - T], Da_1[t - 3T], v_2[t - 2T], d_2[t], d_2[t - 3T], \\
 & Ta_2[t], Ta_2[t - T], Da_2[t], Da_2[t - 2T], Da_2[t - 3T], \\
 & v_5[t - T], v_5[t - 3T], d_5[t], d_5[t - T], d_5[t - 2T], Ta_5[t], \\
 & Ta_2[t - 1], Ta_2[t - 2T], Da_5[t], Da_5[t - 3T])
 \end{aligned}
 \tag{25}$$

where the notation of this virtual sensor model is presented in Table 9.

5.1.2 Comparative Study of Algorithms. The data-mining algorithms mentioned in Sec. 4.1 are also applied here to train the virtual sensor model for measuring the drivetrain acceleration of turbine 1 based on the training dataset in Sec. 5.1.1. Then, the test dataset in Sec. 5.1.1 is used to test the performance of the data-driven models in making the predictions. Table 10 shows the test results achieved by data-driven models that have been trained by all seven data-mining algorithms.

As shown in Table 10, the NNE model provides the best test results for predicting the drivetrain acceleration of turbine 1 based on the test dataset. Therefore, it was selected as the most suitable

Table 9 Parameters before and after selection

Parameter pool				Selected parameters			
$Da_1[t - T]$	Drive train acceleration of turbine 1 at time $t - T$	$Da_2[t - 3T]$	Drivetrain acceleration of turbine 2 at time $t - 3T$	$Da_1[t - T]$	Drivetrain acceleration of turbine 1 at time $t - T$	$Da_5[t]$	Drivetrain acceleration of turbine 5 at time t
$Da_1[t - 2T]$	Drive train acceleration of turbine 1 at time $t - 2T$	$v_5[t]$	Wind speed of turbine 5 at t	$Da_1[t - 3T]$	Drivetrain acceleration of turbine 1 at time $t - 3T$	$Da_5[t - 3T]$	Drivetrain acceleration of turbine 5 at time $t - 3T$
$Da_1[t - 3T]$	Drive train acceleration of turbine 1 at time $t - 3T$	$v_5[t - T]$	Wind speed of turbine 5 at $t - T$	$v_2[t - 2T]$	Wind speed of turbine 2 at $t - 2T$	—	—
$v_2[t]$	Wind speed of turbine 2 at t	$v_5[t - 2T]$	Wind speed of turbine 5 at $t - 2T$	$d_2[t]$	Wind deviation of turbine 2 at time t	—	—
$v_2[t - T]$	Wind speed of turbine 2 at $t - T$	$v_5[t - 3T]$	Wind speed of turbine 5 at $t - 3T$	$d_2[t - 3T]$	Wind deviation of turbine 2 at time $t - 3T$	—	—
$v_2[t - 2T]$	Wind speed of turbine 2 at $t - 2T$	$d_5[t]$	Wind deviation of turbine 5 at time t	$Ta_2[t]$	Tower acceleration of turbine 2 at time t	—	—
$v_2[t - 3T]$	Wind speed of turbine 2 at $t - 3T$	$d_5[t - T]$	Wind deviation of turbine 5 at time $t - T$	$Ta_2[t - T]$	Tower acc. of turbine 2 at time $t - T$	—	—
$d_2[t]$	Wind deviation of turbine 2 at time t	$d_5[t - 2T]$	Wind deviation of turbine 5 at time $t - 2T$	$Da_2[t]$	Drivetrain acceleration of turbine 2 at time t	—	—
$d_2[t - T]$	Wind deviation of turbine 2 at time $t - T$	$d_5[t - 3T]$	Wind deviation of turbine 5 at time $t - 3T$	$Da_2[t - 2T]$	Drivetrain acceleration of turbine 2 at time $t - 2T$	—	—
$d_2[t - 2T]$	Wind deviation of turbine 2 at time $t - 2T$	$Ta_5[t]$	Tower acceleration of turbine 5 at time t	$Da_2[t - 3T]$	Drivetrain acceleration of turbine 2 at time $t - 3T$	—	—
$d_2[t - 3T]$	Wind deviation of turbine 2 at time $t - 3T$	$Ta_5[t - T]$	Tower acceleration of turbine 5 at time $t - T$	$v_5[t - T]$	Wind speed of turbine 5 at $t - T$	—	—
$Ta_2[t]$	Tower acceleration of turbine 2 at time t	$Ta_5[t - 2T]$	Tower acceleration of turbine 5 at time $t - 2T$	$v_5[t - 3T]$	Wind speed of turbine 5 at $t - 3T$	—	—
$Ta_2[t - T]$	Tower acceleration of turbine 2 at time $t - T$	$Ta_5[t - 3T]$	Tower acceleration of turbine 5 at time $t - 3T$	$d_5[t]$	Wind deviation of turbine 5 at time t	—	—
$Ta_2[t - 2T]$	Tower acceleration of turbine 2 at time $t - 2T$	$Da_5[t]$	Drivetrain acceleration of turbine 5 at time t	$d_5[t - T]$	Wind deviation of turbine 5 at time $t - T$	—	—
$Ta_2[t - 3T]$	Tower acceleration of turbine 2 at time $t - 3T$	$Da_5[t - T]$	Drivetrain acceleration of turbine 5 at time $t - T$	$d_5[t - 2T]$	Wind deviation of turbine 5 at time $t - 2T$	—	—
$Da_2[t]$	Drivetrain acceleration of turbine 2 at time t	$Da_5[t - 2T]$	Drivetrain acceleration of turbine 5 at time $t - 2T$	$Ta_5[t]$	Tower acceleration of turbine 5 at time t	—	—
$Da_2[t - T]$	Drivetrain acceleration of turbine 2 at time $t - T$	$Da_5[t - 3T]$	Drivetrain acceleration of turbine 5 at time $t - 3T$	$Ta_5[t - T]$	Tower acceleration of turbine 5 at time $t - T$	—	—
$Da_2[t - 2T]$	Drivetrain acceleration of turbine 2 at time $t - 2T$	—	—	$Ta_5[t - 2T]$	Tower acceleration of turbine 5 at time $t - 2T$	—	—

Table 10 Test performance of data-mining algorithms

Algorithm	MAE	SD of AE	MSE	SD of SE
NNE	7.60	11.01	178.95	786.51
NN	8.07	10.99	185.83	797.65
BT	11.38	11.05	251.52	809.70
SVM	16.85	10.93	403.10	672.54
RF	11.19	18.48	466.19	2219.06
CART	17.08	15.34	526.79	1695.82
kNN	12.79	13.98	358.75	1134.84

algorithm for use in designing the virtual sensor model. The MAPE of NNE is 0.13. Figure 14 shows the accuracy of the prediction based on the model that was based on the NNE algorithm. The test results for the first 100 points in the test dataset are included in Fig. 10.

5.2 Virtual Sensor Model for Tower Acceleration. The datasets described in Sec. 5.1 are used here to build the virtual sensor model for predicting tower acceleration.

5.2.1 Parameter Selection. The same procedure used for parameter selection in Sec. 5.1.1 is used here to select the important parameters for establishing the virtual sensor model. The initial parameter pool is similar to that of Sec. 5.1.1, except that the drivetrain acceleration of turbine 1 has been replaced by the tower acceleration. Table 11 shows the results the two pools of parameters.

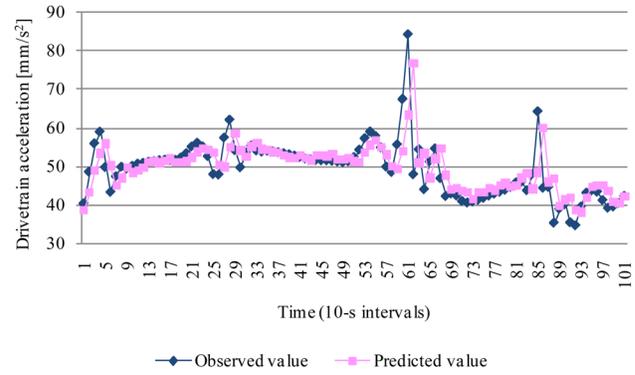


Fig. 14 Predicted and observed values of drivetrain acceleration for the first 100 test points

Then, the formulation of the virtual sensor model for predicting tower acceleration of turbine 1 can be written as

$$\begin{aligned}
 Ta_1[t] = f(Ta_1[t - T], Ta_1[t - 2T], Ta_1[t - 3T], \\
 v_2[t], v_2[t - 3], d_2[t], d_2[t - 3T], \\
 Ta_2[t], Ta_2[t - 2T], Da_2[t - T], \\
 v_5[t - 2T], v_5[t - 3T], d_5[t - T], \\
 Ta_5[t - T], Ta_2[t - 2T], Ta_2[t - 3T], \\
 Da_5[t], Da_5[t - T], Da_5[t - 3T])
 \end{aligned}
 \tag{26}$$

Table 11 Parameters before and after selection

Parameter pool				Selected parameters			
$Ta_1[t - T]$	Tower acceleration of turbine 1 at time $t - T$	$Da_2[t - 3T]$	Drivetrain acceleration of turbine 2 at time $t - 3T$	$Ta_1[t - T]$	Tower acceleration of turbine 1 at time $t - T$	$Da_5[t - 3T]$	Drivetrain acceleration of turbine 5 at time $t - 3T$
$Ta_1[t - 2T]$	Tower acceleration of turbine 1 at time $t - 2T$	$v_5[t]$	Wind speed of turbine 5 at t	$Ta_1[t - 2T]$	Tower acceleration of turbine 1 at time $t - 2T$	—	—
$Ta_1[t - 3T]$	Tower acceleration of turbine 1 at time $t - 3T$	$v_5[t - T]$	Wind speed of turbine 5 at $t - T$	$Ta_1[t - 3T]$	Tower acceleration of turbine 1 at time $t - 3T$	—	—
$v_2[t]$	Wind speed of turbine 2 at t	$v_5[t - 2T]$	Wind speed of turbine 5 at $t - 2T$	$v_2[t]$	Wind speed of turbine 2 at t	—	—
$v_2[t - T]$	Wind speed of turbine 2 at $t - T$	$v_5[t - 3T]$	Wind speed of turbine 5 at $t - 3T$	$v_2[t - 3T]$	Wind speed of turbine 2 at $t - 3T$	—	—
$v_2[t - 2T]$	Wind speed of turbine 2 at $t - 2T$	$d_5[t]$	Wind deviation of turbine 5 at time t	$d_2[t]$	Wind deviation of turbine 2 at time t	—	—
$v_2[t - 3T]$	Wind speed of turbine 2 at $t - 3T$	$d_5[t - T]$	Wind deviation of turbine 5 at time $t - T$	$d_2[t - 3T]$	Wind deviation of turbine 2 at time $t - 3T$	—	—
$d_2[t]$	Wind deviation of turbine 2 at time t	$d_5[t - 2T]$	Wind deviation of turbine 5 at time $t - 2T$	$Ta_2[t]$	Tower acceleration of turbine 2 at time t	—	—
$d_2[t - T]$	Wind deviation of turbine 2 at time $t - T$	$d_5[t - 3T]$	Wind deviation of turbine 5 at time $t - 3T$	$Ta_2[t - 2T]$	Tower acceleration of turbine 2 at time $t - 2T$	—	—
$d_2[t - 2T]$	Wind deviation of turbine 2 at time $t - 2T$	$Ta_5[t]$	Tower acceleration of turbine 5 at time t	$Da_2[t - T]$	Drivetrain acceleration of turbine 2 at time $t - T$	—	—
$d_2[t - 3T]$	Wind deviation of turbine 2 at time $t - 3T$	$Ta_5[t - T]$	Tower acceleration of turbine 5 at time $t - T$	$v_5[t - 2T]$	Wind speed of turbine 5 at $t - 2T$	—	—
$Ta_2[t]$	Tower acceleration of turbine 2 at time t	$Ta_5[t - 2T]$	Tower acceleration of turbine 5 at time $t - 2T$	$v_5[t - 3T]$	Wind speed of turbine 5 at $t - 3T$	—	—
$Ta_2[t - T]$	Tower acceleration of turbine 2 at time $t - T$	$Ta_5[t - 3T]$	Tower acceleration of turbine 5 at time $t - 3T$	$d_5[t - T]$	Wind deviation of turbine 5 at time $t - T$	—	—
$Ta_2[t - 2T]$	Tower acceleration of turbine 2 at time $t - 2T$	$Da_5[t]$	Drivetrain acceleration of turbine 5 at time t	$Ta_5[t - T]$	Tower acceleration of turbine 5 at time $t - T$	—	—
$Ta_2[t - 3T]$	Tower acceleration of turbine 2 at time $t - 3T$	$Da_5[t - T]$	Drivetrain acceleration of turbine 5 at time $t - T$	$Ta_5[t - 2T]$	Tower acceleration of turbine 5 at time $t - 2T$	—	—
$Da_2[t]$	Drivetrain acceleration of turbine 2 at time t	$Da_5[t - 2T]$	Drivetrain acceleration of turbine 5 at time $t - 2T$	$Ta_5[t - 3T]$	Tower acceleration of turbine 5 at time $t - 3T$	—	—
$Da_2[t - T]$	Drivetrain acceleration of turbine 2 at time $t - T$	$Da_5[t - 3T]$	Drivetrain acceleration of turbine 5 at time $t - 3T$	$Da_5[t]$	Drivetrain acceleration of turbine 5 at time t	—	—
$Da_2[t - 2T]$	Drivetrain acceleration of turbine 2 at time $t - 2T$	—	—	$Da_5[t - T]$	Drivetrain acceleration of turbine 5 at time $t - T$	—	—

Table 12 Test performance of data-mining algorithms

Algorithm	MAE	SD of AE	MSE	SD of SE
NNE	6.79	24.95	668.08	7690.09
NN	7.30	25.23	689.51	7804.66
BT	10.71	24.74	726.12	7898.01
SVM	92.61	20.16	8981.93	5452.29
RF	15.18	28.94	1066.82	9041.59
CART	15.92	26.69	964.91	8221.04
kNN	20.84	31.30	1413.21	8795.90

5.2.2 *Comparative Study of Algorithms.* To train the virtual sensor model for predicting tower acceleration, the data-mining algorithms presented in Sec. 5.2.1 are used. Four metrics, i.e., Eqs. (5)–(8), are used to evaluate the performance of the data-mining algorithms. Table 12 illustrates the test results of the models extracted with different data-mining algorithms. The model extracted by algorithm NNE outperforms the others, and algorithm NNE is recognized as the best algorithm to develop the virtual sensor model for predicting tower acceleration (see Table 12). The MAPE of the NNE model is 0.1. Figure 15 shows the prediction results for the first 100 data points in the test dataset based on algorithm NNE.

5.3 **Monitoring Based on the Virtual Sensor Model.** To monitor the drivetrain or tower acceleration, the control-chart approach discussed in Sec. 4.3 is used. In this section, the virtual

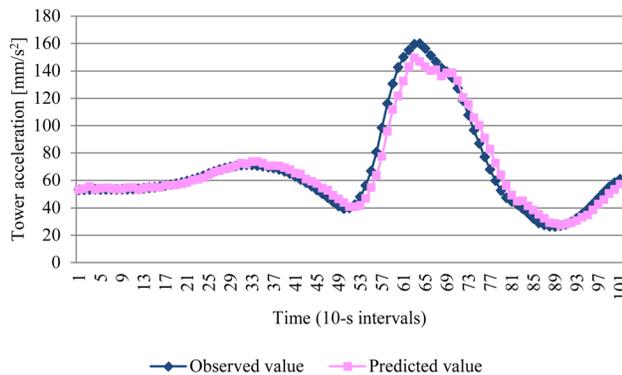


Fig. 15 Predicted and observed values of tower acceleration for the first 100 test points

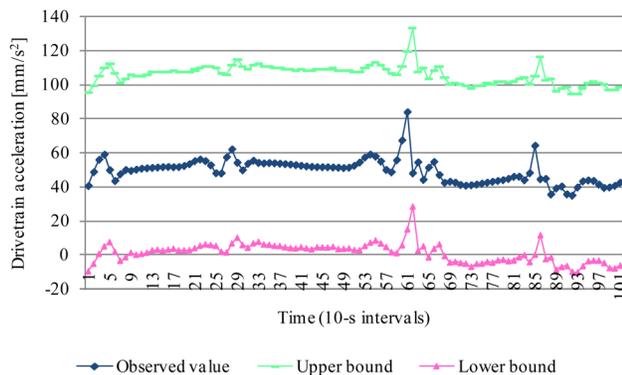


Fig. 16 The control chart for the dataset that contains normal drivetrain acceleration data

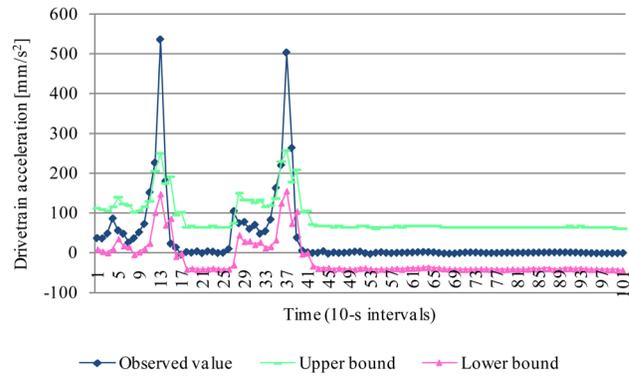


Fig. 17 The control chart for the dataset that contains abnormal drivetrain acceleration data

sensor model for predicting the drivetrain acceleration and tower acceleration is used as the center line of the control chart. Training, test, and error datasets of turbines 1, 2, and 5 are used to demonstrate the monitoring results by the control chart based on the virtual sensor model. Equations (21)–(23) are used to compute the upper control limit, the center line, and the lower control limit of the control chart to monitor the future observed values of drivetrain and tower acceleration. An observed value that falls outside the two boundaries is considered to be an abnormal occurrence of vibration in the drivetrain system or in the turbine tower. Figure 16 illustrates the monitoring results for drivetrain acceleration based on the test dataset that contained only normal data, while Fig. 17 presents the monitoring results for the drivetrain

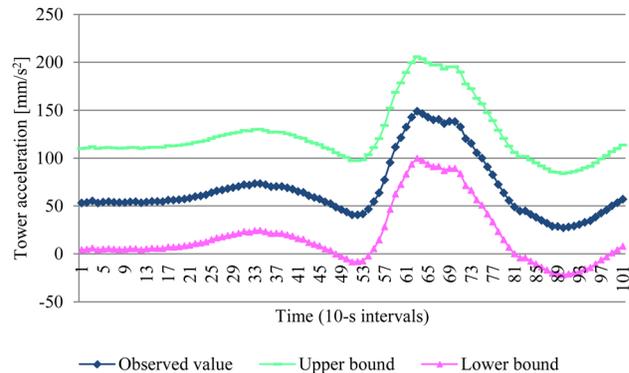


Fig. 18 The control chart for the dataset that contains normal tower acceleration data

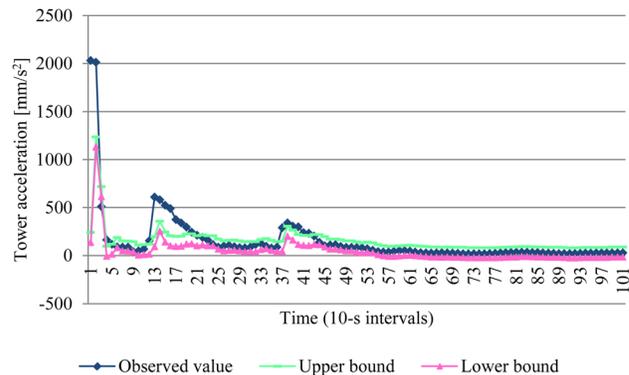


Fig. 19 The control chart for the dataset that contains some abnormal tower acceleration data

acceleration based on the error dataset. The error dataset contains some abnormal occurrences of wind turbine vibration data.

Figure 18 illustrates the monitoring results for tower acceleration of the test dataset based on control chart with center line provided by the virtual sensor model. Figure 19 shows the performance of the monitoring model in detecting abnormal tower acceleration data for the error dataset.

To realize the on-line monitoring of wind turbine vibration, an on-line access to SCADA system is needed for vibration health monitoring and model updates.

6 Conclusions

In this paper, three monitoring models for detecting abnormal vibration of wind turbines in time domain are introduced. Data-mining algorithms were used to develop the monitoring models based on SCADA data collected at six wind turbines operating at a large wind farm. The sampling interval of SCADA data is 10 s, it allows to detect abnormal statuses of wind turbines in the time domain.

A modified k -means clustering algorithm was used to develop the first vibration monitoring model. The k -means algorithm grouped data into clusters by examining their similarity. The clusters were labeled as normal or abnormal statuses of wind turbine vibration based on the error reports of wind turbines. The abnormal data points could be detected by comparing them with the data points assigned to the two class clusters.

The concept of control charts was used to develop models for monitoring of turbine vibration. In the second model, the baseline model, which provides accurate prediction of turbine vibration, was treated as the center line of the control chart. The upper and lower control limits of this control chart were established based on the center line to detect abnormal points that fall outside the two limits. In the third monitoring model, a virtual sensor model acted as the baseline model and provided the center line of the control chart. The virtual sensor model predicted turbine vibration by using information from other wind turbines rather than its own data. The baseline and virtual sensor models were both trained by data-mining algorithms based on a large volume of industrial data.

An approach for detecting abnormal drive train and tower vibration of a wind turbine was addressed in this research. Although the presented approach detects abnormal vibration, detection of root causes of the wind turbine vibration needs further investigation and requires higher frequency data. Incorporating vibration in the frequency domain with the presented framework offers a promising research direction.

Acknowledgment

The research reported in the paper has been supported by funding from the Iowa Energy Center, Grant No. 07-01.

References

- [1] Swedish Standard SS-EN 13306, European Standard EN 13306, 2001, "Maintenance Terminology".
- [2] Nilsson, J., and Bertling, L., 2007, "Maintenance Management of Wind Power Systems Using Condition Monitoring Systems—Life Cycle Cost Analysis for Two Case Studies," *IEEE Trans. Energy Convers.*, **22**(1), pp. 223–229.
- [3] Andrawus, J. A., Watson, J. F., Kishk, M., and Adam, A., 2006, "The Selection of a Suitable Maintenance Strategy for Wind Turbines," *Wind Eng.*, **30**(6), pp. 471–486.
- [4] Hameed, Z., Ahn, S. H., and Cho, Y. M., 2010, "Practical Aspects of a Condition Monitoring System for a Wind Turbine With Emphasis on Its Design, System Architecture, Testing and Installation," *Renewable Energy*, **35**(5), pp. 879–894.
- [5] Caselitz, P., and Giebhardt, J., 2005, "Rotor Condition Monitoring for Improved Operational Safety of Offshore Wind Energy Converters," *ASME J. Sol. Energy Eng.*, **127**(2), pp. 253–261.
- [6] Yang, W., Tavner, P. J., and Wilkinson, M. R., 2009, "Condition Monitoring and Fault Diagnosis of a Wind Turbine Synchronous Generator Drive Train," *IET Renewable Power Generation*, **3**(1), pp. 1–11.
- [7] Wiggelinkhuizen, E., Verbruggen, T., Braam, H., Rademakers, L., Xiang, J., and Watson, S., 2008, "Assessment of Condition Monitoring Techniques for Offshore Wind Farms," *ASME J. Sol. Energy Eng.*, **130**(3), pp. 1–9.
- [8] Kusiak, A., Zheng, H.-Y., and Song, Z., 2009, "Models for Monitoring Wind Farm Power," *Renewable Energy*, **34**(3), pp. 583–590.
- [9] Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S., 2004, "Using Control Charts to Monitor Process and Product Quality Profiles," *J. Quality Technol.*, **36**(3), pp. 309–320. Available at: <http://filebox.vt.edu/users/bwoodall/2004%20JQT%20WOODALL%20et%20a.pdf>
- [10] Mitra, A., 1998, *Fundamentals of Quality Control and Improvement*, 2nd ed., Prentice Hall, Upper Saddle River, New Jersey, NJ.
- [11] Kang, L., and Albin, S. L., 2000, "On-Line Monitoring When the Process Yields a Linear Profile," *J. Quality Technol.*, **32**(4), pp. 418–426.
- [12] Mestek, O., Pavlik, J., and Suchanek, M., 1994, "Multivariate Control Chart: Control Charts for Calibration Curves," *J. Anal. Chem.*, **350**(6), pp. 344–351.
- [13] Daubechies, I., 1992, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- [14] Kobayashi, M., 1998, *Wavelets and their Applications: Case Studies*, 1st ed., Philadelphia, Society for Industrial and Applied Mathematics, PA.
- [15] Tang, Y. Y., Yang, L. H., Liu, J., and Ma, H., 2000, *Wavelet Theory and Its Application to Pattern Recognition*, 1st ed., World Scientific, Singapore.
- [16] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., 2008, *Global Sensitivity Analysis. The Primer*, John Wiley & Sons, Chichester, UK.
- [17] Tan, P. N., Steinbach, M., and Kumar, V., 2006, *Introduction to Data Mining*, Addison Wesley, Boston, MA.
- [18] Devijver, P. A., and Kittler, J., 1982, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, London, UK.
- [19] Mosteller, F., 1948, "A k -Sample Slippage Test for an Extreme Population," *Ann. Math. Stat.*, **19**(1), pp. 58–65.
- [20] Montgomery, D. C., 2005, *Introduction to Statistical Quality Control*, 5th ed., John Wiley, New York.
- [21] Witten, I. H., and Frank, E., 2005, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., Morgan Kaufmann, San Francisco, CA.
- [22] Kohavi, R., and John, G. H., 1997, "Wrapper for Feature Subset Selection," *Artif. Intell.*, **97**(1–2), pp. 273–324.
- [23] Hansen, L. K., and Salamon, P., 1990, "Neural Network Ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, **12**(10), pp. 993–1001.
- [24] Siegelmann, H., and Sontag, E., 1994, "Analog Computation via Neural Networks," *Theor. Comput. Sci.*, **131**(2), pp. 331–360.
- [25] Liu, G. P., 2001, *Nonlinear Identification and Control: A Neural Network Approach*, Springer, London, UK.
- [26] Smith, M., 1993, *Neural Networks for Statistical Modeling*, Van Nostrand Reinhold, New York.
- [27] Friedman, J. H., 2002, "Stochastic Gradient Boosting," *Comput. Stat. Data Anal.*, **38**(4), pp. 367–378.
- [28] Friedman, J. H., 2001, "Greedy Function Approximation: A Gradient Boosting Machine," *Ann. Stat.*, **29**(5), pp. 1189–1232.
- [29] Schölkopf, B., Burges, C. J. C., and Smola, A. J., 1999, *Advances in Kernel Methods: Support Vector Learning*, MIT, Cambridge, MA.
- [30] Steinwart, I., and Christmann, A., 2008, *Support Vector Machines*, Springer-Verlag, New York.
- [31] Breiman, L., 2001, "Random Forests," *Mach. Learn.*, **45**(1), pp. 5–32.
- [32] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 1984, *Classification and Regression Trees*, Wadsworth & Brooks/Cole, Monterey, CA.
- [33] Shakhnarovich, G., Darrell, T., and Indyk, P., 2005, *Nearest-Neighbor Methods in Learning and Vision*, MIT, Cambridge, MA.
- [34] Kusiak, A., Li, M.Y., and Zheng, H.-Y., 2010, "Virtual Models of Indoor-Air-Quality Sensors," *Appl. Energy*, **87**(6), pp. 2087–2094.
- [35] Kusiak, A., and Li, W., 2010, "Virtual Models for Prediction of Wind Turbine Parameters," *IEEE Trans. Energy Convers.*, **25**(1), pp. 245–252.