A Data-Driven Approach for Condition Monitoring of Wind Turbine Pitch Systems

Cong Yang [1], Zheng Qian [1,*], Yan Pei [2] and Lu Wei [1]

[1] School of Instrumentation Science and Opto-electronics Engineering, Beihang University, Beijing 100083, China; yc1617@buaa.edu.cn (C.Y.); weilu@buaa.edu.cn (L.W.)

[2] State Key Laboratory of Operation and Control of Renewable Energy Storage Systems, China Electric Power Research Institute, Beijing 100192, China; peiyan@buaa.edu.cn

* Correspondence: qianzheng@buaa.edu.cn; Tel.: +86-010-8233-9267

check for updates

Abstract: With the rapid development of wind energy, it is important to reduce operation and maintenance (O&M) costs of wind turbines (WTs), especially for a pitch system, which suffers the highest failure rate and downtime. This paper proposes a data-driven method for pitch- system condition monitoring (CM) by only using supervisory control and data acquisition (SCADA) data without any faults, which could be applied to reduce O&M costs of pitch system by providing fault alarms. The pitch-motor temperature is selected as the indicator, and three feature-selection algorithms are employed to select the most appropriate input parameters for modeling. Six data-driven algorithms are applied to model pitch-motor temperature and the support vector regression (SVR) model has the highest accuracy. The control-chart method based on the residual errors between model output and measured value is utilized to calculate the outliers, thus the abnormal condition could be clearly identified once the outliers appear for a period of time. The effectiveness of the proposed method is demonstrated by several case studies, and compared with the classification models. Due to the adaptive ability and low cost, the proposed approach is suitable for online CM of pitch systems, and provides a strategy for CM of new WTs.

Keywords: pitch system; condition monitoring; SVR; feature selection; control chart

1. Introduction

As a commercially viable and environmentally sustainable energy source, wind energy has attracted sustained attention due to its abundance and high social benefits. The number of installed onshore and offshore wind farms is increasing to satisfy the rapidly growing demand. At the end of 2017, the cumulative installed wind-power capacity of China comprised approximately 188,392 MW, followed by USA, Germany, India, and Spain [1]. However, operation and maintenance (O&M) costs are still high due to the harsh environment and the early deterioration of critical components. Condition monitoring (CM) is an effective tool commonly employed to improve the reliability of wind turbines (WTs) and reduce O&M costs. It allows the maintenance to be scheduled based on the conditions of WT components [2].

As one of the critical components of WTs, the pitch system has led to the highest failure rate and downtime according to the results of the Reliawind Project [3]. A pitch-system fault could cause severe downtime events, such as blade fracture, which greatly limits the generating efficiency of WTs and increases O&M costs. Therefore, the CM of pitch systems is crucial for the early detection of pitch faults so as to reduce the costs and ensure the reliability of WTs.

The CM techniques for WTs have been widely studied based on the various signals, such as vibration, electrical, temperature, acoustic emission, lubrication oil parameters and supervisory control

and data acquisition (SCADA) signals. Meanwhile, they have been proved effective in detecting some specific faults of WT components [4,5]. However, it is difficult to apply some of them to monitor the condition of pitch systems due to unsuitability and complexity. For example, vibration analysis requires multiple sensors to be installed and large volumes of data to be collected, which consequently lead to a substantially increasing CM costs. Currently, large WTs are equipped with a SCADA system, providing a great deal of information on WT operating performance. Thus, the CM method based on SCADA data is cost-effective, as no additional sensors are needed, and as a result, a number of CM approaches using SCADA data have been researched in recent years [6,7]. Therefore, SCADA signals were used for pitch-system CM to reduce the expenditure in this paper.

The CM approaches could be generally categorized as the analytical-model-based method, knowledge-based method and data-driven-based method [8]. The analytical-model-based method requires constructing an accurate mathematical model [9]. Considering there are various parameters in the pitch system and the relationship between components is quite complicated, it is hard to construct the accurate mechanism model for the pitch system. Furthermore, there is a risk of model failure with the impact of noise and environmental changes. Consequently, the analytical-model-based method has been rarely used to detect pitch-system faults, while the other two approaches have been presented over the last few years.

The knowledge-based approach does not require a quantitative mathematical model for fault detection compared with the analytical-model-based method. For instance, Chen et al. [10] presented a priori knowledge-based adaptive neuro-fuzzy inference system to detect significant pitch faults automatically based on the 10-min averaged SCADA data. It has a great interpretability for allowing expert to introduce a priori knowledge to the system model. Ran Bi et al. [11] applied the normal behavior models based on the performance curves to detect pitch faults. The normal behavior models were obtained from the WT technical specification. Therefore, no training model is needed, but the knowledge of WT operation and control is required to identify abnormal operation conditions. Obviously, the accuracy of knowledge-based approaches is highly dependent on professional knowledge and long-term accumulation of experience, of which the integrity is difficult to be assured.

Compared with the knowledge-based method, the data-driven approach does not require much priori knowledge and experience. Data-driven models are established by mining information in historical data. It is applicable for the complex system due to its better adaptive ability.

Jamie L et al. [12] proposed a data-driven expert system to detect pitch faults using the 10-min averaged SCADA data. The RIPPER algorithm was used to generate the rules for the diagnosis of pitch-fault classes, including "no pitch fault", "potential pitch fault", and "pitch fault established". A classification accuracy of 85.5% was achieved in this system. For accurate classification, large quantities of data are required, especially for historical fault data.

Andrew Kusiak et al. [13] developed a data-mining-based two-class classifier to monitor blade pitch performance using 1-s SCADA data. By comparing the five data-mining algorithms, the genetic-programming algorithm was selected to perform the prediction of blade-angle implausibility faults with the best classification accuracy in the range of 68.7–87.4% for 13 time stamps. The maximum prediction time is 10 min with the accuracy of 68.7%, which can be improved for better condition-based maintenance.

B. Chen et al. [14] applied an approach for WT SCADA alarm processing and diagnosis using an artificial neural network (ANN). The trained ANN model was generated to identify if any pitch-system fault has occurred. However, the method was performed only based on the SCADA alarm signals, while some other signals with valuable information were potentially ignored. Moreover, the ANN model is a black box that is completely dependent on a large number of training samples and consequently loses insight into new problems.

Considering that these data-driven methods have high requirements for training data, for the effectiveness of these classification models depends on large quantities of labeled data, including historical fault data and healthy data, this paper aims to construct the model by only simply using the healthy historical data, and as a result, avoiding the difficulty of getting large volumes of historical

fault data. Furthermore, it provides a strategy for the diagnosis of new WTs. When constructing the model, the key to success is to select effective parameters. The status labels are commonly used as the model output to represent system health status, while the model input is usually determined based on these labels or research results of other literatures [12,13]. However, the model constructed by only using healthy historical data could not use labels to describe system health status because of the lack of fault status labels. Therefore, this paper evaluates the operating conditions of the pitch system by using the suitable indicator, and the related input parameters are determined with an appropriative feature algorithm in order to improve model efficiency and accuracy. In addition, a control chart is used for the identification of abnormal conditions. The model effect is demonstrated with the specific pitch-system faults.

The rest of the paper is organized as follows: Section 2 gives a brief description of the WT pitch system and related SCADA parameters. The model structure and methodology are introduced in Section 3. The comparative analysis of different data-driven models and monitoring results are discussed in Section 4, and conclusions are drawn in Section 5.

2. Pitch-System Description and Analysis

The pitch system is a vital component of WT, which ensures the effective utilization of wind power and the secure operation of WTs in an emergency by adjusting the blade pitch angle. Generally, pitch systems are divided into hydraulic-pitch systems and electric-pitch systems according to the different types of drive. In recent years, the latter has been used more frequently due to its extended control possibilities and higher precision. Moreover, it could avoid the leakage problems experienced with hydraulic-pitch systems [15]. Therefore, the research object of this article is the electric-pitch system, whose application prospects are better.

2.1. Pitch-System Structure

All parts of the pitch system are installed in the WT hub and rotate with the rotor. Three blades are equipped with an independent drive system in the hub, respectively. An electric variable-pitch drive system contains a pitch controller, pitch motor, gearbox, slewing bearing, limit switch, and battery cabinet, as shown in Figure 1 [16]. Specifically, the central controller calculates the optimal value of blade pitch angle according to the present wind condition and operation status of WT, which is sent to the motor driver afterwards. The blade pitch angle is adjusted by controlling the pitch motor. Thereby, the blade is controlled to the optimal position, driven by the gearbox and the slewing bearing. Besides, the power supply of the system is obtained via a slip ring, which is mounted between the rotor and the nacelle. When the external power supply is not available due to a slip ring or grid failure, the backup battery works to motivate the pitch motor.
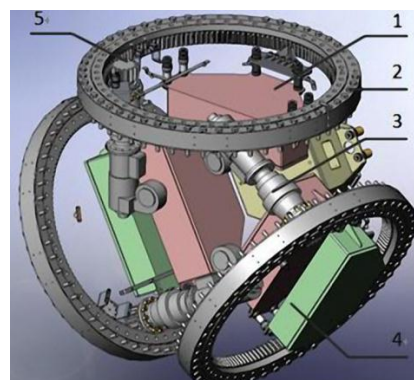
Figure 1. The structure of a pitch system [16]. (1) Pitch controller. (2) Slewing bearing. (3) Pitch motor. (4) Battery cabinet. (5) Gearbox.

2.2. Related Parameters of Pitch Systems

A typical SCADA system collects information of WT subassemblies, including various signals and alarms. However, there are over one hundred parameters, and only some of them are related to the pitch system. As for the pitch system, the monitored parameters are mainly associated with the operation of its components, such as the pitch motor and battery cabinet. Meanwhile, the pitch system is also related to the operation of other WT subassemblies and environment parameters, such as generator torque and wind speed. In addition, each SCADA message contains a time record and corresponding state code, which reflect the WT operation status. Based on the domain knowledge, the parameters which are associated with the pitch system are selected as shown in Table 1.

Table 1. The initial parameters related to the pitch system.

Environment Parameters	Electrical Parameters	Control Parameters	Temperature Parameters	Status Parameters
wind speed	pitch-motor current	blade pitch angle	pitch-motor temperature	state code
ambient temperature	pitch-motor power	generator torque	pitch-inverter temperature	limited power value
	generator power	generator speed	battery-cabinet temperature	
		hub speed	hub temperature	

3. Methodology

Because the fault samples are not contained in the training dataset, the status labels are not available for representing system health status. As a result, a suitable status indicator was applied for condition evaluation in this paper. Temperature parameter is a great condition indicator due to its thermal inertia and strong anti-interference capacity, which means that wind-speed uncertainty and uncontrollable noise have hardly any disturbance to it [17]. In addition, a large number of temperature data can be obtained directly from the SCADA system. Several temperature parameters of the critical WT components have been successfully applied to CM as a deterioration indication. For example, in Reference [18], a generator-temperature trend-analysis method was proposed to monitor the generator condition of WTs. In Reference [19], a generator-bearing-temperature model was generated using neural network algorithms to analyze the generator bearing failures of WTs. In Reference [20], the gearbox faults were predicted using SCADA oil temperature. This paper develops a pitch-motor temperature model to monitor the condition of WT pitch systems as the pitch-system operation is mainly driven by the pitch motor. It means that the pitch-motor temperature is used as a target for monitoring WT pitch system.

3.1. Modeling Process

The pitch-motor-temperature model for pitch-system CM is presented in Figure 2, which mainly includes the following steps:

1. Data Preprocessing

The SCADA system collects both normal and abnormal data. The training set for modeling only includes the healthy data collected in the normal operation, excluding downtime, failure, and maintenance. It could be obtained by removing the abnormal data according to the SCADA state code and maintenance record. Meanwhile, the data collected during the limited power period were not considered, which could be determined according to the power limit value in the SCADA system.

2. Feature Selection

In order to improve model performance, it is critical to select proper parameters to construct an efficient CM system for the pitch system of WTs. The related parameters used for the model input could be selected from Table 1 with a valid feature-selection algorithm.

3.   Model Training and Test

Based on the selected features, the model could be established using healthy historical SCADA data, which represent the relationship between the features and the condition indicator. Therefore, the target variable value could be estimated with the trained model and be close to the real value if the pitch system is normal. Otherwise, there is a possible failure in the pitch system.

4.   Residual Analysis

In the interest of a better interpretation for the model results, a proper residual-analysis method should be applied to identify the abnormal condition. This paper employs the control chart to get the bound. This means the pitch system might be abnormal when the residual error exceeds the bound.
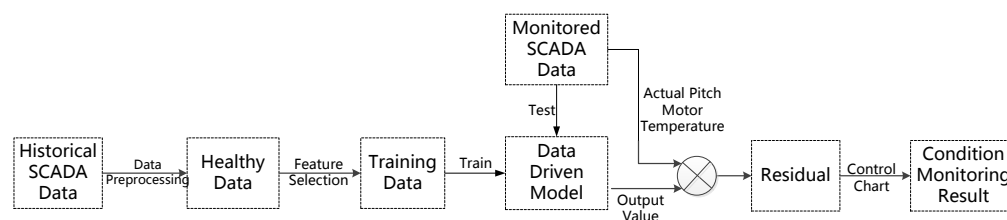

Figure 2. The framework of the data-driven model.

3.2. Feature Selection

For constructing an efficient data-driven model, it is crucial to select related parameters as the input of the pitch-motor-temperature model. The feature-selection algorithm could be applied to evaluate the importance of initial parameters and reduce modeling complexity. As an important preprocessing technique, it is widely used in data mining, which can be divided into the wrapper, embedded, and filter methods [21]. The wrapper model determines the optimal feature subset according to the objective function, which is usually defined as the performance index such as mean square error (MSE), while the filter model gets the feature ranking on the basis of relevance. With respect to the embedded model, the features are automatically selected with the specific machine-learning model. In this paper, three parameter-selection algorithms were applied, respectively, to acquire the optimal feature subset. The SCADA data of a turbine collected in four months (from 1 September, 2016 to 31 December, 2016) are illustrated for feature selection. During this period, there was no failure.

1.   Sequential Forward Selection

As one of the wrapper methods, sequential forward selection requires a proper learning algorithm and evaluation criteria to determine the feature subset. Here, support vector regression (SVR) is applied for modeling, and the MSE between the model output and monitored pitch-motor temperature is used to assess the importance of the parameters. It begins with an empty set. The feature making the MSE minimum is added into the set sequentially until the MSE does not decrease anymore. Finally, battery-cabinet temperature, blade pitch angle, hub temperature, ambient temperature, and pitch-motor current constitute the feature set, which makes the MSE minimum as shown in Table 2.

Table 2. The results of the sequential forward selection.

Input Number	Added Feature	Mean Square Error (MSE)
1	battery cabinet temperature	1.1786
2	blade pitch angle	0.8206
3	hub temperature	0.7562
4	ambient temperature	0.7146
5	pitch-motor current	0.5692
6	pitch-inverter temperature	0.7357
7	pitch-motor power	0.7458

2. Gradient-Boosting Regression Trees (GBRT)

As a flexible nonparametric machine-learning approach, GBRT can be used to train a regression model. This paper mainly takes advantage of the automatic feature selection of the GBRT algorithm as a typical embedded model. In addition, this method considers the correlation between the input variables compared with mutual information. The relative importance of the features is given in Table 3. It was found that the MSE is smallest when the first five parameters are selected as the model input, including hub temperature, battery-cabinet temperature, ambient temperature, pitch-motor current, and blade pitch angle.

Table 3. The relative importance of features.

Features	Relative Importance
hub temperature	0.4783
battery-cabinet temperature	0.4321
ambient temperature	0.0200
pitch-motor current	0.0120
blade pitch angle	0.0113
pitch-inverter temperature	0.0091
generator torque	0.0082
pitch-motor power	0.0078
wind speed	0.0070
hub speed	0.0043
generator speed	0.0042
generator power	0.0040

3. Mutual Information

This algorithm measures the mutual dependence between two variables as one of the filter models. In this research, it was used to calculate the correlation between the pitch-motor temperature and the related parameters in Table 1. The results are shown in Table 4.

Obviously, the former two algorithms have the same result. Namely, the pitch-motor-temperature model has an optimal performance with the following five parameters as the input variables: battery-cabinet temperature, hub temperature, ambient temperature, pitch-motor current, and blade pitch angle. In addition, these five parameters have a higher correlation with the pitch-motor temperature according to the result of mutual information. Considering the criteria that input variables should be highly correlated with the output and less correlated with each other, as a redundant variable, pitch-inverter temperature is not considered because it is highly corrected with the battery-cabinet temperature. Therefore, these five parameters are selected to construct the regression model eventually. In this case, the R-squared is 0.9056, the MSE is 0.5692, and the mean absolute error (MAE) is 0.5370.

Table 4. Mutual information between the parameters and the pitch-motor temperature.

Features	Mutual Information
battery-cabinet temperature	0.93
hub temperature	0.92
pitch-inverter temperature	0.89
ambient temperature	0.62
pitch-motor current	0.14
blade pitch angle	0.13
generator torque	0.11
generator power	0.11
hub speed	0.10
generator speed	0.10
wind speed	0.08
pitch-motor power	0.01

3.3. Model Construction

With the selected five parameters as the input variables, a pitch-motor-temperature model is established using the healthy historical SCADA data. There are multiple data-driven algorithms that could be used to model the pitch-motor temperature, such as ridge regression (Ridge) [22], least absolute shrinkage and selection operator (Lasso) [23], k-Nearest Neighbors (kNN) [24], random forest (RF) [25], ANN [26], and SVR [27]. Considering SVR has the advantages of solving problems such as nonlinear and local minima point with finite samples [27], this paper applies SVR to establish the model.

SVR principles are as follow [27].

Given a training dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, x_i is the i-th m-dimensional input vector and $y_i \in R$ is the corresponding target of x_i. The SVR model can be represented as Equation (1):

$$f(x) = w^T \phi(x) \tag{1}$$

where $f(x)$ is the model output, w is the normal vector, b is the bias parameter, and $\phi(x)$ denotes a fixed feature-space transformation, which transforms the nonlinear problems in the input space into the linear problems in the feature space.

SVR assumes the absolute difference between the target y and the model output $f(x)$ is less than ϵ. Thus, the SVR optimization-objective function is given by Equation (2):

$$min \; \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} l_\epsilon (f(x_i) - y_i) \tag{2}$$

where C is the regularization parameter, and l_ϵ is the ϵ-insensitive loss function.

$\tilde{\xi}_i$ and $\hat{\tilde{\xi}}_i$ are defined as two slack variables. Then, the optimization problem can be rewritten as Equation (3):

$$min \; \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} (\tilde{\xi}_i + \hat{\tilde{\xi}}_i) \tag{3}$$

subject to:

$$f(x_i) - y_i \leq \epsilon + \tilde{\xi}_i \; , \; \tilde{\xi}_i \geq 0, i = 1, 2, \ldots, n \tag{4}$$

$$y_i - f(x_i) \leq \epsilon + \hat{\tilde{\xi}}_i, \; \hat{\tilde{\xi}}_i \geq 0, i = 1, 2, \ldots, n \tag{5}$$

This problem can be achieved by introducing Lagrange multipliers and optimizing the Lagrange function. Therefore, the dual problem of Equation (3) can be obtained as Equation (6):

$$max \sum_{i=1}^{n} y_i(\hat{\alpha}_i - \alpha_i) - \epsilon(\hat{\alpha}_i + \alpha_i) - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j + \alpha_j)k(x_i, x_j) \tag{6}$$

subject to:

$$\sum_{i=1}^{n}(\hat{\alpha}_i - \alpha_i) = 0, \ 0 \le \alpha_i, \hat{\alpha}_i \le C \tag{7}$$

where $\alpha_i, \hat{\alpha}_i$ are Lagrange multipliers, and $k(x_i, x_j)$ stands for the kernel function.

Eventually, the regression function obtained is shown in Equation (8) by solving the equations above:

$$f(x) = \sum_{i=1}^{n}(\hat{\alpha}_i - \alpha_i)k(x, x_i) + b \tag{8}$$

where the kernel function $k(x, x_i)$ is used to map the training set to high dimension space. Thus, SVR is capable of solving both linear and nonlinear regression problems. In this paper, the SVR parameters are determined by grid search.

3.4. Residual Analysis

In order to determine whether a sustained change in the condition of the pitch system has occurred, an exponentially weighted moving average (EWMA) control chart was applied to identify the abnormal condition of the pitch system. The EWMA can be viewed as a weighted average of all past and current observations. It has a smoothing effect on the uncontrollable noise. In addition, it is very effective against small process shifts. Consequently, the EWMA control chart is typically applied with individual observations [28]. In this paper, an EWMA-based residual analysis method is adopted for the CM of WT pitch system. The outliers of the EWMA control chart that exceed the control limits are viewed as out of control, which could indicate impending pitch-system failure.

The raw SCADA data were collected at 10-min intervals. In order to avoid the misidentification of fault events, the moving-average approach is used prior to the control chart to reduce the data noise. In this research, the window length is set to 6.

The EWMA statistic is defined as Equation (9):

$$z_i = \lambda e_i + (1 - \lambda)z_{i-1} \tag{9}$$

where $0 < \lambda \le 1$ is the smoothing parameter, and e_i is the i-th residual error, defined as the deviation between the measured pitch-motor temperature and the model output value. The starting value z_0 is the process target μ_0, which is set to the average of historical residual errors $\bar{e}$ in the normal operations. If the observations e are independent random variables with variance σ^2, the variance of z_i can be calculated as Equation (10):

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda}\right) \left[1 - (1 - \lambda)^{2i}\right] \tag{10}$$

Thus, the EWMA control chart can be constructed by drawing the relationship between z_i and i. The center line and control limits can be represented as Equations (11)–(13):

$$UCL = \mu_0 + L\sigma\sqrt{\frac{\lambda}{(2 - \lambda)} \left[1 - (1 - \lambda)^{2i}\right]} \tag{11}$$

$$Center \ Line \ (CL) = \mu_0 \tag{12}$$

$$LCL = \mu_0 - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)}\left[1 - (1-\lambda)^{2i}\right]} \qquad (13)$$

where L stands for the width of the control limits. The performance of the EWMA control chart mainly depends on the reasonable selection of design parameters, including λ and L. It was found that 0.05, 0.1, and 0.2 were commonly used values of λ in practical application, and $L = 3$ works reasonably well, which corresponds to the usual 3σ limits [28]. In this study, λ and L are set to 0.2 and 3, respectively.

4. Validation and Test Results

In order to validate the effectiveness of the proposed approach, the model was applied to the CM of the pitch system of a total of 24 variable-speed and variable-pitch WTs from a commercial wind farm, which is located in China. The 10-min averaged SCADA data collected from June 2016 to July 2017 are available. The CM results were compared to the different models and the original event logs. In addition, all the computations were carried out on the computer using Python programming language (The computer configuration is as follows: CPU: Inter (R) Core (TM) i5-4590 CPU @ 3.30 GHz 3.30 GHz; RAM: 4.00 GB; System type: 64 bit). The data-driven algorithms used in this paper are from the scikit-learn package in Python.

4.1. Comparative Analysis

4.1.1. Data-Driven Algorithms Comparison

In order to demonstrate the capability of SVR in modeling the pitch-motor temperature, the SVR model was compared with five data-driven models developed by five well-known algorithms, including Ridge, Lasso, kNN, RF, and ANN. In addition, MSE and MAE were used as metrics to evaluate the model performance.

In this research, the parameters of each model were determined by grid search with a 10-fold cross validation in the training. For the Ridge model and Lasso model, the optimal regularization parameter was selected from a set {0.01, 0.1, 1.0, 10.0}. For the kNN model, the Euclidean distance was considered, and $k = 1, 2, 3 \ldots 10$ were evaluated to determine the optimal value. For the RF model, the number of features to consider when looking for the best split was selected from a set {*'auto'*, *'sqrt'*, *'log2'*}, and the number of trees in the forest was determined from a set {10, 100, 200 ... 500}. For the ANN model, three layers was considered, and three activation functions were evaluated, including logistic sigmoid function, hyperbolic tan function, and rectified linear unit function. A stochastic gradient-based optimizer was used for weight optimization. The maximum number of iterations was selected from a set {100, 200 ... 500}. For the SVM model, a Gaussian kernel was considered. The kernel coefficient *gamma* = 0.01, 0.1, 1, 10 and penalty parameter C = 1, 10, 100, 1000 were evaluated, and the combination of them with the smallest 10-fold cross-validation error was selected.

The results in Table 5 indicate that SVR is better than the other five data-driven algorithms in this paper. In Table 5, four datasets obtained from four turbines, respectively, were tested to evaluate the performance of these six models, and each test dataset contains the healthy SCADA data collected in one month. For each data-driven model, the healthy historical SCADA data collected in three months were considered as training dataset. Obviously, the SVR-based model provided the lowest MSE and MAE. Considering the modeling accuracy, SVR is more suitable for estimating the pitch-motor temperature due to its excellent generalization ability, and, as a result, it was employed to model the pitch-motor temperature eventually.

Table 5. MSE and mean absolute error (MAE) of different data-driven algorithms.

Algorithm	Test Set 1		Test Set 2		Test Set 3		Test Set 4	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Ridge regression (Ridge)	0.618	0.570	0.796	0.697	0.998	0.795	0.636	0.610
Least absolute shrinkage and selection operator (Lasso)	0.656	0.598	0.815	0.691	0.991	0.788	0.682	0.633
k-Nearest Neighbors (kNN)	1.056	0.718	0.744	0.613	1.145	0.818	0.803	0.620
Random forest (RF)	0.977	0.618	0.680	0.556	1.421	0.856	0.614	0.533
Artificial neural network (ANN)	0.678	0.589	0.569	0.577	0.932	0.740	0.628	0.605
Support vector regression (SVR)	0.569	0.530	0.523	0.550	0.849	0.719	0.614	0.591

4.1.2. Comparison with Classification Models

As the classification models are often employed to identify the status patterns of a wind-turbine pitch system [10–12], the classification models were also compared in this section to further prove the effectiveness of the proposed method. The classification approaches construct the models by mining the differences between large quantities of historical fault data and healthy data, and the status label is used to represent the health status of pitch system. In this paper, ANN and SVM were selected for constructing the classification models. The input parameters of the model include battery-cabinet temperature, hub temperature, ambient temperature, pitch-motor current, blade pitch angle, and pitch-motor temperature. Using the status labels as the model output, two conditions were identified by the classification model, including normal operation and pitch-system fault.

In order to compare the performance of different models, three datasets obtained from three turbines, respectively, were tested. Each dataset contains labeled healthy samples and pitch system fault samples. Due to the limited volume of pitch-fault data, there were 3856 sets of SCADA data as the training dataset for the classification models, including 1508 pitch fault samples and 2348 healthy samples. In order to make an accurate comparison, the number of training samples of different models in the comparison experiment was the same. Similarly, the model parameters were determined by grid search with a 10-fold cross validation in training, and the ones with the smallest error were selected. The monitoring results are presented in Table 6. Obviously, the proposed model could identify the conditions of the pitch system with higher accuracy. Although the accuracy of classification models could be improved by increasing the training dataset, it is difficult to provide sufficient historical fault samples covering different failure types in practical applications, especially for the new wind farms. Compared with the classification models, the proposed model could identify the abnormal conditions of pitch systems by only using the same number of healthy samples, which is more effective and applicable in practical applications.

**Table 6.** The monitoring accuracy of different models.

| Dataset	Classification Model		Proposed Model
	ANN	SVM	
1	0.717	0.739	0.874
2	0.767	0.777	0.858
3	0.601	0.624	0.887

4.2. Monitoring Results

In this section, four case studies have been analyzed specifically to demonstrate the feasibility of the proposed approach, including a normal event and three pitch-system-failure events derived from four WTs, respectively.

4.2.1. Case 1: Normal Operation

A turbine with normal operations between November 2016 and February 2017 was illustrated. The SCADA data collected in the previous three months (from 1 November, 2016 to 31 January, 2017) were trained to construct the pitch-motor-temperature model. Afterwards, the feasibility of the model was tested with the data collected from 1 February, 2017 to 28 February, 2017. The estimated value of the model and the measured pitch-motor temperature are shown in Figure 3. The R-Squared, root MSE (RMSE), and MAE were 0.9674, 0.7781 and 0.6007, respectively. These results indicate that the pitch-motor temperature estimated by the proposed model was close to the actual value. In addition, the residual errors of the pitch-motor temperature are normally distributed as shown in Figure 4, which demonstrates that residual errors could be analyzed with the EWMA control chart. In summary, the regression model works well in the normal operations.
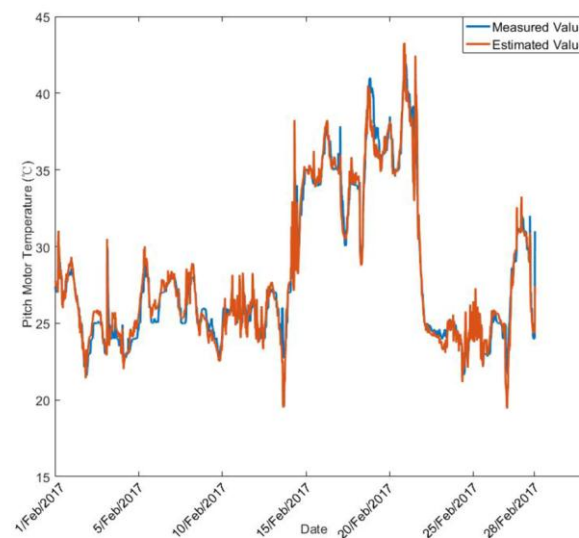
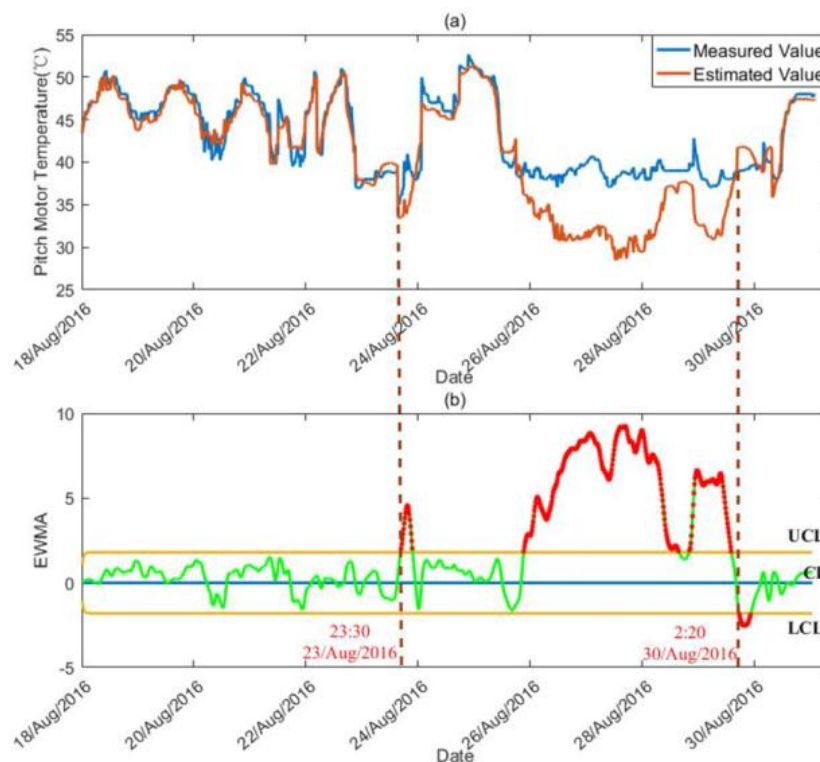**Figure 3.** The comparison of the model output and the actual temperature.
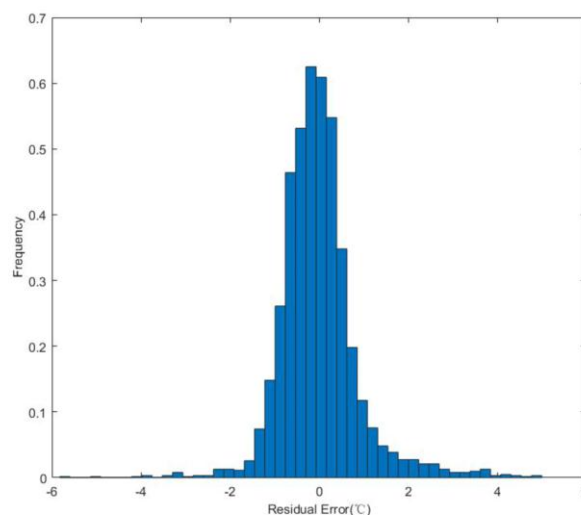
Figure 4. The distribution of the residual errors.

4.2.2. Case 2: Limit-Switch Failure

According to the maintenance record, a turbine was shut down exceptionally at 2:20 on 30 August, 2016. It has found that there was a rag near the limit switch that caused the switch to be stuck. The

fault was cleared at 11:20. The limit switch is responsible for the stroke control and position protection of pitch operations. Due to the limit-switch fault, the pitch controller received wrong pitch signals, leading to undesired pitch-control operations. However, there was no alarm or fault detected by the SCADA system prior to the shutdown.

With the proposed method, the abnormal condition was detected ahead of the shutdown. The SCADA data collected from 18 August, 2016 to 30 August, 2016 were tested to illustrate the effectiveness of the model. The model output and the actual temperature during this period are presented in Figure 5a. It is obviously observed that the actual temperature is higher than the output value in the later period, which means a pitch-system fault has occurred. Specifically, the model firstly detected the abnormal data at 23:30 on 23 August, 2016 with the moving average approach and the EWMA control chart. At this time, the EWMA statistic began to exceed the upper control limit as shown in Figure 5b. Furthermore, it returned to normal at 11:30 on 30 August, 2016 due to maintenance. It indicates that the proposed method can detect limit-switch failure about six days earlier than the fault occurrence.
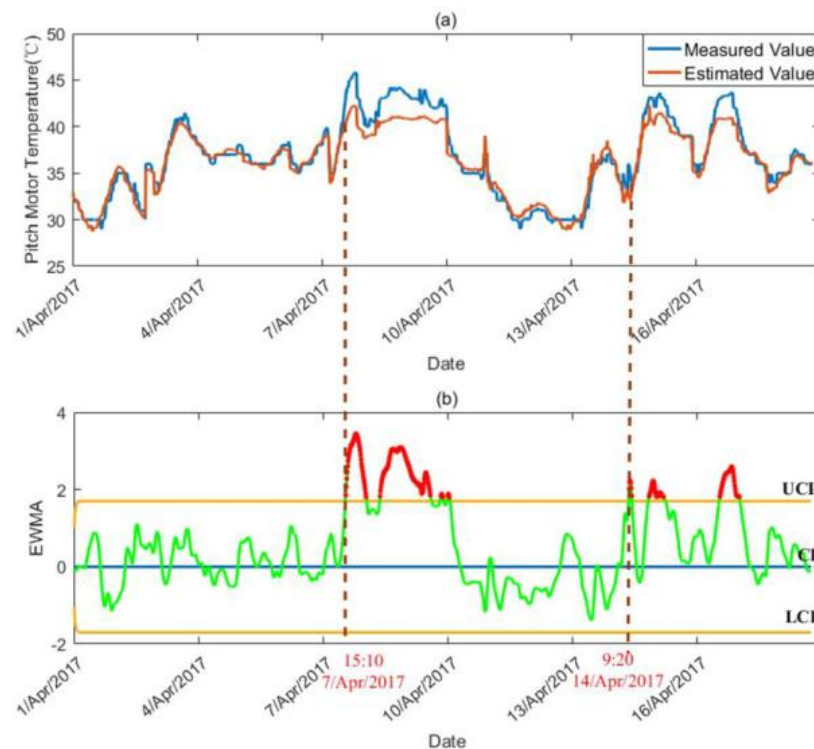

Figure 5. The monitoring results of Case 2. (**a**) The comparison of the model output and the actual temperature. (**b**) The residual analysis with the exponentially weighted moving average (EWMA) control chart.

4.2.3. Case 3: Angle-Encoder Failure

A pitch-system fault of a turbine was alarmed by the SCADA system at 9:20 on 14 April, 2017. An angle-encoder failure was diagnosed based on the inspection to pitch system carried out on 19 April, 2017. The Angle encoder is an important component of pitch systems, permitting the measurement of pitch angles that are served for blade control. The angle-encoder fault caused the undesired pitch action because the pitch controller and driver received wrong pitch signals. The encoder was eventually replaced at 15:00 on 19 April, 2017.

The angle-encoder failure was tested using SCADA data from 1 April, 2017 to 20 April, 2017. As shown in Figure 6b, the estimated pitch-motor temperature given by the model obviously deviated from the measured value. With the proposed residual analysis method, the pitch-system fault

was firstly detected at 15:10 on 7 April, 2017, which is nearly seven days earlier than the SCADA alarm system.
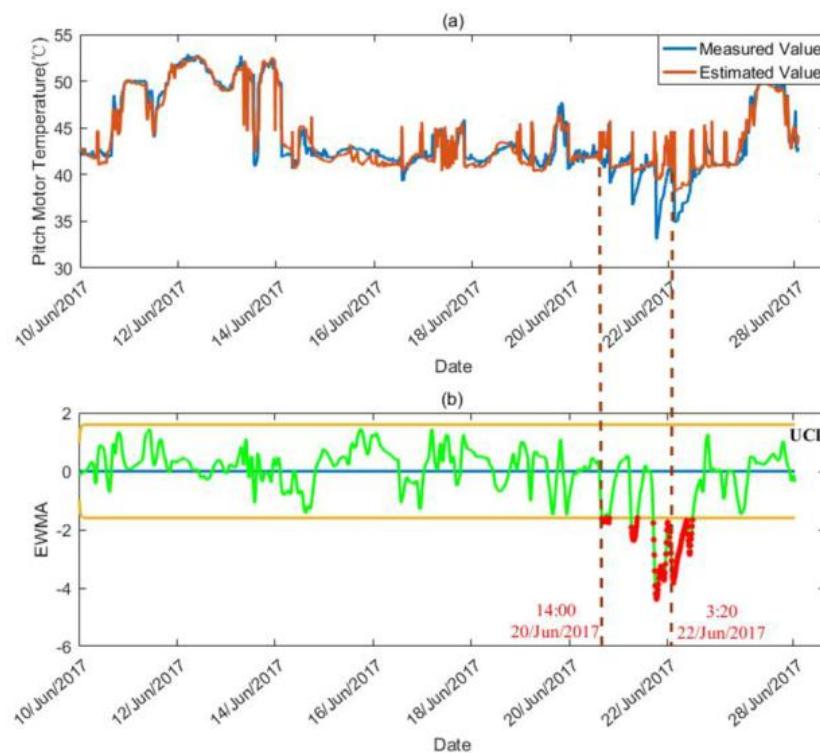

Figure 6. The monitoring results of Case 3. (**a**) The comparison of the model output and the actual temperature. (**b**) The residual analysis with EWMA control chart.

4.2.4. Case 4: Slip-Ring Failure

A sustained communication alarm of a turbine was given by the SCADA alarm system, which started from 3:20 on 22 June, 2017. An inspection was carried out afterwards. It diagnosed that the slip ring was damaged. The slip ring is responsible for the communication between the main controller and the pitch controller. Due to the slip-ring fault, the blade pitch angle could not change to the optimal value because the pitch actuator did not receive the correct pitch signal. Therefore, the turbine almost stopped operation from 22 June, 2017 to 26 June, 2017, which caused heavy power loss. According to the maintenance record, the slip ring was replaced at 18:20 on 26 June, 2017. It was found there was no SCADA alarm after the maintenance.

The SCADA data collected from 10 June, 2017 to 28 June, 2017 were tested to illustrate the feasibility of the approach. It was observed that there was a large deviation between the model output and the measured pitch-motor temperature as shown in Figure 7a. With the moving-average approach and the EWMA control chart, the exceptional value was firstly detected at 14:00 on 20 June, 2017 when the EWMA statistic started to exceed the lower control limit, which is presented in Figure 7b. It indicates that the proposed model can detect the slip-ring failure 37.3 h earlier than the SCADA system. In addition, it resumed to normal operation after maintenance.

Figure 7. The monitoring results of Case 4. (**a**) The comparison of the model output and the actual temperature. (**b**) The residual analysis with EWMA control chart.

4.2.5. Analysis of Monitoring Results

In order to evaluate the adaptive ability of the proposed approach, a total of 24 normal cases and eight pitch-system failure cases were tested in this paper. To avoid making false alarms and improve the performance of the proposed monitoring framework, this paper ultimately defined that an alarm occurs when five consecutive points exceed the control limits according to the validation results. The test results show that 24 normal cases are all correctly identified as no abnormal condition. Meanwhile, the abnormal conditions of actual faults were all identified with the EWMA control chart. The monitoring results of pitch-system failure cases are specifically shown in Table 7. All eight pitch-system failures were successfully identified at least one day prior to the SCADA alarm system. The early detections of abnormal conditions provide more time for workers to make timely maintenance decisions. Consequently, the undesired downtime and cost could be reduced or even avoided.

Table 7. The monitoring results of pitch-system failure cases.

Case	Alarm Time		Time Ahead
	Proposed Model	**Supervisory Control and Data Acquisition (SCADA) Alarm System**	
1	23 August, 2016 23:30	30 August, 2016 2:20	146.8 h
2	19 December, 2016 17:00	21 December, 2016 17:30	48.5 h
3	25 January, 2017 15:50	28 January, 2017 12:00	68.2 h
4	7 April, 2017 15:10	14 April, 2017 9:20	162.2 h
5	26 April, 2017 16:10	29 April, 2017 16:30	72.3 h
6	20 June, 2017 7:10	21 June, 2017 21:20	38.2 h
7	20 June, 2017 14:00	22 June, 2017 3:20	37.3 h
8	24 June, 2017 00:10	26 June, 2017 7:10	55.0 h

Additionally, the proposed model is applicable for online CM. According to the results of six repeated experiments, it costs an average of 420.21 s to finish the model training, including 10,562 sets of 10-min averaged SCADA data over a period of three months. Meanwhile, the model-training time would be longer as the dataset increased. Although it takes a lot of time for training due to the data preprocessing and parameter optimization, the model could be established in advance, and could be used directly in the monitoring process. In addition, it could be updated to achieve better performance with newly collected data after a certain period of time. During the monitoring process, for each 10-min averaged SCADA data collected, the model gives a result on whether there is an outlier. The computing time is less than 1 s for each sample, which fully satisfies the requirement of online CM. In addition, the computation effect could be improved with better computer configurations.

5. Conclusions

A data-driven approach for the CM of WT pitch systems using SCADA data has been presented. The pitch motor temperature was applied to monitor the condition of pitch systems as the status indicator. Then, a regression model was established to represent the relationship between the pitch-motor temperature and the selected features in the normal operations, including battery-cabinet temperature, hub temperature, ambient temperature, pitch-motor current and pitch blade angle, which were determined based on three feature-selection algorithms. An SVR algorithm was employed to model the pitch-motor temperature and compared with five data-driven algorithms, Ridge, Lasso, kNN, RF, and ANN. As a result, SVR was determined to construct the data-driven model because of its excellent generalization ability. It was found there is little difference between the model output and the measured value in the normal operations, and the residual errors are normally distributed. Therefore, the abnormal monitored temperature indicates a potential pitch-system fault. With the moving-average approach and the EWMA control chart, the abnormal condition could be identified clearly once there are five residual-based statistics exceed the control limits. The results demonstrate that pitch-system failures are successfully detected earlier than the SCADA alarm system with the proposed approach. Moreover, it is more effective and applicable than the classification models.

Compared with the knowledge-based method, much professional knowledge and experience are not needed in this paper. The priori knowledge of physical characteristics for modeling is also not required compared with the analytical-model-based method. The proposed approach could construct the regression model automatically with the healthy historical SCADA data. It provides a strategy for the CM of new WTs. In conclusion, the proposed method is applicable in industrial applications due to its great adaptive ability and low cost.

This paper only focuses on the CM of WT pitch systems due to a small number of pitch-system-fault samples. In the SCADA system, various alarms and fault logs are recorded. However, it is difficult to accurately determine the fault types depending on the SCADA system. Therefore, an advanced fault-identification system should be developed. As a result, further studies on fault isolation will be summarized in the next research by collecting sufficient fault samples and investigating the identification approaches.

Author Contributions: Methodology, C.Y. and Z.Q.; experiments, C.Y. and L.W.; data analysis, Y.P.; writing—original draft, C.Y.; draft—review, Z.Q., Y.P., and L.W.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

O&M	operation and maintenance
WT	wind turbine
CM	condition monitoring
SCADA	supervisory control and data acquisition
SVR	support vector regression
ANN	artificial neural network
MSE	mean square error
GBRT	gradient boosting regression trees
MAE	mean absolute error
Ridge	ridge regression
Lasso	least absolute shrinkage and selection operator
kNN	k-Nearest Neighbors
RF	random forest
EWMA	exponentially weighted moving average

References

1. Global Wind Energy Council. Global Statistics. Available online: http://www.gwec.net/global-figures/graphs/ (accessed on 27 July 2018).
2. Márquez, F.P.G.; Tobias, A.M.; Pérez, J.M.P.; Papaelias, M. Condition monitoring of wind turbines: Techniques and methods. *Renew. Energy* **2012**, *46*, 169–178. [CrossRef]
3. Wilkinson, M.; Harman, K.; Hendriks, B. Measuring wind turbine reliability: Results of the reliawind project. *Wind Energy* **2011**, *35*, 102–109. Available online: http://proceedings.ewea.org/annual2011/allfiles2/901_EWEA2011presentation.pdf (accessed on 15 August 2018).
4. Qiao, W.; Lu, D. A survey on wind turbine condition monitoring and fault diagnosis—Part I: Components and subsystems. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6546–6557. [CrossRef]
5. Qiao, W.; Lu, D. A survey on wind turbine condition monitoring and fault diagnosis—part II: signals and signal processing methods. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6546–6557. [CrossRef]
6. Yang, W.; Court, R.; Jiang, J. Wind turbine condition monitoring by the approach of SCADA data analysis. *Renew. Energy* **2013**, *53*, 365–376. [CrossRef]
7. Leahy, K.; Gallagher, C.; O'Donovan, P.; Bruton, K.; O'Sullivan, D.T.J. A Robust Prescriptive Framework and Performance Metric for Diagnosing and Predicting Wind Turbine Faults Based on SCADA and Alarms Data with Case Study. *Energies* **2018**, *11*, 1738. [CrossRef]
8. Ms, L.H.C.; Russell, E.L.; Braatz, R.D. Fault Detection and Diagnosis in Industrial Systems. *Technometrics* **2002**, *44*, 197–198. [CrossRef]
9. Cho, S.; Gao, Z.; Moan, T. Model-based fault detection, fault isolation and fault-tolerant control of a blade pitch system in floating wind turbines. *Renew. Energy* **2018**, *120*, 306–321. [CrossRef]
10. Chen, B.; Matthews, P.C.; Tavner, P.J. Automated on-line fault prognosis for wind turbine pitch systems using supervisory control and data acquisition. *IET Renew. Power Gener.* **2015**, *9*, 503–513. [CrossRef]
11. Bi, R.; Zhou, C.; Hepburn, D.M. Detection and classification of faults in pitch-regulated wind turbine generators using normal behaviour models based on performance curves. *Renew. Energy* **2016**, *105*, 674–688. [CrossRef]
12. Godwin, J.L.; Matthews, P. Classification and detection of wind turbine pitch faults through SCADA data analysis. Available online: https://core.ac.uk/download/pdf/25851339.pdf (accessed on 15 August 2018).
13. Kusiak, A.; Verma, A. A Data-Driven Approach for Monitoring Blade Pitch Faults in Wind Turbines. *IEEE Trans. Sustain. Energy* **2011**, *2*, 87–96. [CrossRef]
14. Chen, B.; Qiu, Y.N.; Feng, Y.; Tavner, P.J.; Song, W.W. Wind turbine SCADA alarm pattern recognition. In Proceedings of the IET Conference on Renewable Power Generation, Edinburgh, UK, 6–8 September 2011. [CrossRef]
15. Hau, E.; Renouard, H.V. Wind turbines: Fundamentals, technologies, application, economics. *IEEE Electr. Insul. Mag.* **2013**, *19*, 48. [CrossRef]
16. Jian, K. Wind turbine pitch system. *China Eletr. Equip. Ind.* **2011**, *7*, 52–57. [CrossRef]

17. Hu, Y.; Li, H.; Liao, X.; Song, E.; Liu, H.; Chen, Z. A probability evaluation method of early deterioration condition for the critical components of wind turbine generator systems. *Mech. Syst. Signal Proc.* **2016**, *76–77*, 729–741. [CrossRef]

18. Peng, G.; Infield, D.; Yang, X. Wind Turbine Generator Condition-Monitoring Using Temperature Trend Analysis. *IEEE Trans. Sustain. Energy* **2011**, *3*, 124–133. [CrossRef]

19. Kusiak, A.; Verma, A. Analyzing bearing faults in wind turbines: A data-mining approach. *Renew. Energy* **2012**, *48*, 110–116. [CrossRef]

20. Feng, Y.; Qiu, Y.; Crabtree, C.J.; Long, H.; Tavner, P.J. Monitoring wind turbine gearboxes. *Wind Energy* **2013**, *16*, 728–740. [CrossRef]

21. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182. [CrossRef]

22. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **2000**, *12*, 55–67. [CrossRef]

23. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* **2011**, *73*, 267–288. [CrossRef]

24. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185. [CrossRef]

25. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

26. Jain, A.K.; Mao, J.; Mohiuddin, K.M. *Artificial Neural Networks: A Tutorial*; Academic Press: Cambridge, MA, USA, 1996; pp. 31–44.

27. Bishop, C.M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*; Springer: New York, NY, USA, 2007; pp. 325–345.

28. Montgomery, D.C. *Introduction to Statistical Quality Control*; Wiley: New York, NY, USA, 2009; pp. 419–428.