Energies **2011**, *4*, 2077-2093; doi:10.3390/en4112077

OPEN ACCESS

energies

ISSN 1996-1073

www.mdpi.com/journal/energies

Article

Wind Turbine Gearbox Condition Monitoring with AAKR and Moving Window Statistic Methods

Peng Guo and Nan Bai *

School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China; E-Mail: huadiangp@163.com

* Author to whom correspondence should be addressed; E-Mail: bainan66@163.com; Tel./Fax: +86-10-6177-2654.

Received: 30 August 2011; in revised form: 24 October 2011 / Accepted: 21 November 2011 / Published: 23 November 2011

Abstract: Condition Monitoring (CM) of wind turbines can greatly reduce the maintenance costs for wind farms, especially for offshore wind farms. A new condition monitoring method for a wind turbine gearbox using temperature trend analysis is proposed. Autoassociative Kernel Regression (AAKR) is used to construct the normal behavior model of the gearbox temperature. With a proper construction of the memory matrix, the AAKR model can cover the normal working space for the gearbox. When the gearbox has an incipient failure, the residuals between AAKR model estimates and the measurement temperature will become significant. A moving window statistical method is used to detect the changes of the residual mean value and standard deviation in a timely manner. When one of these parameters exceeds predefined thresholds, an incipient failure is flagged. In order to simulate the gearbox fault, manual temperature drift is added to the initial Supervisory Control and Data Acquisitions (SCADA) data. Analysis of simulated gearbox failures shows that the new condition monitoring method is effective.

Keywords: wind turbine condition monitoring; gearbox; Autoassociative Kernel Regression; residual analysis; moving window statistics

1. Introduction

Challenging environmental factors combined with high and turbulent winds place serious demands on wind turbines and result in significant component failure rates, highlighting the importance of maintenance. Appropriate use of condition monitoring can, by detecting faults at an early stage, reduce turbine repair and maintenance costs.

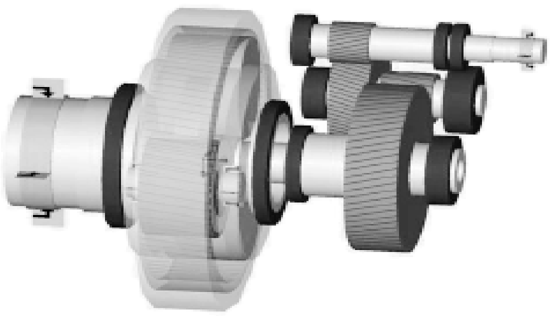
The gearbox is one of the most important components of a wind turbine. In [1] the authors show that the maintenance cost for the gearbox is very high compared with the other higher failure rate components such as electric system and hydraulic system, especially for the offshore wind farms where maintenance will need a boat, crane and nice weather. The gearbox undergoes varying speed and load. With the varying wind speed, the rotating speed and load of different stages of gearbox change from time to time, bringing great challenges to the condition monitoring of gearbox. Cost performance is another factor that should be taken in to account in gearbox condition monitoring. Compared with the steam turbine in a thermal power plant, the building cost for each wind turbine is relatively low, and the price for the condition monitoring system must be acceptable. Comprehensive introductions and analyses of condition monitoring methods for different components of wind turbine have been published [2–4]. Temperatures are important and easily measured indicators of the health of many wind turbine component such as the gearbox and are often recorded automatically by the SCADA system. An unexpected increase in component temperature may indicate an overload, poor lubrication or possibly ineffective passive or active cooling [5]. Previous work used a Back Propagation (BP) neural network to construct the normal behavior temperature models of gearbox based on SCADA data [6]. When the residual between the model prediction and the measured value becomes very large, a potential fault is identified. In [7], the authors proposed a method using a Multiple Layer Perception (MLP) to build a temperature model of the gearbox. When the measured temperature value increases and is outside the confidence range of the value, a fault is registered. However, the Artificial Neural Network (ANN) including BP and MLP has demerits of requiring a time consuming training process and there can be local minima problems that may limit the improvement of model accuracy. Several authors [8–11] have built a test rig for gearbox and generator. Wavelets are used to analyze the high-speed sampling vibration signals. However, a test rig is quite different from a real wind turbine. There are only limited numbers of acceleration sensors to measure the drive train and gearbox vibration to provide a magnitude alarm and the sample frequency (10 s or 10 min) is too slow to meet the high frequency demands of vibration analysis. This paper uses temperature trend analysis to monitor a gearbox's operating conditions. The Autoassociative Kernel Regression (AAKR) method is used to model the normal behavior of gearbox temperature and give temperature estimates. When the gearbox has an incipient failure, the temperature residual between the AAKR model estimate and real measurement will become significant. With moving window residual statistics, these incipient failures can be detected in a timely way.

The paper is arranged as follows: Section 2 provides an introduction to turbine gearbox and SCADA data. Section 3 explains how the AAKR gearbox temperature model is constructed. The fourth section focuses on the moving window residual statistical analysis method. In Section 5, the AAKR model is validated and two simulated gearbox failure cases are studied. The final section provides discussion and conclusions, including suggestions for further research.

2. Structure of a Wind Turbine Gearbox and the SCADA Parameters

The wind turbines studied in this paper are located at Zhangjiakou in Northern China. SCADA data covering the period 04/01/2006 to 24/12/2006 was available for these units. The turbines, manufactured by GE, are variable speed, with a rated power of 1.5 MW. The cut-in and cut out wind speed are respectively 3 m/s and 12 m/s. The rated rotating speeds for rotor and DFIG are 20 rpm and 1800 rpm, respectively. The ratio for the gearbox is 1:90. This wind turbine uses a three-stage gearbox. The first stage is a planetary gear, and the second and third stages are conventional parallel gears, as shown in Figure 1. There is a circulating lubricating oil system to lubricate and cool the gearbox.

Figure 1. Structure of a gearbox for a wind turbine.



The SCADA system at the wind farm records all wind turbine parameters every 10 min. Each record includes a time stamp, output power, stator current and voltage, wind speed, ambient and nacelle temperature, gearbox temperature amongst many others; in total 47 parameters are recorded. At the same time, the SCADA system keeps a record of wind turbine operation and fault information, such as start up, shutdown, generator over temperature, pitch system fault, *etc*. Each fault record includes a time stamp, state code, and fault information. For example: at 2:28, 02/04/2006, the state code is 77 indicating that a gearbox oil over temperature alarm occurred and wind turbine shut down. In the wind turbine operation handbook, if the gearbox oil temperature is above 80 Celsius and this condition lasts for 60 s, the wind turbine will be shut down. When the oil temperature cools below 65 Celsius, the wind turbine will start up again.

3. Gearbox Temperature AAKR Model and Estimate

3.1. Autoassociative Kernel Regression Model Construction

In this paper, the Autoassociative Kernel Regression (AAKR) method is used to model the normal behavior of gearbox temperature. AAKR is based upon the multivariate inferential kernel regression derived by Wand and Jones in [12]. It is now mainly used in nuclear power plant sensor calibration [13].

AAKR is a non-parametric, empirical modeling technique that uses historical, fault free observations to estimate the output of a process or device. Let there be n variables of interest in a process or device. At time i, a single observation of the variables can be written as an observation vector:

$$\boldsymbol{X}(i) = \begin{bmatrix} \boldsymbol{X}_1(i) & \boldsymbol{X}_2(i) & \cdots & \boldsymbol{X}_n(i) \end{bmatrix}^{\mathrm{T}} \tag{1}$$

Construction of a memory matrix $\boldsymbol{D}$ is the first step of AAKR modeling. In a period of normal operation of the process or device, m historical observation vectors are collected covering the range of different operating conditions (such as high or low load, start up, before shut down, *etc.*) to construct the memory matrix, denoted as:

$$\mathbf{D} = \begin{bmatrix} \mathbf{X}(1) & \mathbf{X}(2) & \cdots & \mathbf{X}(m) \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1(1) & \mathbf{X}_1(2) & \cdots & \mathbf{X}_1(m) \\ \mathbf{X}_2(1) & \mathbf{X}_2(2) & \cdots & \mathbf{X}_2(m) \\ \vdots & \vdots & & \vdots \\ \mathbf{X}_n(1) & \mathbf{X}_n(2) & \cdots & \mathbf{X}_n(m) \end{bmatrix} \tag{2}$$

Each observation vector in the memory matrix represents a measured operating state of the process or device. With proper selection of the m historical observation vectors from an extended period of normal (un-faulted) operation of the process or device, the subset space spanned by the memory matrix $\boldsymbol{D}$ can be taken to represent the whole normal working space of the process or device. The construction of memory matrix is actually the procedure of learning and memorizing the normal behavior of the process or device analogous but different from the training of ANN.

During subsequent operation, the input to AAKR at each time step is a new observation vector $\boldsymbol{X}_{\text{obs}}$, and the output from AAKR is an estimate $\boldsymbol{X}_{\text{est}}$ for this input vector for the same moment in time. First, the distance between the observation vectors $\boldsymbol{X}_{\text{obs}}$ and each vector $\boldsymbol{X}(i)$ in the memory matrix is computed. There are several distance functions that may be used [14], but the most commonly used function is the Euclidean distance, as follows:

$$d_i(\boldsymbol{X}(i), \boldsymbol{X}_{\text{obs}}) = \sqrt{(\boldsymbol{X}_1(i) - \boldsymbol{X}_{1\text{obs}})^2 + (\boldsymbol{X}_2(i) - \boldsymbol{X}_{2\text{obs}})^2 + \cdots + (\boldsymbol{X}_n(i) - \boldsymbol{X}_{n\text{obs}})^2} \tag{3}$$

For the new observation vector, this calculation is repeated for each vector in the memory matrix, and resulting in a $m \times 1$ vector of distances:

$$\boldsymbol{d} = [d_1 \quad d_2 \quad \cdots \quad d_m]^{\mathrm{T}} \tag{4}$$

Next, these distances are transformed to similarity measures used to determine weights by evaluating the Gaussian kernel, expressed by:

$$\boldsymbol{W} = [w_1 \quad w_2 \quad \cdots \quad w_m]^{\mathrm{T}} = K_h(\boldsymbol{d}) = \frac{1}{\sqrt{2\pi h^2}} e^{-\boldsymbol{d}^2/h^2} \tag{5}$$

where h is the kernel bandwidth, $\boldsymbol{W}$ are the weights for the m vectors in the memory matrix. Finally, these weights are combined with the memory matrix $\boldsymbol{D}$ to make estimate according to:

$$\boldsymbol{X}_{\text{est}} = \frac{\sum_{i=1}^{m} (w_i \boldsymbol{X}_i)}{\sum_{i=1}^{m} w_i} \tag{6}$$

If the scalar a is defined as the sum of the weights, *i.e.*,

$$a = \sum_{i-1}^{m} w_i \tag{7}$$

Then Equation (6) can be represented in a more compact matrix notation:

$$\boldsymbol{X}_{\text{est}} = \frac{\boldsymbol{DW}}{a} \tag{8}$$

When the process or device works normally, the input observation vectors of the AAKR are most likely to be located in the normal working space that is represented by the memory matrix $\boldsymbol{D}$, in that it is similar to some historically measured vectors in the memory matrix. As a result, the estimate of AAKR will have a very high accuracy. When problems arise with the process or device, its dynamic characteristics will change, and the new observation vector will deviate from the normal working space. In this case the linear combination of the historical vectors in the memory matrix will not provide an accurate estimate of the input and the residual will increase in magnitude.

3.2. The Variable Selection for Gearbox Temperature AAKR Model

In order to construct the AAKR model of gearbox temperature, the variables included in the observation vector should be carefully chosen. Because we are concerned with the gearbox temperature, variables that have close relationship with gearbox temperature should be taken into account. Following a review of the 47 variables recorded by the SCADA system, the following five variables were selected to construct the observation vector.

(1) Power (P); power has a great influence on the gearbox temperature. When the power is high, the gearbox will endure a high load which leads to high gearbox temperatures.

(2) Wind speed (V); Wind is the energy source for wind turbines. For the variable speed wind turbine, in order to reach maximum power point tracking, the drive train rotating speed is proportional to the wind speed. The higher the wind speed, the higher the rotation speed of the gearbox which also leads to high gearbox temperatures.

(3) Ambient temperature (T); Because the local temperature that the wind turbine experiences changes greatly in the short term (from day to night for example) and in the longer term (weeks to months) due to passing weather systems and seasons it must be taken into account explicitly. At Zhangjiakou during March and April, ambient temperature changes can be as large as 30 Celsius because of fast moving weather fronts.

(4) Gearbox Temperature itself (T_{G}).

(5) Gearbox Temperature of last time (T_{G1}). Because the gearbox is a closed structure and the gearbox temperature has big inertia, the gearbox temperature of the last sampling time has an important influence on the current moment gearbox temperature.

The time span for the gearbox temperature model should take the local climate into account. The wind turbine studied in the paper is located at the Zhangjiakou area, north of Beijing, and the ambient temperature and wind speed distribution are quite different from season to season, as shown in Figure 2 and Figure 3. Such meteorological parameters have a great influence on the operation of the gearbox. For example, in winter, the temperature can be as low as minus 20 Celsius and frequently results in gearbox lubrication problems. If the normal behavior AAKR model covers a long period such as from January to July, it accuracy will be quite low for two reasons. One reason is that in different seasons the gearbox operating conditions are quite different as the result of these meteorological parameters. The other reason is that if the time span for model is too long. The range of each above variable will be very large that also leads to low model accuracy. In order to get satisfactory model accuracy, the time span for AAKR should be one month or one season.

Figure 2. Ambient temperature in different seasons.

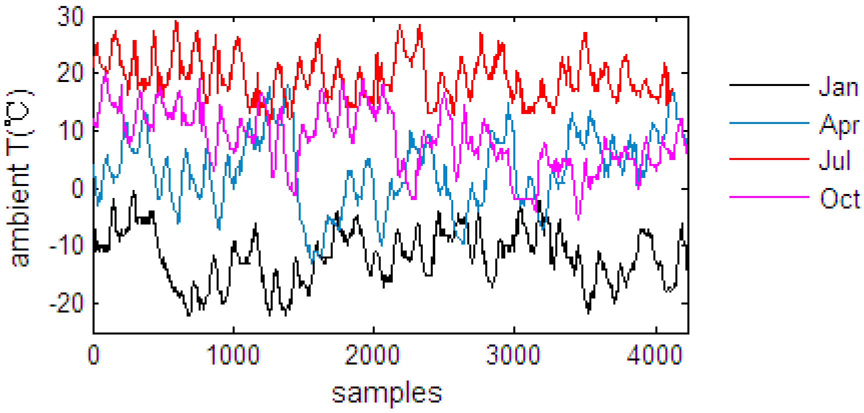
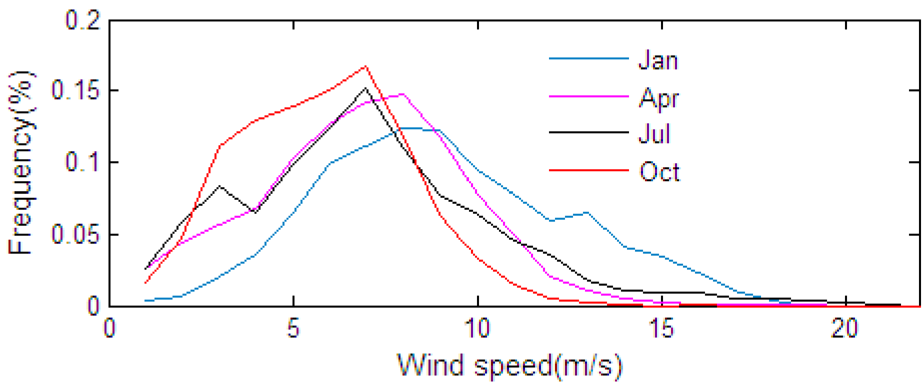
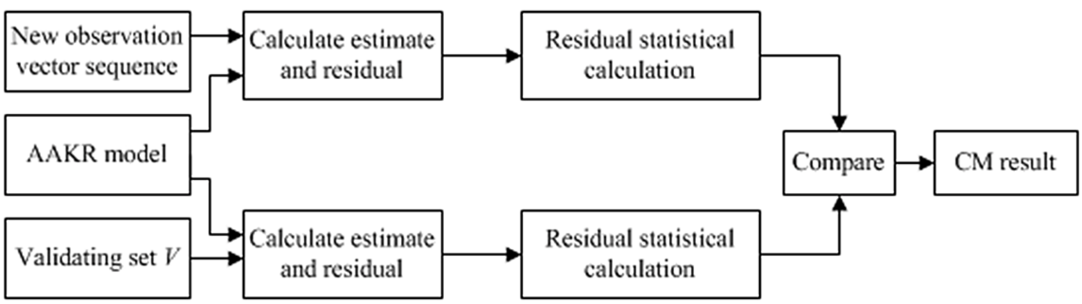


Figure 3. Wind speed distribution in different seasons.



In this paper, the time span for the gearbox temperature AAKR model is one month. After the construction of the AAKR model, model estimates and residual analysis to monitor the gearbox operating condition can be carried out as shown in Figure 4.

Figure 4. Gearbox condition monitoring using the AAKR model.



3.3. Construction of Memory Matrix **D**

Because the five variables in the observation vector have different units and their absolute values are quite different, the initial values of the five variables are rescaled to the range [0 1] according to their maximum and minimum values. After rescaling, each variable has same weight in the calculation of Euclidean distance. Especially for wind speed, because the wind turbine's cut-in and cut-out wind speeds are 3 m/s and 12 m/s, respectively, wind speeds below 3 m/s and above 12 m/s should be set as 3 m/s and 12 m/s before scaling.

In previous work [15], two algorithms are used to extract observation vectors from the normal working period to construct the memory matrix. In the first algorithm, vectors are selected that correspond to the extreme normal working states. For each variable in the observation vector, the algorithm finds the minimum and maximum measurements from the normal working period. The observation vectors containing these measurements are added to the memory matrix. In the second algorithm, the remaining observation vectors from the normal working period are ordered by their Euclidean norms. For n dimensional vector $\boldsymbol{X}$, the Euclidean norm is:

$$\|\boldsymbol{X}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{9}$$

The algorithm then selects evenly spaced elements from the ordered set and adds their corresponding observation vectors to the memory matrix. This construction method is simple, but it has some problems. Observation vectors can exist with Euclidean norm values that are similar or even exactly the same, but the vectors may be quite completely different, such as for the following two vectors:

$$\boldsymbol{X}_1 = [1 \quad 0 \quad 0 \quad 0 \quad 0]^{\mathrm{T}} \text{ and } \boldsymbol{X}_2 = [0 \quad 0 \quad 1 \quad 0 \quad 0]^{\mathrm{T}}$$

Selecting one of such similar Euclidean norm equivalent vectors will result in other equivalent norm vectors being discarded. Then only the working space near the selected observation vector is covered by the memory matrix, while parts of the parameter space near the discarded ones will be neglected. In order to minimize this effect a new memory matrix construction method is proposed that builds on the method used in [15] to give much improved NSET model accuracy. The new method is as follows.

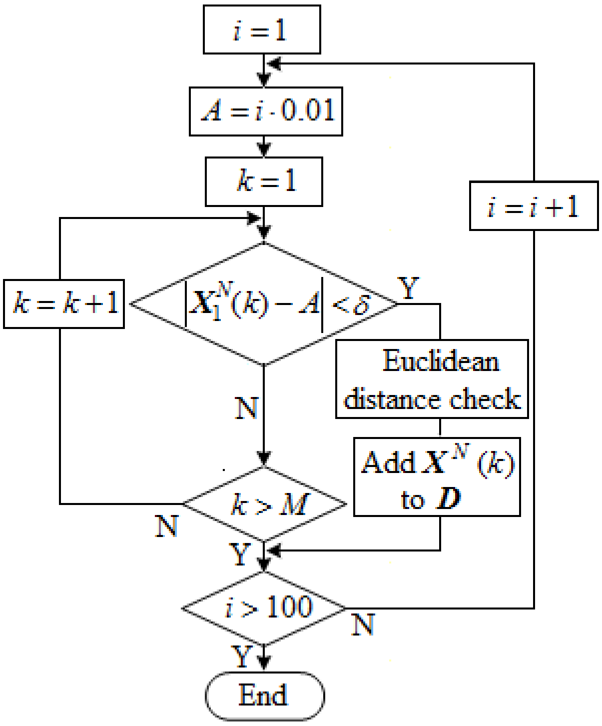
Assume the historical observation vectors during the normal working period are $\mathbf{X}^N(1), \mathbf{X}^N(2), \cdots,$ $\mathbf{X}^N(M)$ and make the matrix:

$$\mathbf{K} = \begin{bmatrix} \mathbf{X}^N(1) & \mathbf{X}^N(2) & \cdots & \mathbf{X}^N(M) \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1^N(1) & \mathbf{X}_1^N(2) & \cdots & \mathbf{X}_1^N(M) \\ \mathbf{X}_2^N(1) & \mathbf{X}_2^N(2) & \cdots & \mathbf{X}_2^N(M) \\ \vdots & \vdots & & \vdots \\ \mathbf{X}_5^N(1) & \mathbf{X}_5^N(2) & \cdots & \mathbf{X}_5^N(M) \end{bmatrix}_{5\times M} \tag{10}$$

The number of vectors in matrix $\boldsymbol{K}$ is M. Each observation vector includes power, wind speed, ambient temperature, gearbox temperature, gearbox temperature of last time, denoted $x_1, x_2, \cdots, x_5$ respectively, and n in Equation (1) takes the value $n = 5$.

Every observation vector in the normal working space covers the five variables and is normalized as described above. In order to ensure that memory matrix $\boldsymbol{D}$ covers the vectors with different variable values in the normal working space, for each of the five variables, the range [0 1] is equally divided into 100 sections and the observation vectors from matrix $\boldsymbol{K}$ selected at steps of 0.01. Observation vectors at each step of 0.01 for each variable in turn are added into memory matrix $\boldsymbol{D}$ so long as the variable in question is sufficiently close to the value in the observation vector. For example, for variable x_1 (turbine power), the method of adding observation vectors to $\boldsymbol{D}$ is shown in Figure 5.

Figure 5. Construction of memory matrix ***D*** for variable x_1.



In Figure 5, δ is a small positive number, taken to be 0.001 in this work. With this method, observation vectors with different variable values can be included in the memory matrix ***D***. Memory matrix construction thus has two steps: the first step uses the method in [15] to select observation vectors from the normal working period. The second step uses the new method outlined in this paper to add further observation vectors to ***D***. Before adding a new vector to the memory matrix, the Euclidean distances between the new potential vector and the vectors already in the memory matrix are checked. If the distance is too small, that means the memory matrix already has a vector that is quite similar to the one being considered and consequently it should not be added into the memory matrix. This check will limit the size of memory matrix and ensure that it is well-conditioned.

After the memory matrix construction is complete, the AAKR model is ready to estimate the gearbox temperature for the current moment measurement. For comparison, the reader is referred to the performance of a three-layer neural network for SCADA data modeling shown in [16]. It is quite complex and time consuming.

4. Gearbox Temperature AAKR Model Residual Statistical Analysis

4.1. Moving Window Calculation of Residual Mean Value and Standard Deviation Statistics

When the wind turbine gearbox suffers from some abnormality, the new observation vector will deviate from the normal working space and the distribution and time development of the estimate residual will change significantly from the normal condition. The mean value and standard deviation will reflect a change in the distribution of the residuals. In order to detect the changes in the variables in a timely manner, a moving average calculation is used. At a certain instant, the residual sequence of gearbox temperature from the AAKR model is:

$$\boldsymbol{\varepsilon}_{\mathrm{GT}} = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_N \quad \cdots],\ \varepsilon_i = x_4 - \hat{x}_4 \tag{11}$$

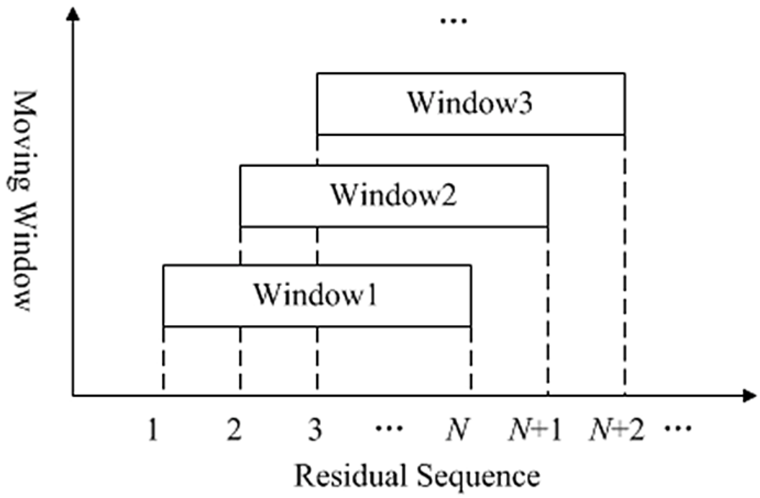
where x_4 is the gearbox temperature measurement in the observation vectors, and $\hat{x}_4$ is the AAKR model estimate for x_4.

A time window with width N is adopted to calculate the moving average or mean value and standard deviation for the N successive residuals in the window:

$$\overline{X}_\varepsilon = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i\ ,\ S_\varepsilon = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\varepsilon_i - \overline{X}_\varepsilon\right)^2} \tag{12}$$

The moving window is shown in Figure 6.

Figure 6. Residual moving average window used in residual analysis.



4.2. Residual Statistical Distribution when Gearbox Has an Abnormality

When the gearbox works normally, the AAKR model provides very accurate estimates of gearbox temperature. The residual sequence has a mean value near zero and the standard deviation is small. When a problem occurs with the gearbox, the new observation vector may deviate from the normal condition and the gearbox temperature residual distribution will thus also change. Abnormal gearbox operation can be identified as follows:

(1) The residual mean value remains near zero, but the standard deviation increases dramatically. In this condition, the distribution of the residuals becomes wider.

(2) The residual mean value deviates from zero by some obvious magnitude and the standard deviation remains small. In this condition, the residual systematically departs from zero.

(3) A combination of the above two situations.

In order to detect early gearbox faults, failure thresholds are needed for both the residual mean value and its standard deviation. Assume that the thresholds for them are set respectively as E_Y and S_Y where these are determined according to operator experience or determined through model validation as presented below.

The residual sequence has been obtained using observation vectors from the validation set as input to the already identified AAKR model with subsequent application of the moving average window to the time sequence of residuals. By trial and error the optimal size of the moving window is determined. The maximum absolute mean value and standard deviation for validating set residual sequence are, as defined above, respectively E_V and S_V but note that these will be dependent on averaging window size. Then the thresholds for gearbox failure detection are as follows:

$$E_Y = \pm k_1 \cdot E_V, \; S_Y = k_2 \cdot S_V \tag{13}$$

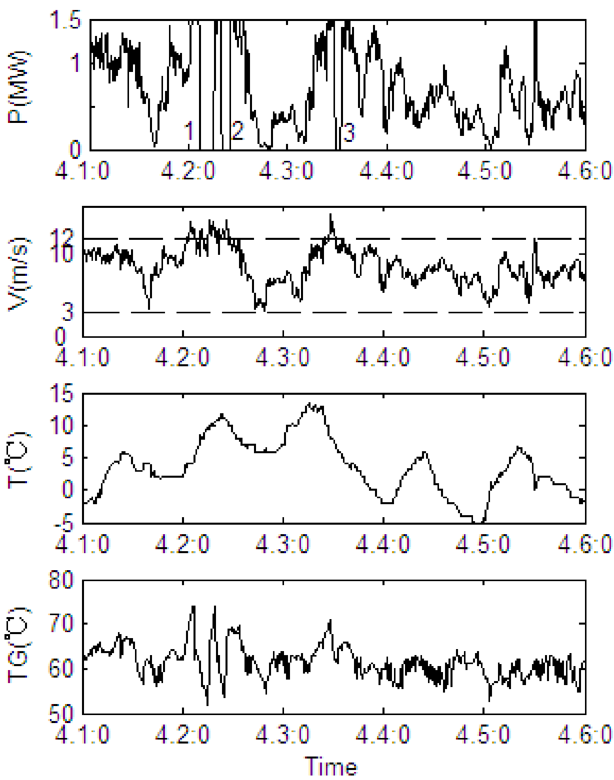
where k_1 and k_2 are positive coefficients and can be chosen based on operator experience. Relatively larger k_1 and k_2 will increase the robustness of CM method and reduce the false alarms. In this paper, $k_1 = 3$ and $k_2 = 2$.

5. Gearbox AAKR Model Validation and Failure Case Studies

5.1. Gearbox AAKR Model Validation

SCADA data of one wind turbine for April 2006 is used in this paper. The 10-min data trends of the five related variables mentioned in Section 3.2 from 01/04/2006 to 06/04/2006 are shown in Figure 7.

Figure 7. Trends from 01/04/2006 to 06/04/2006.



During this period, three shutdowns of this wind turbine occurred, as shown in Figure 7 and Table 1.

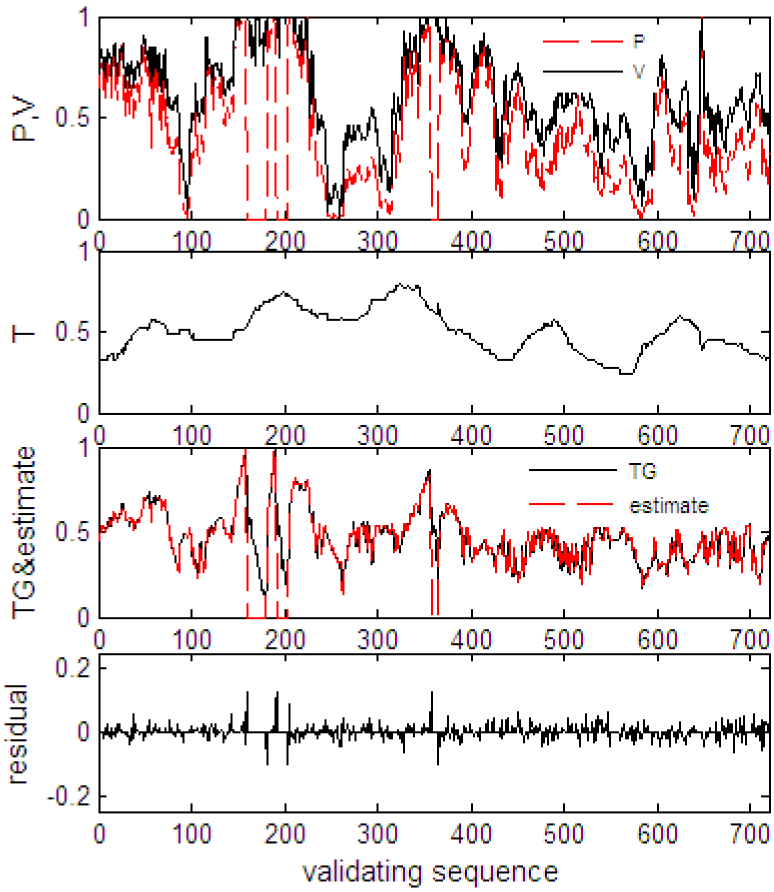
Table 1. Shutdown records from 01/04 to 06/04.

	Date	Time	State code	Reason
1	2006/04/02	2:28:43	77	Gearbox oil over temperature
2	2006/04/02	7:42:13	77	Gearbox oil over temperature
3	2006/04/03	11:14:35	147	Manual shutdown

But from the maintenance log, there was no record indicating that the gearbox had a failure or had been repaired. It is quite usual for the wind turbine to shut down due to some parameter alarms. During April 2006, the gearbox worked properly. The total 10-min SCADA data point number for April is 4320, and among them, the useful data number (data with power above zero) is 3731. Normal working data from 07/04/2006 to 30/04/2006 is used to construct the gearbox temperature normal behavior AAKR model. Data from 01/04/2006 to 06/04/2006 shown in Figure 7 is used as the validating set to validate the AAKR model.

In April 2006, the maximum and minimum gearbox temperatures were 74.1 Celsius and 50.2 Celsius, respectively. The maximum and minimum ambient temperatures were 20 Celsius and 13 Celsius, respectively. The validation result is shown in Figure 8. Values for different variables are scaled values. When the turbine power in an observation vector is zero, the AAKR model will not work, resulting in zeros for estimate and residual.

Figure 8. Validation results for the AAKR model.



In Figure 8, at some isolated positions, the residuals are much larger than other positions. These large residuals are in pairs, and the number of the large residual pairs is three. The positions and reasons for them are as given in Table 2.

Table 2. Large residuals' positions and reasons.

Pair	Position	Time	Residual	Reasons
1	160	04/02 2:30	0.103	2006/04/02 2:28:43 shutdown due to gearbox oil over temperature
	182	04/02 6:10	−0.124	2006/04/02 5:58:56 system OK startup
2	192	04/02 7:50	0.1205	2006/04/02 7:42:13 shutdown due to gearbox oil over temperature
	205	04/02 10:00	−0.1189	2006/04/02 9:56:56 system OK startup
3	357	04/03 11:20	0.1134	2006/04/03 11:14:35 manual stop
	366	04/03 12:50	−0.1187	2006/04/03 12:43:03 system OK startup

At the above positions in Table 2 where the wind turbine would shut down or just start up, the relationship between the five variables in the observation vectors is quite different from the normal working condition (for example, at position 160, the wind speed is high, while the power near zero due to shutdown), and the observation vectors at these positions deviate from the normal working space represented by the memory matrix ***D***, the combination of the historical observations in the memory matrix cannot provide an accurate estimate for the gearbox temperature at these positions and the residuals become quite large. These large residuals caused by the shutdown or startup of the wind turbine in some extent prove that the AAKR model is working and they should not be recognized as an indication of incipient gearbox failure.

After removing the above six large residuals, the AAKR model has a very high estimate accuracy, and most absolute residual values are below 0.05. The validation results shows that AAKR has satisfied the modeling accuracy for gearbox temperature dynamic characteristics.

The moving window calculation outlined in Section 4.1 is applied to the residual sequence of the validation set in Figure 8. The window width N should be properly selected so that the influence of the occasional isolated large residual caused by the imperfect coverage ability of memory matrix ***D*** can be minimized. At the same time, a moving window with a properly selected N must be able to detect the changes of mean value and standard deviation in a quick and effective manner. In this paper, the window width $N = 50$ reflects a useful balance between these two conflicting requirements. Figure 9 shows the trends of the mean value and standard deviation for the validating set after moving average filtering. It should be noticed that the six large residuals caused by turbine shutdown and startup in Table 2 have been removed during moving window averaging.

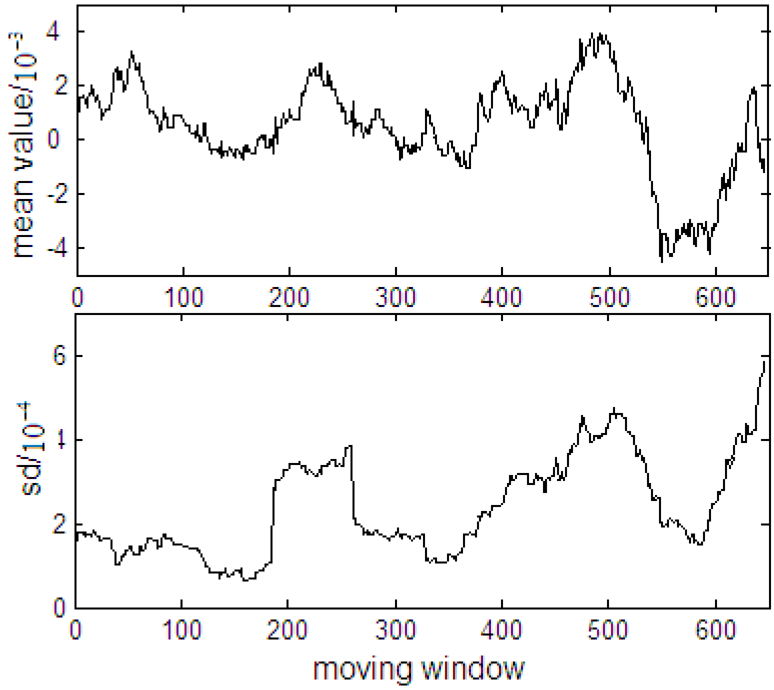
From the trends, we estimate:

$$E_{\mathrm{V}} = 3.9\times10^{-3},\ S_{\mathrm{V}} = 5.9\times10^{-4} \tag{14}$$

and the thresholds for the mean value and standard deviation are chosen respectively as:

$$E_{\mathrm{Y}} = \pm1.17\times10^{-2},\ S_{\mathrm{Y}} = 1.18\times10^{-3} \tag{15}$$

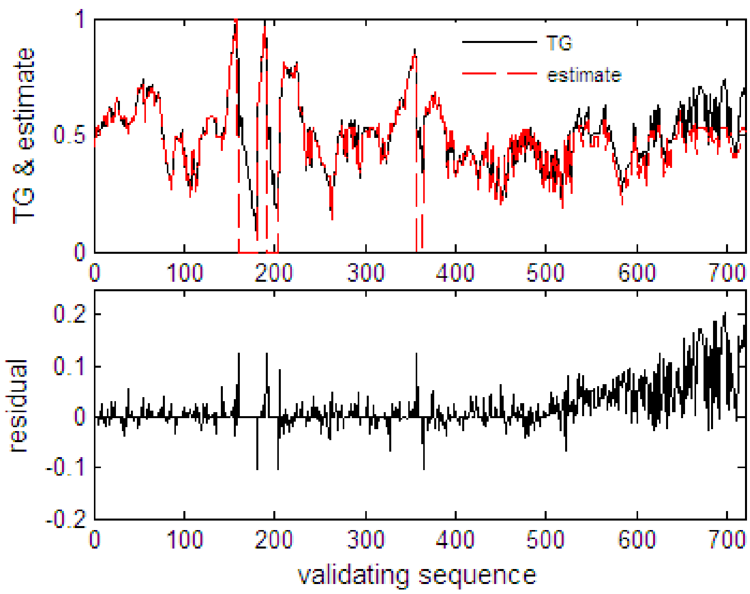
Figure 9. Statistical characteristics for the validation sequence.



5.2. Gearbox Failure Analysis

Actual failure of major components like gearboxes is relatively rare and sometimes it is also very difficult to access these components' failure data and maintenance records as the operator or manufacturer usually keep them as commercial secrets. From [5–7], we know that abnormal temperature changes or rises are an effective indication of incipient gearbox failure. In [17], authors gives a proof that the gearbox temperature rise will be proportional to the power output if the gearbox works normally, that is, the gearbox transmission efficiency has not changes.

Figure 10. AAKR model estimate and residual in Case 1.

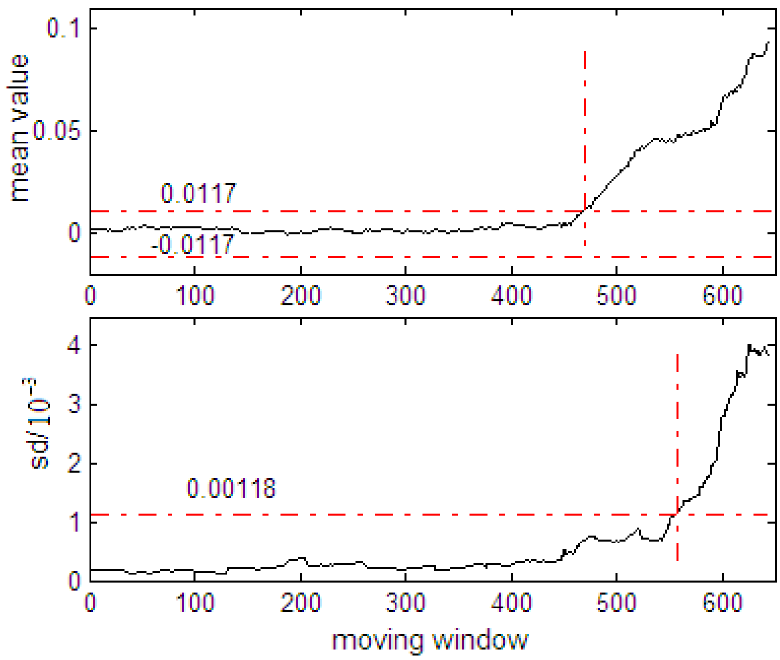


When an incipient fault occurs in the gearbox, it efficiency decreases, and the gearbox temperature will have extra increase. Unfortunately, in the SCADA data we have, there is no gearbox failure record. Based on [5–7,17], it is representive to simulate a real incipient gearbox failure by adding extra temperature drift on the initial SCADA data. And these manual drift data is used to test the effectiveness of this AAKR CM method in the flowing two cases.

Case 1:

A fixed temperature rise of 0.001 per time step is added at position 501 to the initial gearbox temperature from 01/04/2006 to 06/04/2006 in the validation set. The AAKR model estimate and the residual sequence are shown in Figure 10. The trends for mean value and standard deviation for the above residual sequence are shown in Figure 11.

Figure 11. Statistical characteristics for Case 1.

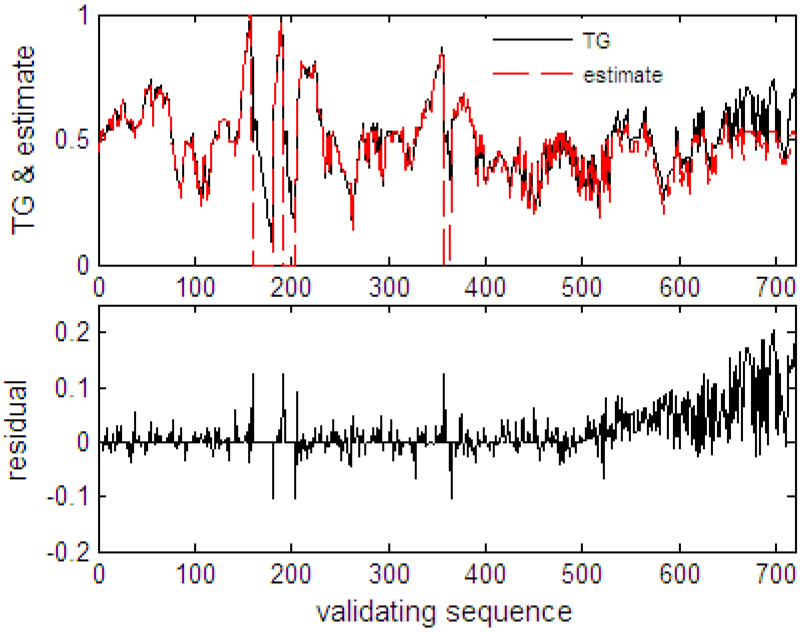


In the trend for mean value, at position 471, the mean value exceeds the failure threshold. Because the moving window width is $N = 50$, the new CM method detects the simulated gearbox failure at time point of 471 + 50 = 521, and only 521 − 501 = 20 points after the manual temperature drift began being added at time point 501. Before the simulated gearbox failure is detected, the absolute temperature drift can be calculated by 20 × 0.001 × (74.1 − 50.2) = 0.478 Celsius where 74.1 and 50.2 are the gearbox maximum and minimum temperatures in April 2006, respectively. From this studied case, we can see that the AAKR condition monitoring method can detect in a timely fashion the abnormal changes in the gearbox temperature and give incipient failure alarms.

Case 2:

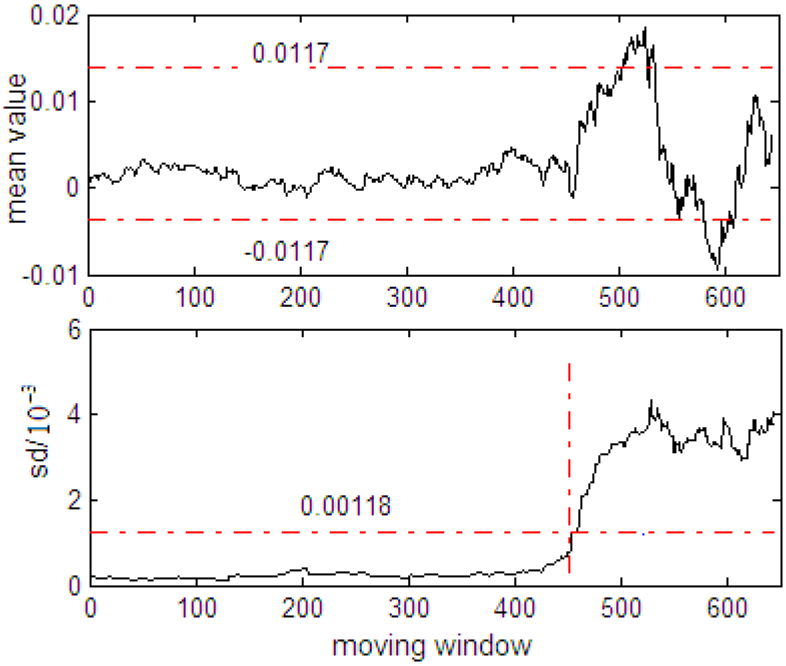
In studied case 2, white noise with variance 0.01 is added from position 501 to the initial validation set with SCADA data from 01/04/2006 to 06/04/2006. The AAKR model estimate and residual are shown in Figure 12.

Figure 12. AAKR model estimate and residual in Case 2.



The trends of mean value and standard deviation for case 2 are shown in Figure 13.

Figure 13. Statistical characteristics for Case 2.



In Figure 13, the standard deviation exceeds the failure threshold at position 462, and only 462 + 50 − 501 = 11 points after the white noise is added. In this case, the AAKR CM method also detects the simulated abnormality on time.

6. Discussion and Conclusions

This paper uses Autoassociative Kernel Regression (AAKR) to construct a gearbox temperature model. Details of how to construct the required memory matrix are provided. The method has the advantage of being computationally remarkably simple and should thus have immediate appeal for wind industry practitioners. When the gearbox experiences a fault, new observation vectors will generally deviate from the normal working space and the AAKR estimate of the residual distribution and its time development will change. A moving average window filter has been adopted to reveal the trends in mean value and standard deviation of the residual sequence; this is also very easy to implement computationally. Gearbox condition can be determined from crossing of predetermined thresholds. With the effective selection of the memory matrix from data from normal operation, the presented method can identify simulated wind turbine gearbox failures in a timely way. With suitable selection of the parameters available to the user (window size, vector parameter selection, thresholds *etc.*), the method should prove useful for wind turbine condition monitoring more widely. There is still a small flaw in this paper because the validation was carried out with the man-made gearbox failure data. In future research, real failure data will be collected and this CM method further validated.

Acknowledgments

This paper was supported by the Fundamental Research Funds for the Chinese Central University under Grant 09MG18 and National Natural Science Foundation of China under Grant 61074094.

References

1. Crabtree, C.J.; Feng, Y.; Tavner, P.J. Detecting Incipient Wind turbine Gearbox Failure: A Signal Analysis Method for On-Line Condition Monitoring. In *Proceedings of the European Wind Energy Conference*, Warsaw, Poland, 20–23 April 2010.
2. Hameed, Z.; Hong, Y.S.; Cho, Y.M.; Song, C.K. Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renew. Sustain. Energy Rev.* **2009**, *1*, 1–39.
3. Amirat, Y.; Benbouzid, M.E.H.; Al-Ahmar, E.; Bensaker, B.; Turri, S. A brief status on condition monitoring and fault diagnosis in wind energy conversion systems. *Renew. Sustain. Energy Rev.* **2009**, *9*, 2629–2636.
4. Lu, B.; Li, Y.; Wu, X. A Review of Recent Advance in Wind Turbine Condition Monitoring and Fault Diagnosis. In *Proceedings of 2009 IEEE Power Electronics and Machines in Wind Application*, Lincoln, NE, USA, 24–26 June 2009; pp. 1–7.
5. Tavner, P.; Ran, L.; Penman, J.; Sedding, S. *Condition Monitoring of Rotating Electrical Machines*, 2nd revised ed.; Institute of Engineering and Technology: London, UK, 2008; pp. 133–134.

6. Zaher, A.S.A.E.; McArther, S.D.J.; Infield, D.G. Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy* **2009**, *6*, 574–593.
7. Garcia, M.C.; Sanz-Bobi, M.A.; del Pico, J. SIMAP: Intelligent system for predictive maintenance: Application to health condition monitoring of a wind turbine gearbox. *Comput. Ind.* **2006**, *6*, 552–568.
8. Yang, W.; Tavner, P.J.; Crabtree, C.J.; Wilkinson, M. Cost- effective condition monitoring for wind turbines. *IEEE Trans. Ind. Electron.* **2010**, *1*, 263–271.
9. Watson, S.J.; Xiang, J.P.; Yang, W.; Tavner, P.J.; Crabtree, C.J. Condition monitoring of the power output of wind turbine generators using wavelets. *IEEE Trans. Energy Convers.* **2010**, *3*, 715–721.
10. Yang, W.; Tavner, P.J.; Wilkinson, M.R. Condition monitoring and fault diagnosis of a wind turbine synchronous generator drive train. *Renew. Power Gener.* **2009**, *1*, 1–11.
11. Wilkinson, M.R.; Spinato, F.; Tavner, P.J. Condition Monitoring of Generators & Other Subassemblies in Wind Turbine Drive Trains. In *Proceedings of IEEE International Symposium on Diagnostics for Electric Machines, Power Electrics and Drives*, Cracow, Poland, 6–8 September 2007; pp. 388–392.
12. Wand, M.P.; Jones, M.C. *Kernel Smoothing*; Chapman & Hall: London, 1995; Monographs on Statistics and Applied Probability, Volume 60.
13. Hines, W.J.; Garvey, D. Process and equipment monitoring methodologies applied to sensor calibration monitoring. *Qual. Reliab. Eng. Int.* **2007**, *23*, 123–135.
14. Black, C.L.; Uhrig, R.E.; Hines, J.W. System Modeling and Instrument Calibration Verification with a Nonlinear State Estimate Technique. In *Proceedings of Maintenance and Reliable Conference*, Knoxville, TN, USA, 12–14 May 1998.
15. Bockhorst, F.K.; Gross, K.C.; Herzog, J.P.; Wegerich, S.W. MSET Modeling of Crystal River-3 Venturi Flow Meters. In *Proceedings of 6th International Conference on Nuclear Engineering*, San Diego, CA, USA, 10–15 May 1998.
16. Guo, P.; Infield, D.; Yang, X. Wind turbine generator condition monitoring using temperature trend analysis. *IEEE Trans. Sustain. Energy* **2011**, in press.
17. Feng, Y.; Qiu, Y.; Crabtree, C.J.; Long, H.; Tavner, P.J. Use of SCADA and CMS Signals for Failure Detection and Diagnosis of a Wind Turbine Gearbox. In *Proceedings of Europe Premier Wind Energy Conference*, Brussels, Belgium, 14–17 March 2011.