An integrated data-driven model-based approach to condition monitoring of the wind turbine gearbox

Peng Qian, Xiandong Ma*, Philip Cross
Engineering Department, Lancaster University, Lancaster LA1 4YW, United Kingdom
Emails: p.qian@lancaster.ac.uk, xiandong.ma@lancaster.ac.uk

Abstract: Condition Monitoring (CM) is considered an effective method to improve the reliability of wind turbines and implement cost-effective maintenance. This paper presents a single hidden-layer feed forward neural network (SLFN), trained using an extreme learning machine (ELM) algorithm, for condition monitoring of wind turbines. Gradient-based algorithms are commonly used to train SLFNs; however, these algorithms are slow and may become trapped in local optima. The use of an ELM algorithm can dramatically reduce learning time and overcome issues associated with local optima. In this paper, the ELM model is optimized using a genetic algorithm. The residual signal obtained by comparing the model and actual output is analyzed using the Mahalanobis distance measure due to its ability to capture correlations among multiple variables. An accumulated Mahalanobis distance value, obtained from a range of components, is used to evaluate the heath of a gearbox, one of the critical subsystems of a wind turbine. Models have been identified from supervisory control and data acquisition (SCADA) data obtained from a working wind farm. The results show that the proposed training method is considerably faster than traditional techniques, and the proposed method can efficiently identify faults and the health condition of the gearbox in wind turbines.

1. Introduction

There has been a dramatic increase in the construction of wind farms over the past decade in UK, especially offshore wind installations, contributing to the UK achieving national targets for reducing CO_2 emissions and the production of sustainable energy. Compared to their onshore counterparts, the major advantages of offshore wind turbines (WTs) include increased turbine size, improved wind conditions due to higher wind speed and lower turbulence, and reduced visual impact and noise intrusion. However, the high cost of routine inspection and maintenance has been problematic, particularly when the WTs are operating in harsh environments and are sited in deep sea waters. Over an operating life of 20 years, maintenance costs of wind farm may reach 15% and 30% of the total income for onshore and offshore wind farms, respectively [1]. Condition monitoring (CM) is considered an effective method to schedule cost-effective maintenance activities and enhance the reliability of wind turbines [2]-[5]. Clearly, it is essential to develop effective CM techniques for wind turbines, providing information regarding the past and current condition of the turbines and to enable the optimal scheduling of maintenance tasks.

Among CM techniques, data-driven model-based methods (referred to as data-based methods thereafter in this paper) do not need to consider the mathematical model of the physical system; instead models are purely based on data obtained by investigating the relationship between measured inputs and outputs. In the data-based method, data gathered using a CM system or equivalent are used as the inputs for models predicting the output signals of a physical process. Actual output signals generated by the system are then compared to the predicted outputs for the corresponding input signals. Any differences between these output signals could be caused by changes to the system, and may be caused by the occurrence of a fault [6]. In this regard, the residual signal can provide an early warning of imminent component failure.

Although the residual signal can show impending component failure, it does not provide accurate details regarding the failure of components or subsystems in a wind turbine. One of the important aims of a CM system is to assist the operators to operate safely and reliably the wind turbines in order to avoid unnecessary operating outages. The outputs from such condition monitoring systems allow turbine operators to make decisions with regards to maintenance scheduling through improved understanding of the turbine's health condition. Reasonable maintenance strategies can therefore be implemented, which can significantly reduce the maintenance cost and enhance the availability and reliability of a wind turbine [7].

This paper proposes a new method for condition monitoring and fault diagnosis of the gearbox in the wind turbines. The faults associated with the gearbox account for a considerable proportion of total faults, which could contribute to approximately 20% of the downtime of a doubly-fed induction generator-based wind turbine, particularly for offshore wind farms [8]-[9]. For data-based condition monitoring systems, accurate models are essential for the relationships between those parameters being monitored. In this regard, artificial intelligence (AI) techniques are utilized by many researchers for data-based CM schemes, such as artificial neural networks (ANNs) [10]-[12], support vector machines (SVMs) [13]-[14] and fuzzy logic [15]-[17]. ANN-based methods are robust to signal noise, making them suitable for dealing with data acquired in noisy environments. However, the long training times associated with ANN models can limit their application. SVMs tend to have better generalized performance and more accurate training results than neural network models; however, training SVM models with large datasets is not straightforward. A fuzzy logic system, based on fuzzy sets of linguistic variables, uses predefined rules to enable reasoning. A fuzzy logic system is based upon fuzzified features of the faults and then uses these features to diagnose faults by using the predefined rules. It is clear that a fuzzy logic system requires full knowledge of failure mechanisms of a wind turbine in order to design these rules, which is usually unfeasible in practice. In this paper, an extreme learning machine (ELM) algorithm is employed to train a neural network model for data-based condition monitoring, overcoming the drawbacks of a traditional feedforward ANN. The preliminary results obtained by the authors of this paper using the ELM for condition monitoring of wind turbines can be found in reference [18]. In this paper, the ELM algorithm is firstly optimized by a genetic algorithm in order to optimize the initial weight values and the biases of the hidden neurons; then a classification method based on the accumulated value of the Mahalanobis distance (MD) from multiple components are used as the measure to assess the health condition of the wind turbine gearbox. The proposed method is able to integrate the optimized ELM algorithm with an appropriate classification method utilizing different components in the gearbox system, facilitating fast and reliable condition monitoring and fault diagnosis of the wind turbines.

The remainder of this paper is organized as follows. The working principle of the extreme learning machine algorithm is presented in Section 2, while Section 3 describes the genetic algorithm employed to optimize the ELM model. Section 4 demonstrates the Mahalanobis distance method and proposes an accumulated MD method in order to diagnose the health condition of a gearbox. Case studies using SCADA data obtained from a working wind farm are discussed in Section 5. Finally, Section 6 contains conclusions and suggestions for further research.

2. The extreme learning machine algorithm

Feed-forward neural networks with a single hidden layer (SLFNs) are particularly efficient and are used widely in several research areas, including mode recognition and state prediction [19-21]. Gradient-based back-propagation training algorithms, traditionally used during the learning procedure for a SLFN, have some disadvantages, which can cause long training times of the model during the learning process. Other issues include being stuck in local optima, improper learning rate, and over-fitting. In this regard, the extreme learning machine (ELM) algorithm was first proposed by Huang as a non-iterative algorithm to improve the learning process of a SLFN [22]. Compared with gradient-based learning methods, the ELM algorithm incorporates the following merits [23]-[24]:

(i) It arbitrarily initializes the weights on the input and the biases, and calculates analytically the weights on the output. Note that the output weights do not need be iterated repeatedly during training, resulting in faster learning than other algorithms.

(ii) Traditional gradient-based learning algorithms are iterative and may become trapped in local optima. Other problems include overtraining and overfitting. These issues may interfere

This article has been accepted for publication in a future issue of this journal, but has not been fully edited.
Content may change prior to final publication in an issue of the journal. To cite the paper please use the doi provided on the Digital Library page.

with the training result, especially when modelling a nonlinear system. In contrast, the ELM algorithm is better at the generalization of training, thus overcoming these issues.

Fig.1 shows a diagram of a feed forward neural network with a single hidden-layer. The network consists of an input layer, a hidden layer and an output layer of neurons. For this example, the input layer has n neurons; the hidden layer has L neurons, and the output layer has m neurons. Finally, $x_1, x_2, \cdots, x_n$ are the inputs to the network and $y_1, y_2, \cdots, y_m$ are the outputs from the network.

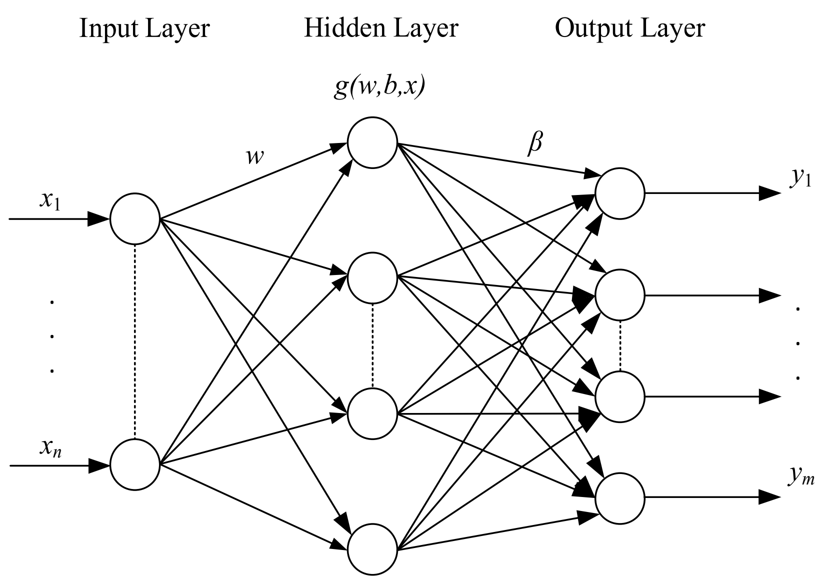


Fig. 1 *Diagram of a feedforward neural network with a single hidden-layer (SLFN)*

Consider an ELM based upon the network illustrated in Fig. 1 with an activation function $g(.)$. It is assumed that the ELM is able to estimate N training outputs with zero error. The algorithm can be represented by the following expression:

$$M = \begin{bmatrix} \sum_{i=1}^{L} \beta_{i1} g(w_{i1}x_1 + b_i) & \sum_{i=1}^{L} \beta_{i1} g(w_{i2}x_2 + b_i) & \dots & \sum_{i=1}^{L} \beta_{i1} g(w_{in}x_N + b_i) \\ \sum_{i=1}^{L} \beta_{i2} g(w_{i1}x_1 + b_i) & \sum_{i=1}^{L} \beta_{i2} g(w_{i2}x_2 + b_i) & \dots & \sum_{i=1}^{L} \beta_{i2} g(w_{in}x_N + b_i) \\ \dots & & & \\ \sum_{i=1}^{L} \beta_{im} g(w_{i1}x_1 + b_i) & \sum_{i=1}^{L} \beta_{im} g(w_{i2}x_2 + b_i) & \dots & \sum_{i=1}^{L} \beta_{im} g(w_{in}x_N + b_i) \end{bmatrix}_{m\times N} \quad (1)$$

where w_{ij} is the weight between the ith hidden neuron and jth input neuron; $\beta_i = [\beta_{i1} \quad \beta_{i2} \quad \cdots \quad \beta_{im}]$ is the vector of output weights connecting the ith hidden neuron and m output neurons; $x_j = [x_{1j} \quad x_{2j} \quad \cdots \quad x_{nj}]^T$ $(j = 1, 2, \dots, N)$ are the input signals; $b_i = [b_1 \quad b_2 \quad \cdots \quad b_L]^T$ is the bias of the ith hidden neuron.

Eq. (1) can be rewritten,

$$H\beta = M^T \tag{2}$$

in which M^T is the transpose of matrix M and H is the output matrix of the hidden layer. The matrix H can be represented as,

$$H = \begin{bmatrix} g(\sum_{j=1}^{n} w_{1j}x_1 + b_1) & g(\sum_{j=1}^{n} w_{2j}x_1 + b_2) & \dots & g(\sum_{j=1}^{n} w_{Lj}x_1 + b_L) \\ g(\sum_{j=1}^{n} w_{1j}x_2 + b_1) & g(\sum_{j=1}^{n} w_{2j}x_2 + b_2) & \dots & g(\sum_{j=1}^{n} w_{Lj}x_2 + b_L) \\ \dots & & & \\ g(\sum_{j=1}^{n} w_{1j}x_N + b_1) & g(\sum_{j=1}^{n} w_{2j}x_N + b_2) & \dots & g(\sum_{j=1}^{n} w_{Lj}x_N + b_L) \end{bmatrix}_{N\times L} \tag{3}$$

where the ith column of H is the vector of outputs of the ith hidden neuron given inputs x_1, x_2, · · ·, x_n. Following initialization of the input weight matrix w ($L\times n$ dimensions) and the hidden layer bias vector b (length L), the matrix H ($N\times L$ dimensions) is uniquely determined. The matrix of output weights, β ($L\times m$ dimensions), can then be calculated by simply finding a matrix $\hat{\beta}$ in order to minimize the error function,

$$\min_{\beta} \left\| H\beta - M^T \right\| \tag{4}$$

It is worth noting that the input weights w and the hidden layer biases b are not changed during this procedure. The solution is expressed as the following:

$$\hat{\beta} = H^{+} M^{T} \tag{5}$$

Minimizing this function is equivalent to obtaining the unique smallest norm least-squares solution of the linear system in eq. (4). The matrix H^+ is the generalized Moore-Penrose inverse of the matrix H, which can be found using the singular value decomposition (SVD) method. Details about the SVD method can be found in reference [25].

3. Genetic Algorithm Optimization

As described in Section 2, arbitrary values are assigned to the weights of the inputs and the biases of the hidden neurons of the ELM model at the beginning of learning; clearly these parameters may not be the optimum values for the ANN. However, the training results of the ELM model largely depend on both the input-to-hidden weights and hidden-to-output weights, hence the ANN tends to have better generalization performance given small values for the weights. The selection of optimal initial input weights and biases would therefore be essential for an effective ELM model. Thus, a genetic algorithm (GA) is adopted to optimize these weights and biases. GAs were originally proposed by Holland [26], and are a kind of parallel adaptive search algorithm based on the mechanics of natural selection and genetic systems, where individuals are usually represented by binary strings, as here. The algorithms have unique advantages, particularly in the fields of searching, optimization, and machine learning [27]. The purpose of using a genetic algorithm in this study is to obtain optimum values for the initial input weights and the initial hidden neuron biases so that the weights β can be calculated using eq. (5). In general, a genetic algorithm has five steps, including initialization, fitness evaluation, selection, crossover and mutation operations.

The purpose of the selection operation is to obtain the probability of an individual being able to contribute to the next generation, This is based upon each individual's 'fitness', in this case, the optimum values for the initial input weights and biases. In order to achieve this, a roulette wheel selection technique is employed in the GA. There needs to be a balance in order to maintain the selection pressure and the diversity of the population. The crossover operation obtains new individuals from two 'parents'. Here a kind two-point crossover is used where two points are chosen on the parent chromosome strings. Two child chromosomes are obtained by swapping the elements between two points on the parent binary strings. Finally, the mutation operation introduces a random element to the individuals of the population. The rate of mutation decreases exponentially as the number of generations increases. For each mutation, a random number is generated. If the random number is smaller than the mutation rate, the value of the bit is flipped; otherwise, the value remains the same. More details about the GA can be found in reference [28].

When the internal weights and biases are initialized, the ELM model calculates a predicted output. The fitness value can be found by calculating the sum of the absolute errors of the expected output and actual output of the ELM,

$$F = k(\sum_{i=1}^{m} |y_i - o_i|) \tag{6}$$

where m is the number of outputs; y_i is the ith predicted output of the ELM model; o_i is the ith actual output of ELM model; although k is an application dependent constant, k=1 is normally selected [29].

The steps of the optimal extreme learning machine incorporating a genetic algorithm are described as follows:

Step 1: Define the structure of the SLFN, including the number of input layer neurons and hidden layer neurons, n and L respectively; arbitrary initial values are assigned to input weights w and hidden neuron biases b.

Step 2: The input weights and hidden neuron biases are forwarded to the genetic algorithm. Through the five steps of the GA described above, optimal initial values of the input weights and biases are determined. It is worth emphasizing that when the input weights and biases are initialized, the optimal output weights are uniquely determined, as described in the above section; thus output weights need not to be optimized by the GA.

Step 3: The ELM model is then updated using the initial values of w and b. The model is subsequently trained with the training data, with the hidden-to-output weights β being adjusted until the output data from the model match the target output data.

Step 4: A set of input data are then used to test the model to observe how well the corresponding outputs are predicted. In this case, the output values are predicted signals of the process being modelled. The actual outputs are then compared with the model prediction for given input signals, and the residual signals between them are obtained.

4. Health Condition Identification

In this section, faults in a wind turbine gearbox are investigated by comparing the difference between the actual signal detected in real time and the predicted signal from the optimized extreme learning machine. Although a method relying on residual signals alone can detect faults effectively, it is not able to provide accurate characteristics about the failure of components. Furthermore, the gearbox in a wind turbine generally has several components, and traditional methods have only focused on detecting faults or identifying the health of an individual component [10]. Clearly, it would be desirable to use a more appropriate method in order to identify the health condition of the gearbox system as a whole.

A minimum-redundancy maximum-relevance feature approach is adopted in this paper to optimize the residual signal, taking into account interactions between signals measured from different components in the gearbox. The Mahalanobis distance (MD) is a measure of the distance between a point and a distribution without consideration of the units used for the measurement. This means that the MD measure has the capability to describe correlations among variables in a process or a system. Thus, the MD measure can provide a univariate distance value for multivariate data, which is ideal for estimating the deviation values of a complex system [30] [31]. Consequently, the MD measure is selected to help obtain the deviation from the group data, which can be used to identify the health condition of the gearbox. For the ith observation vectors $X_i = (x_{1i}, x_{2i}, ..., x_{ni})$ and $Y_i = (y_{1i}, y_{2i}, ..., y_{ni})$, the MD is given by matrix

$$MD = \sqrt{(X_i - Y_i)C^{-1}(X_i - Y_i)^T} \quad (7)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited.
Content may change prior to final publication in an issue of the journal. To cite the paper please use the doi provided on the Digital Library page.

where n is the number of parameters $x_1, x_2, \cdots, x_n$ to be analyzed, for example, the temperatures and pressures of oil in the gearbox; the matrix C is the covariance matrix of X_i and Y_i , i.e., $C = cov(X_i, Y_i)$, where cov is a function for calculating covariance matrix . In this paper, the residual signals from the ELM are used to form an observation vector X_i. Y_i is regarded as the reference vector with a reasonable deviation value. In ideal conditions, the values in the reference vector can be considered to be zero.

MD values can be accumulated over a period of time t, indicating the deviation of the calculated MD value from the expected value for different components in the gearbox. However, it is necessary for a confidence band for the accumulated MD values to be defined. In this paper, the value of the confidence band is set to unity. If the accumulated MD values are below this level, the deviations are attributed to signal interference, which are therefore ignored in the accumulation of MD values. Otherwise, the values are added to the accumulated MD value. Three relationships are considered in this study, including gearbox pump oil pressure with gearbox oil temperature, gearbox pump oil pressure with gearbox bearing 1 (main speed shaft bearing connected to the rotor) temperature, and gearbox pump oil pressure with gearbox bearing 2 (high speed shaft bearing connected to the electric generator) temperature, assessing the condition of each component in the gearbox. The definition of these signals will be described in the subsequent section. The MD values described in this section can therefore be extended to multiple processes.

However, the durability and failure modes of each component in a gearbox can be different; thus weights are allocated to represent the health impact of each component on the performance of a gearbox. Here, a multiple MD model is defined as sum of all MD values above the confidence band observed during a defined period of time. This multiple MD model can be used as the basis of an early warning system, with an alarm raised if the threshold is exceeded.

The accumulated MD model with multiple components is described as follows:

$$RIV = \int_0^t (\alpha MD_1 + \beta MD_2 + \gamma MD_3)dt; \quad \alpha + \beta + \gamma = 1 \tag{8}$$

where RIV is the risk indicator value of the gearbox as a whole; MD_1 is the MD value of the gearbox pump oil pressure to the gearbox oil temperature; MD_2 is the MD value of the gearbox pump oil pressure to the gearbox bearing 1 temperature; MD_3 is the MD value of the gearbox pump oil pressure to the gearbox bearing 2 temperature; α, β *and* γ are the weights of these MD values, respectively. The RIV takes the variability of each MD value into account when determining its distance from the multivariate center of the distribution, thus providing a more sensitive indicator. As can be seen from eq. (8), the *RIV* and its derivative change over time and a higher value of the derivative represents an indication of higher risk, indicating worsening health of the gearbox.

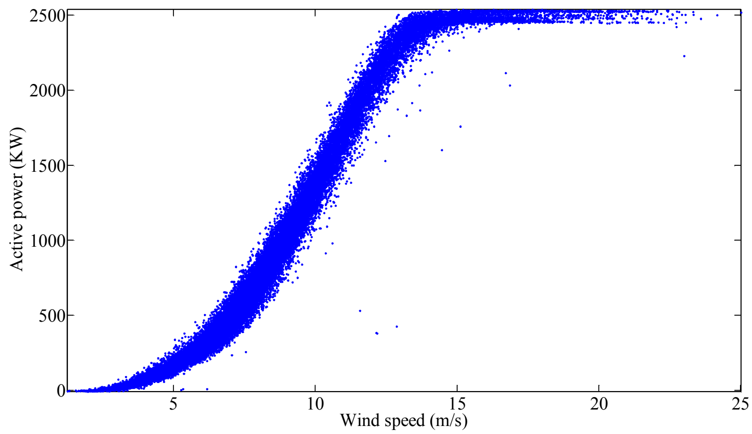
5. Case Studies

5.1 SCADA data

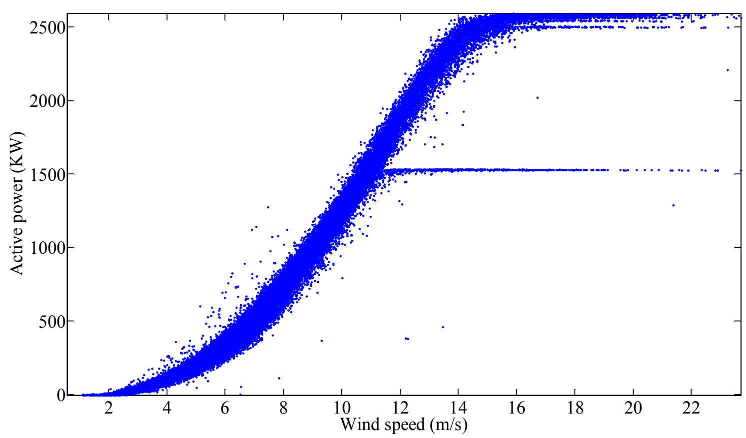
Supervisory control and data acquisition (SCADA) systems utilize hardware and software elements and IT technologies to monitor, gather, and process data. In power systems, SCADA systems are used for a range of functions, including data acquisition, control, adjustment of parameters, and generating warning signals. The SCADA data used here have been obtained from a working wind farm. The use of operational SCADA data is an effective way to demonstrate the algorithms described in this paper. These data represent 12 months' operation and consist of 128 variables, comprising temperatures, pressures, vibrations, power outputs, wind speed, and digital control signals. Note that SCADA signals are usually processed and stored at 10 minute intervals, although sampled in the order of 2 s.

Power curves of two wind turbines, obtained from the SCADA data, are shown in Fig. 2. Fig. 2 (a) illustrates a power curve of a healthy turbine. It can be seen that power varies with the cube of wind speed below the rated speed of 15 m/s. When the wind speed is below the cut-in speed of 4 m/s, the rotor torque is not sufficient for the turbine to produce any power. When the speed of the wind is greater than the cut-out speed of 25 m/s, the turbine is shut down and does not generate any power. At wind speeds above the rated speed but below the cut-out speed, power output is restricted to the rated power of the turbine. This turbine has been chosen as the 'reference turbine', and forms the basis of the ELM model.

In contrast, Fig. 2 (b) shows the power curve of a faulty wind turbine. It can be seen that this turbine has, at some point, operated with reduced power output. After studying the fault log of the turbine, it has been concluded that this power reduction followed a fault with the gearbox.



a



b

Fig. 2 *Examples of two power curves of wind turbines*
a Power curve of a fault-free turbine
b Power curve of a faulty turbine

5.2 Gearbox

This paper is focused on gearbox faults and the health condition of the wind turbine gearbox. The gearbox is one of the key components in indirect-drive wind turbines because the turbine rotor cannot match the synchronous speed of the generator. The gearbox is used to transmit kinetic energy from the turbine rotor to the electric generator, adjusting rotational speed and torque accordingly. However, the gearbox can be a major contributor to a turbine's downtime, with common failure modes being bearing faults and gear teeth faults. Surveys have shown that the root cause of gearbox failure is due to rapid changes of torque from stochastic wind profiles, which create an uneven load for the bearing and misalignment of gear teeth. Other causes of bearing and gear teeth failure are elevated operating temperature and excessive contamination of the cooling lubricant due to failure of the gearbox cooling system.

Any fault from the gearbox can result in an abnormal input to the generator, reducing efficiency or, in extreme cases, damaging the generator [32] [33].

Fig. 3 shows a schematic diagram of the structure of a three-stage gearbox. The gearbox consists of three types of components, specifically, gears, bearings and the cooling system (usually oil cooling). In this paper, gearbox temperature and oil pressure measurements at different locations of the gearbox obtained from the SCADA data [34] [35] are selected to monitor the condition of gearbox, which contain specifically temperature readings for gearbox bearing 1 (main speed shaft bearing connected to the rotor), gearbox bearing 2 (high speed shaft bearing connected to the electric generator) and the gearbox oil (the temperature of gearbox oil is close to actual gear temperature) and the pressure in the oil pump. The oil pressure shows the operating condition of the gearbox cooling system.

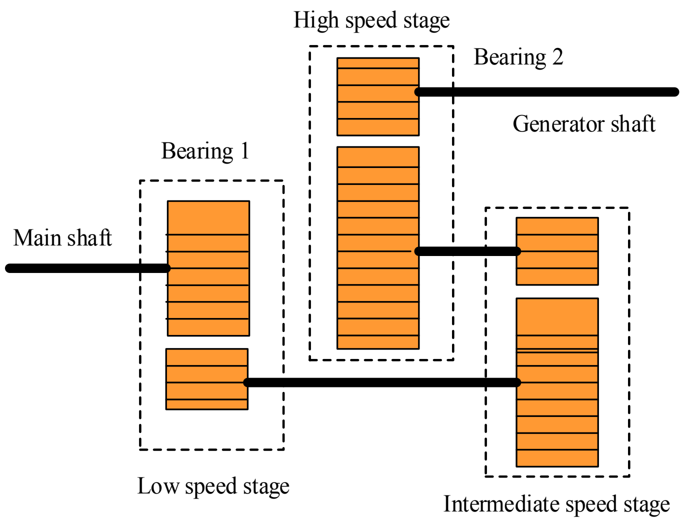
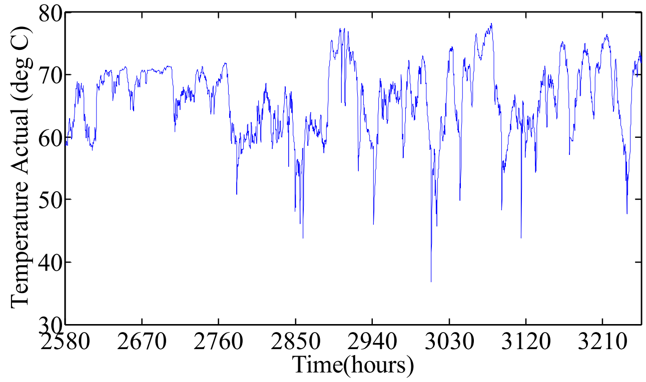


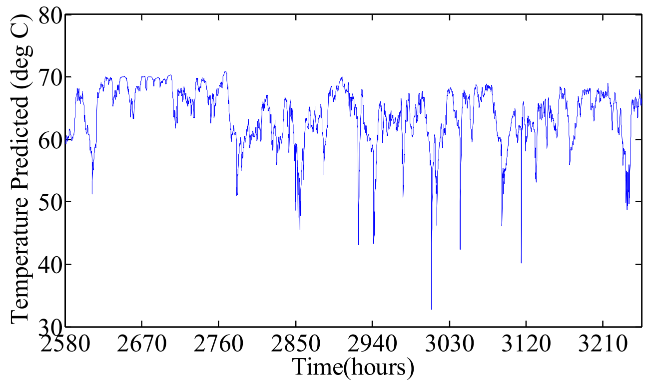
Fig. 3 *Schematic diagram of gearbox structure*

5.3 Model predictions

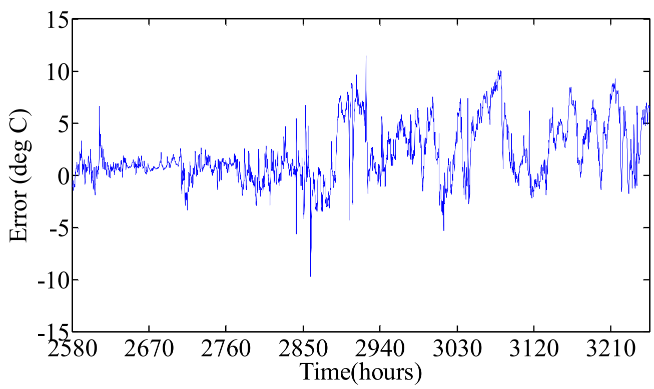
The model predictions for the gearbox oil temperature, gearbox bearing 1 temperature and bearing 2 temperature using the optimized ELM model are illustrated in figures 4 to 6. Fig. 4 (a) shows the gearbox oil temperature obtained from the SCADA data for the faulty turbine. Fig. 4 (b) illustrates the predicted gearbox oil temperature obtained from the ELM model. Fig. 4 (c) illustrates the residual signal between the actual temperature and predicted temperature of the gearbox oil. It can be seen that the actual temperature deviates from the prediction at hour 2850 indicating the onset of the fault. Fig. 5 and Fig 6 show actual SCADA data, the signals predicted by the model, and the residual signals of the temperatures of gearbox bearing 1 and gearbox bearing 2, respectively. The temperatures of gearbox bearing 1 and bearing 2 deviate from the model predictions at hour 2850. At the same time, the actual gearbox oil temperature deviates from the predicted temperature. Clearly, it can be concluded that the models provide a reliable and effective indication of the onset of the gearbox fault.



a

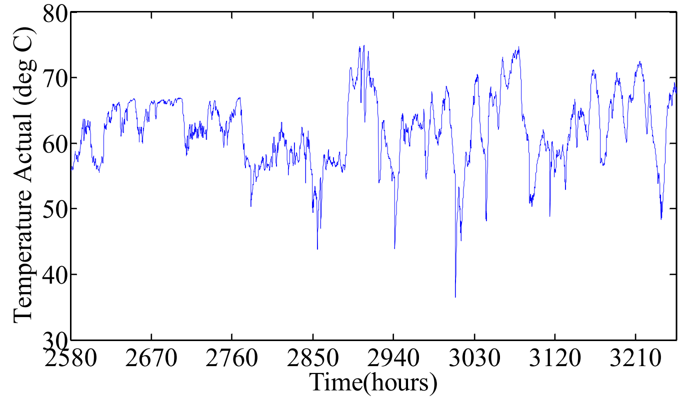


b

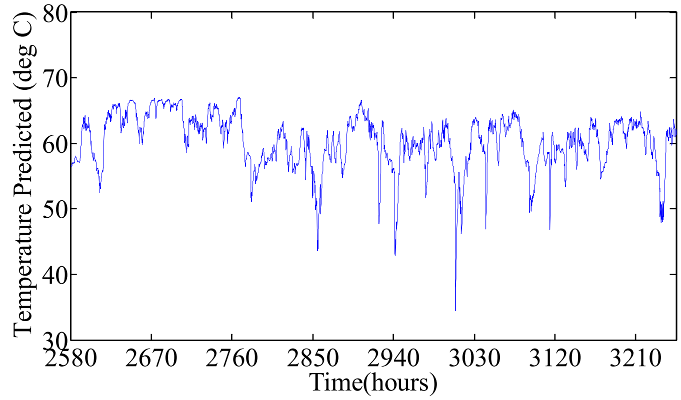


c

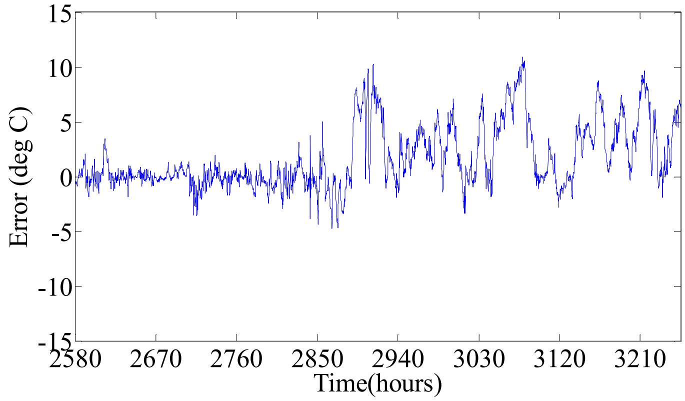
Fig. 4 *ELM model prediction compared to SCADA data for gearbox oil temperature*
a SCADA output
b Model output
c Residual signal



a

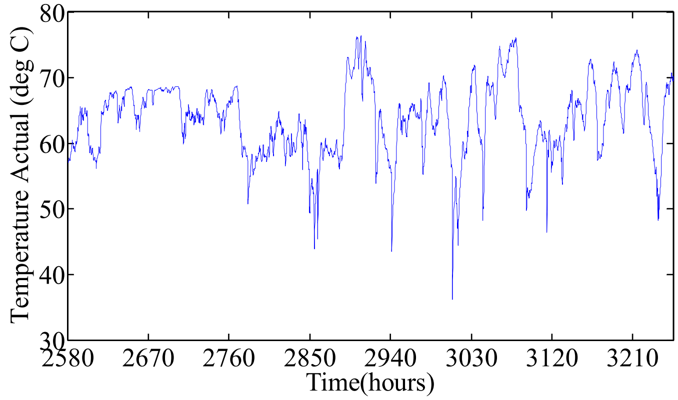


b

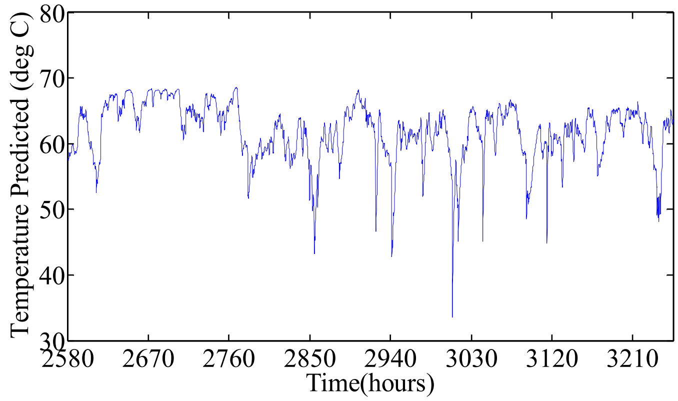


c

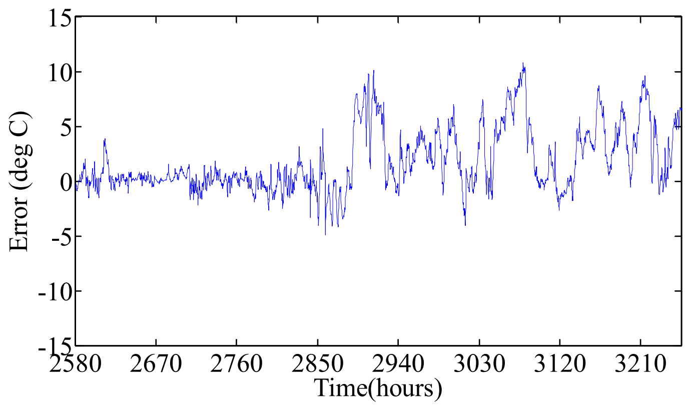
Fig. 5 *ELM model prediction compared to SCADA data for gearbox bearing 1 temperature*
a SCADA output
b Model output
c Residual signal



a



b

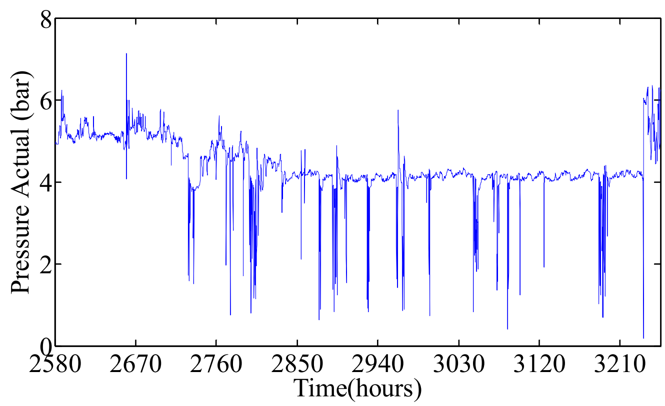


c

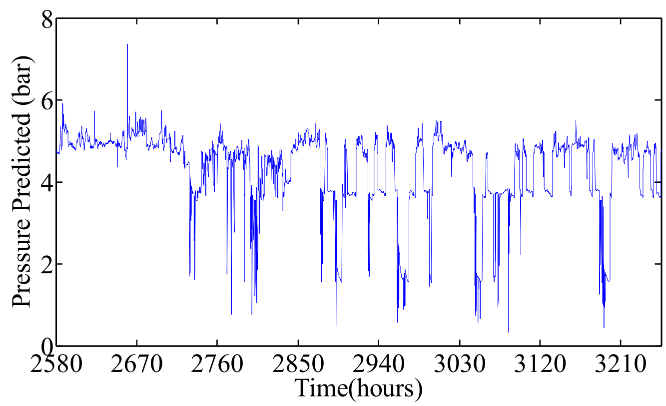
Fig. 6 *ELM model prediction compared to SCADA data for gearbox bearing 2 temperature*
a SCADA output
b Model output
c Residual signal

In addition to the temperature of the gearbox, the pressure of oil in the gearbox pump is another important signal that can be used to detect the faults of the gearbox in a wind turbine.

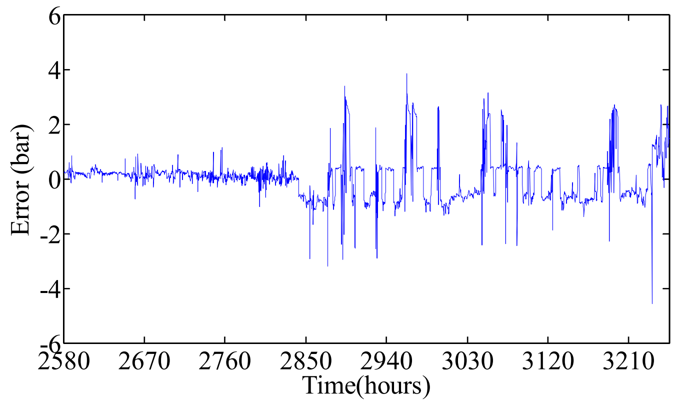
Abnormal levels of oil pressure in the gearbox pump will affect heat dissipation from the gearbox, which is usually caused by faults in the gearbox oil pump, filter blocking of oil-conveying pipes or deterioration of the condition of the cooling oil. Thus, the modelled predictions for the oil pressure in the oil pump are also considered here. Note that the gearbox pump oil pressure changes with the power output of the turbine. Fig. 7 (a) shows the actual oil pressure in the oil pump, while Fig. 7 (b) illustrates the pressure of the oil as predicted by the ELM model. At 2850 hours, the residual signal in Fig. 7 (c) shows that the oil pressure begins to deviate from the model prediction. In general, the cooling system is able to keep the gearbox at the normal operating temperature to ensure that no damage is caused, but when the temperature of the gearbox becomes abnormal, the residual signal of the oil pressure in Fig. 7 (c) fluctuates between positive and negative values. This indicates that the cooling system is attempting to restore the normal working conditions of the gearbox, but it is unable to do so effectively.



a



b



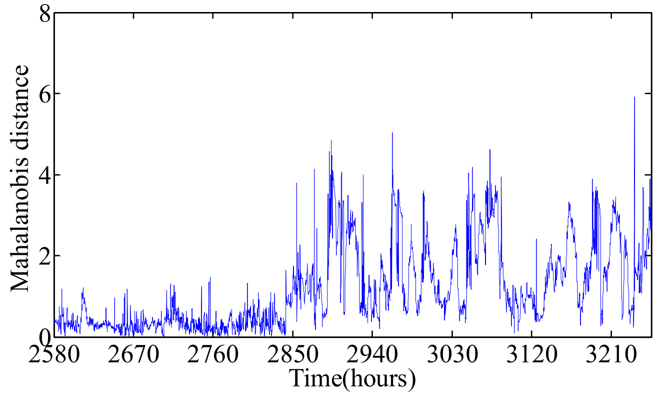
c

Fig. 7 *ELM model prediction compared to SCADA data for gearbox pump oil pressure*
a SCADA output
b Model output
c Residual signal

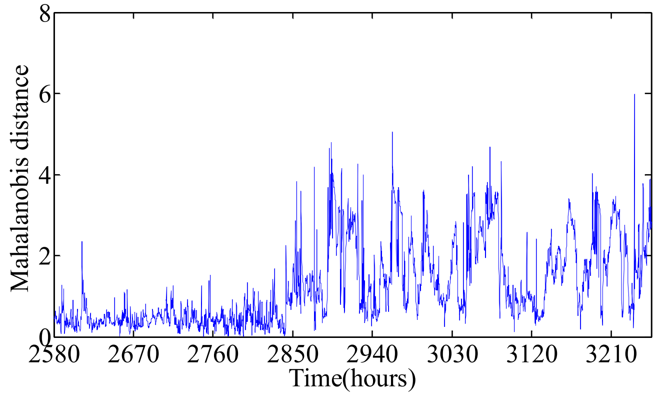
A desktop PC with a Xeon E3-1271 v3 3.6GHz CPU and 16GB RAM was used to implement the ELM. The time taken to train the ELM was compared with that taken to train a traditional BP back propagation neutral network using a threshold training algorithm, an algorithm commonly used to train ANNs. The ELM algorithm learns on an average of 0.16s compared to 22s using the BP method for the same training sets. Consequently, the ELM learning algorithm run around 138 times faster than the BP method. The root mean square error (RMSE) is also employed here as a measure of how well the models explain the actual output data. The RMSE values for the models with ELM and BP are 0.0915 and 0.0862 respectively. This indicates that the ELM model also provides a good fit with considerably reduced learning time.

5.4 Fault identification

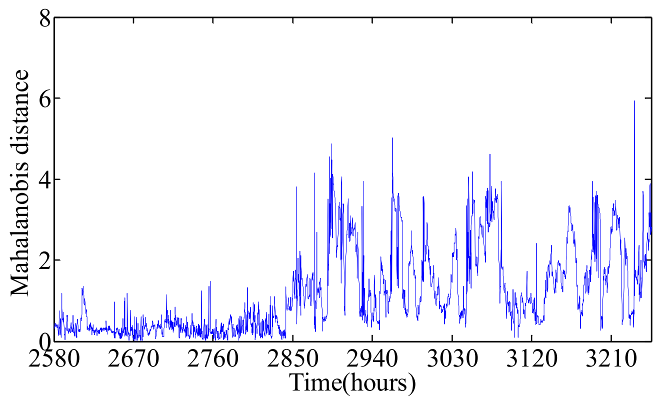
In order to assess further the condition of gearbox components, a MD measure of residual signals is used in this section to establish a relationship between the temperature change of gearbox components and oil pressure in the gearbox oil pump. The residual signal of the oil pressure is shown in Fig. 7(c). The gearbox component residual temperatures, shown in Fig. 4(c), 5(c) and 6(c), have been selected as the observation vectors. Hence, MD values of temperatures for the gearbox oil, gearbox bearing 1, and gearbox bearing 2 in relation to the working condition of the cooling system are obtained. Figure 8 shows the MD values calculated using equation (7) for these gearbox components. It can be seen that the MD values increase significantly at hour 2850, indicating the onset of the fault. Compared to individual residual signals from the predicted models shown in the figures in Section 5.3, these MD values can identify the fault more clearly by taking into account different monitoring signals from the system.



a



b



c

Fig. 8 *MD calculated for gearbox components*
a Gearbox bearing 1
b Gearbox bearing 2
c Gearbox oil

This article has been accepted for publication in a future issue of this journal, but has not been fully edited.
Content may change prior to final publication in an issue of the journal. To cite the paper please use the doi provided on the Digital Library page.

The accumulated MD values, referred to here as the 'risk indicator', describing relationships between the oil pressure and the bearing temperature changes are shown in Fig. 9. As can be seen from the figure, the risk indicators of pump oil pressure to gearbox oil temperature, pump oil pressure to gearbox bearing 1 temperature and pump oil pressure to gearbox bearing 2 temperature demonstrate an almost same change in the derivative over time, representing an approximately equal share of risk of failure of each component. Therefore, for this case, the weightings of gearbox pump oil pressure to gearbox oil temperature, α, gearbox pump oil pressure to gearbox bearing 1 temperature, β, and gearbox pump oil pressure to gearbox bearing 2 temperature, γ, are each set to 1/3. The accumulated MD values from these components are then calculated using eq. (8) to indicate the health condition of the gearbox as a whole. Fig 10 shows the observed risk indicator values of oil pressure to bearing 1 temperature for the gearboxes of one faulty and two fault-free wind turbines over a period of 1 month; the gearbox failure in the faulty wind turbine occurs at the middle of the month. When the fault begins to occur, the risk indicator value increases dramatically to 3500, after 16 days of the fault occurring. Conversely, the observed risk indicator values for the two fault-free wind turbines over the same month increases slowly, simply because of component aging.

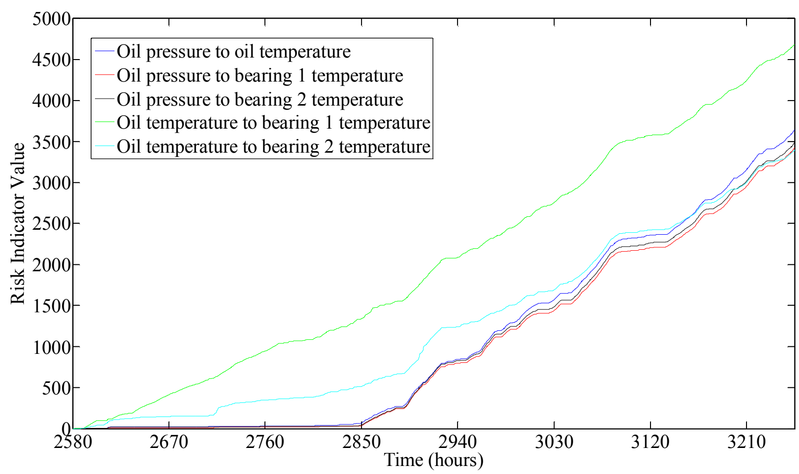


Fig. 9 *Observed risk indicators for the gearbox of a faulty turbine in relation to oil pressure and oil temperature, respectively*

Fig. 9 also shows the observed risk indicators describing the relationship between the bearing temperature changes and the oil temperature. Even though these risk indicators have demonstrated a similar change over time, the MD values associated with the oil temperature increase monotonically with the time, and hence do not show the onset of the fault at hour 2850. It can therefore be concluded that the fault occurs in the cooling system, and the oil pressure should be selected to diagnose the operating condition of the cooling system in the gearbox. As is well known, active cooling systems are the main means for dissipating heat, which, for a wind turbine, include the oil lubrication system of the gearbox and the ventilation system of the generator. A typical gearbox lubrication system in a wind turbine consists of an oil pump unit, a heat exchanger, and an oil filter. Oil filters are used to remove impurities or metal particles within the lubrication oil in order to maintain oil quality and to prevent further wear of gearbox components. Pressure sensors are installed at both ends of the filters to monitor their status, while a temperature sensor is installed in the oil sump to measure lubrication oil temperature. The oil cooling system is started if the oil temperature is over a certain threshold, usually 60°C [34]. In this paper, the increase in gearbox temperature is due to an oil filter becoming blocked, as indicated in the alarm log and from an investigation of the SCADA data. The heat emission efficiency is reduced due to the oil filter blockage, leading to a rise in gearbox temperature.

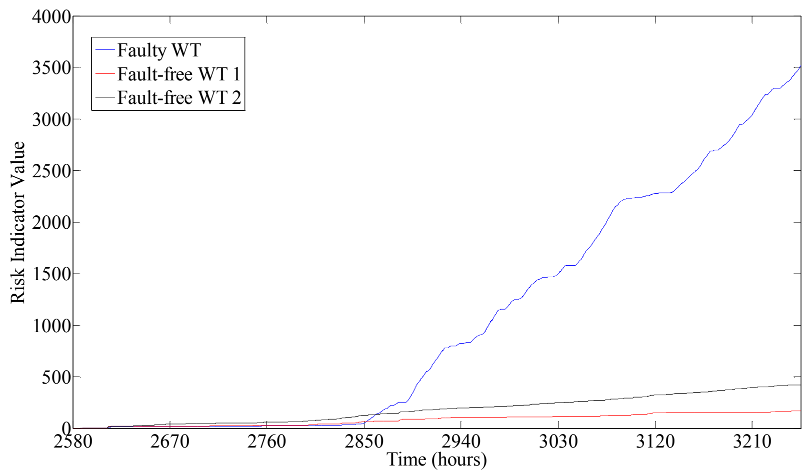


Fig. 10 *Observed risk indicator value of oil pressure to bearing 1 temperature for the gearbox of a faulty and two fault-free wind turbines over a period of 1 month*

6. Conclusions

In this paper, a data-based approach using an extreme learning machine (ELM) algorithm optimized with a genetic algorithm has been proposed for condition monitoring of the gearbox in wind turbines. SCADA data, acquired from a working wind farm, have been used to demonstrate the effectiveness of the ELM method. These data include the temperature of the oil in the gearbox, the temperature of the gearbox bearings, and the pressure in the gearbox oil pump. Models derived from these data have been used to identify faults. It has been shown that the residual signals between the actual output and the predicted output are caused by a gearbox fault, providing an early warning of impending failure. The results also demonstrate that the ELM learning algorithm can provide a good fit with a considerably reduced learning time compared to a BP algorithm.

Moreover, Mahalanobis distance (MD) values and accumulated MD values, obtained from multiple components, are employed to identify the health condition of the gearbox. These MD values can detect the fault more effectively by taking into account a range of different monitoring signals from the system. Observed risk indicator values, describing relationships between different components in the gearbox, have shown that the cooling system has a significant effect on the performance of the gearbox system.

Note that the data used in this paper are mostly representative of the normal operation of wind turbines and do not contain a great deal of information regarding the occurrence of faults; consequently, this paper employs static ELM models only. Future work will therefore consider dynamic models by taking into account the effect of more past inputs on the model output, and the different effect each component has on the health condition of the gearbox. In this paper, the same value is used as the risk indicator for several different gearbox components. It is clearly worth evaluating different risk indicator values, taking into account the residual signal produced from the ELM model and the contributions to the downtime caused by failure of each component. A real-time early warning system, employing an online sequential ELM, will also be developed in order to predict faults in the operational wind turbines.

Acknowledgement

The SCADA data were obtained from Wind Prospect Ltd. and permission to use these data is gratefully acknowledged.

Reference

[1]. Walford, C.A.: 'Wind Turbine Reliability: understanding and minimizing wind turbine operation and maintenance costs', technical report SAND 2006-1100, Sandia National Laboratories, USA, 2006.

[2]. Yang, W., Tavner, P.J., Tian, W.: 'Wind Turbine Condition Monitoring Based on an Improved Spline-Kernelled Chirplet Transform', IEEE Transactions on Industrial Electronics, 2015, 62(10), pp.6565-6574.

[3]. Yang, W., Tavner, P.J., Crabtree, C.J., Wilkinson, M.: 'Cost-Effective Condition Monitoring for Wind Turbines', IEEE Transactions on Industrial Electronics, 2010, 57(1), pp.263-271.

[4]. Bhaskar, K., Singh, S.N.: ‘AWNN-assisted wind power forecasting using feed-forward neural network’, IEEE Transactions on Sustainable Energy, 2012, 3(2), pp. 306-315

[5]. Nizar, A.H., Dong, Z.Y., Wang, Y.: ‘Power utility nontechnical loss analysis with extreme learning machine method’, IEEE Transactions on Power Systems, 2008, 23(3), pp. 946-955

[6]. Cross, P., Ma, X.: ‘Nonlinear system identification for model-based condition monitoring of wind turbines’, Renewable Energy, 2014, 71, pp.166-175.

[7]. Ribrant, J., Bertling, L.M.: ‘Survey of failures in wind power systems with focus on swedish wind power plants during 1997-2005’, IEEE Transactions on Energy Conversion., 2007, 22(1), pp. 167-173.

[8]. Qiao, W., Lu, D.: ‘A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis—Part I: Components and Subsystems’, IEEE Transactions on Industrial Electronics, 2015, 62(10), pp.6536-6545

[9]. Qiao, W., Lu, D.: ‘A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis—Part II: Signals and Signal Processing Methods’, IEEE Transactions on Industrial Electronics, 2015, 62(10), pp.6546-6557

[10]. Yang, S., Li, W., Wang C.: ‘The intelligent fault diagnosis of wind turbine gearbox based on artificial neural network’, Proc. Int. Conf. Condition Monitoring and Diagnosis, Beijing，China, April. 2008, pp. 1327-1330.

[11]. Wang, Z., Guo, Q.: ‘The diagnosis method for converter fault of the variable speed wind turbine based on the neural networks’, Proc. Int. Conf. Innovative Computing, Information and Control, Kumamoto, Japan, Sept. 2007, pp. 615-615.

[12]. Hou, G., Jiang, P., Wang, Z., Zhang, J.: ‘Research on fault diagnosis of wind turbine control system based on Artificial Neural Network’, Proc. Int. Conf. Intelligent Control and Automation, Jinan , China, Jul. 2010, pp. 4875-4879.

[13]. Tang, B., Song, T., Li, F., Deng, L.: ‘Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine’, Renewable Energy, 2014, 62, pp. 1-9.

[14]. Zeng, J., Lu, L., Zhao, Y., Zhang, Z., Qiao, W., Gong, X,: ‘Wind turbine fault detection and isolation using support vector machines and residual-based method’, Proc. Int. Conf. American Control Conference (ACC), Washington, DC, USA, Jun. 2013, pp. 3661-3666.

[15]. Li, H., Hu, Y., Yang, C., Chen, Z., Ji, H., Zhao, B.: ‘An improved fuzzy synthetic condition assessment of a wind turbine generator system’, International Journal of Electrical Power & Energy Systems, 2013, 45, pp. 468-476.

[16]. Chen, B., Matthews, P.C., Tavner, P.J.: ‘Wind turbine pitch faults prognosis using a-priori knowledge-based ANFIS’, Expert Systems with Applications, 2013, 40(17), pp. 6863-6876.

[17]. Liu, H., Dong, X., Yang, Z., Zheng, K.: ‘The application of intelligent fuzzy inference to the fault diagnosis in pitch-controlled system’, Energy Procedia, 2012, 16, Part C, pp. 1839-1844.

[18]. Peng, Q., Ma, X., Wang, Y.: ‘Condition monitoring of wind turbines based on extreme learning machine”, Proc. 21st Int. Conf. on Automation and Computing (ICAC), Glasgow, 11-12 September 2015.

[19]. Elobaid, L.M., Abdelsalam, A.K., Zakzouk, E.E.: ‘Artificial neural network-based photovoltaic maximum power point tracking techniques: a survey’, Renewable Power Generation, IET, 2015, 8, pp.916-924.

[20]. Zaher, A., McArthur, S.D.J., Infield, D.G., Patel, Y.: ‘Online wind turbine fault detection through automated SCADA data analysis’, Wind Energy, 2009, 12(6), pp. 574-593.

[21]. Bhaskar K., Singh, S.N.: 'AWNN-assisted wind power forecasting using feed-forward neural network', IEEE Transactions on Sustainable Energy, 2012, 3(2), pp. 306-315.
[22]. Huang, G.B., Zhu, Q.Y., Mao, K.Z.: 'Can threshold networks be trained directly?', IEEE Transactions on Circuits and Systems II: Express Briefs, 2006, 53(3), pp.187-191.
[23]. Nizar, A.H., Dong, Z.Y., Wang, Y.: 'Power utility nontechnical loss analysis with extreme learning machine method', IEEE Transactions on Power System, 2008, 23(3), pp. 946-955.
[24]. Huang, G.B., Ding, X.J., Zhou, H.M.: 'Optimization method based extreme learning machine for classification', Neurocomputing, 2010, 74, pp. 155-163.
[25]. Alexander, G., Yoan, M., Maarit, K., Amaury, L.: 'Singular Value Decomposition update and its application to (Inc)-OP-ELM', Neurocomputing, 2016, 174, pp. 99-108.
[26]. Holland, J.H.: 'Genetic algorithms', Sci. American., 1992, 267, pp. 66-72.
[27]. McCabe, A.P., Aggidis, G.A., Widden, M.B.: 'Optimizing the shape of a surge-and-pitch wave energy collector using a genetic algorithm' renewable energy, 2010, 35, pp.2767-2775.
[28]. Sun, X., Yang, Z., Wang, Z.: 'The application of BP neutral network optimized by genetic algorithm in transportation data fusion', Proc. Int. Conf. Advanced Computer Control, Shenyang, China, March 2010, pp. 560 -563.
[29]. Zeng, J., Zhang, C., Xie, N., Yang, P., Xu, C., Zhang Z.: 'A Hybrid Model for Short-Term Wind Power Forecasting Based on MIV, Tversky Model and GA-BP Neural Network', Proc. Int. Conf. Manufacturing and Industrial Technologies, Istanbul, Turkey, MAY, 2016.
[30]. Wang, Y., Miao, Q., Ma, E.W.M., Tsui, K.L., Pecht, M.G.: 'Online anomaly detection for hard disk drives based on Mahalanobis distance', IEEE Transactions on Reliability, 2013, 62, pp. 136-145.
[31]. Zhang, Y., Du, B., Zhang, L., Wang, S.: 'A Low-Rank and Sparse Matrix Decomposition-Based Mahalanobis Distance Method for Hyperspectral Anomaly Detection', IEEE Transactions on Geoscience and Remote Sensing, 2015, 99, pp.1-14.
[32]. Qiu, Y., Zhang, W., Cao, M., Feng, Y. and Infield, D.: 'An Electro-Thermal Analysis of a Variable-Speed Doubly-Fed Induction Generator in a Wind Turbine,' Energies, 2015, 8, pp. 3386-3402.
[33]. Feng, Y., Qiu, Y., Crabtree, C. J., Long, H. and Tavner, P. J. : 'Monitoring wind turbine gearboxes,' Wind Energy, 2013, 16, pp. 728-740.
[34]. Qiu, Y., Zhang, W., Infield, D., Feng, Y and Sun, J.: 'Applying thermophysics for wind turbine drivetrain fault diagnosis using SCADA data,' IET Renewable Power Generation, 2016, 10, pp. 661-668.
[35]. Kaidis, C., Uzunoglu, B., Amoiralis, F.: 'Wind turbine reliability estimation for different assemblies and failure severity categories', Renewable Power Generation, IET, 2015, 8, pp.892-899.