CONDITION MONITORING OF WIND TURBINE GEAR-BOXES USING VIBRATION BASED METHODS

Jiayao Lin

The department of Mechanical Engineering, The University of Sheffield, United Kingdom
email: ljy3326126@163.com

During 2003 to 2012, the aggregated downtime of gearboxes is the longest for wind turbines. The maintenance costs of gearboxes are the highest in the wind turbines. This thesis is motivated by the increasing demand for new repair methodologies in wind turbine gearboxes, and the need for the monitoring of these expensive equipment. Its aim is to distinguish the damaged state data from the undamaged state data. There are several steps to monitor the condition of the gearboxes, including operational evaluation, data acquisition, feature extraction and feature discrimination. The operational evaluation is to justify the performance of condition monitoring and define the damage. The data acquisition is to select the excitation methods, the sensor types, number and locations, and the hardware of acquisition. In this paper, the Welch’s method and the principal component analysis (PCA) method are adopted to extract the features. The outlier analysis, multi-layer perceptron (MLP) and radial basis function (RBF) are applied to discriminate the data respectively to compare with their advantages and disadvantages to others. The results show that the data are successfully processed, extracted the features, and classified whether the data are damaged or not by these methods. However, each method has its own drawbacks.

Keywords: Condition monitoring, Wind turbine gearboxes.

1. A Brief Introduction

1.1 Background and motivation

Wind turbines are pieces of equipment that transform kinetic energy (from wind) into electrical power. According to the Global Wind Energy Council [1], it is projected that wind energy will supply 20% of global electricity by 2030. In 2016, global wind energy increased by more than 54Gw to reach a total of 486.8GW (cumulative capacity rose by 12.6%). Europe will continue to lead in low-cost offshore wind energy and it is strongly supported by the public authorities [2]. The European Wind Energy Association predicts that 31% of EU electricity will be produced by wind energy by 2030 (19.7% from onshore wind and 11.3% from offshore wind) [3]. In addition, the costs of offshore production are reducing dramatically because of new developments [1]. The average cost of wind power energy will decrease by 20% from 2012 to 2030 in the US, according to the National Renewable Energy Laboratory [4].

There are three main reasons that wind turbines require condition monitoring (CM). Firstly, the harsh working environment (such as humidity, salinity, fluctuating temperature, ice, etc.) and the fluctuating loads have a massive impact on the performance of wind turbines, even causing the loss of basic function. Moreover, wind turbines are structures that are difficult to access and are built in remote sites. Additionally, the economic losses due to wind turbine downtime increase with the increasing size of the turbines [5]. The total cost of operation and maintenance accounts for about a quarter to one third of the total cost of offshore wind energy [6, 7]. The maintenance costs for offshore wind turbines are 20-25% higher than onshore wind generation [6]. Therefore, it is crucial to use CM for wind turbines, because of the benefits [7-9]. There is a significant amount of literature about CM for wind turbines [8, 10-13].
Pattern recognition is a useful discipline to identify damage by using machine learning algorithms, which assigns class labels to samples of measured data. The main approaches are statistical, neural and syntactic in mathematical terms of pattern recognition. It is a natural approach for SHM using a statistical approach. The neural network approaches are similarly interpreted in statistical approaches. As SHM is referred to as CM, when applied to rotating machinery [14], it uses same procedures as CM. There are three main procedures in this case, namely: Data acquisition, Feature extraction and Feature discrimination.

The vibration analysis is that a certain vibration signal in standard condition changes in a way due to the fault development [15]. The machines generate vibrations which are directly related to periodic or non-periodic events such as rotating shafts, meshing gear teeth, rotating electric fields etc. during the machine’s operation. There are various signals to analyse for condition monitoring including vibration, acoustics, oil, strain measurement and thermography. However, the vibration-based condition monitoring method is one of the most popular methods for wind turbine gearboxes [13, 16-17]. One advantage of vibration analysis is that this analysis suffers low signal-to-noise ratio, because the accelerometers are generally located in the case [16]. However, this technique requires the installation of multiple vibration sensors on the wind drivetrain and high-end data acquisition equipment, which is expensive [18]. Moreover, vibration analysis is able to detect issues at the high speed (>600 rpm) stage of the drivetrain. Nevertheless, it is a challenge to detect faults at low speed [16-18].

1.2 Aim and outline of thesis

The aim of this thesis is to describe the detection of damage to wind turbine gearboxes using vibration-based methods. Vibration measurement and spectrum analysis are traditional techniques for condition monitoring of gearboxes [18]. As noted above, wind turbines suffer from varying loads and speeds due to the highly transient nature of wind. This means that the vibration signals are typically non-stationary. Furthermore, the non-linear signals cause by the dynamics. These challenges faced by wind turbine gearboxes require more complicated tools than just a spectrum analysis to process signals. In this paper, the Welch’s method and PCA are used to process signals as feature extraction methods. The Welch’s method is to produce the power spectrum of the vibration data of gearboxes. The PCA is a tool to reduce the dimension of data.

The outline of the thesis will be as follows: The gearbox datasets will be described in Section 2. The results and discussion of feature extraction and feature discrimination will be described in Section 3. Finally, the thesis conclusions and suggestions for future work are included in Section 4.

2. Gearboxes Datasets

There are two different sets of experimental gearbox data from an NEG Micon NM 1000/60 wind turbine in Germany. One is damaged state data and the other is undamaged state data (simply write as damaged data and undamaged data in following thesis), taken by members of the company EC Grupa. This is a Polish company which maintains the wind turbine. Thanks to Doctor Antoniadou for sharing these data.

One of the most common standards for design and specification of gearboxes for wind turbines is ANSI/AGMA/AWEA 6006-A03 [19]. This standard for gearboxes is applied for power from 40kW to 2MW wind turbines. This gearbox is used in all parallel axes with one stage epicyclic and combined one stage epicyclic and parallel shaft. It is important to design gearboxes to standard criteria, because the standard shows the differences and provides information which is based on field experience, theory and laboratory data.

The measured data are from a single accelerometer which carried more information for the damaged gearbox case. The sample frequency of the measurements was 25 kHz with 10×10^6 number of points. The measured gearbox has a 28-tooth gear (smaller wheel) which meshes with an 86-tooth gear (bigger wheel) at ISS, where one tooth was observed to be damaged.
It is important to normalise the data because the different input produces a different system. In this paper, the acceleration data measured by the accelerator sensor are used for normalisation. The normalisation equation is as follows:

\[ \text{normalization} = \frac{x - \mu}{\sigma} \]  

(1)

Where, \( x \) is matrix of acceleration data; \( \mu \) is matrix of mean of acceleration data; \( \sigma \) is matrix of standard deviation of acceleration data.

**Figure 1:** The normalisation of damaged data.  
**Figure 2:** The normalisation of undamaged data.

The results of damaged data and undamaged data, which are calculated and plotted by MATLAB, are showed in Figure 1 and Figure 2, respectively. From these figures, it is observed that the total test time is 40s and it is hard to discern whether there is damaged or not. Therefore, the next step is to extract features by using signal processing methods and discriminates between features. This is done in the next section. As noted above there are two sorts of data (damaged and undamaged data) provided in this paper. The \( 10 \times 10^6 \) number of acceleration points for each kind of data are divided into 100 observations for each.

3. Results And Discussion

3.1 Feature extraction results

3.1.1 Results of Welch’s methods.

The window used in this case is a Hamming window, which was briefly introduced in the previous section. Non-smooth means the noise in the PSD results, therefore, the length of the Hamming window and the amount of overlap are modified to find the better solutions. The amount of overlap must be a positive integer 50% less than the length of window. The parameters for smooth case used in the Welch’s method are presented in the Table 1.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of Hamming window</td>
<td>400</td>
</tr>
<tr>
<td>Amount of overlap</td>
<td>200</td>
</tr>
<tr>
<td>Number of FFT points</td>
<td>10000</td>
</tr>
<tr>
<td>Sample frequency</td>
<td>25000 Hz</td>
</tr>
</tbody>
</table>

**Table 1:** The parameters for smooth case used in the Welch’s method

**Figure 3** Results of 200 observations by using Welch’s method.

The frequency in this figure is 12.5 kHz, which is half of sample frequency. This is explained by the Nyquist frequency. The energy in the original signal exceeds the Nyquist frequency will also cause the folding, which called aliasing. The red curves are damaged data, and the blue curves are healthy data. There are two three characteristics: The largest power value is in the low frequency
(near 2 kHz) for both sets of data. However, comparing with these sets of data, the biggest difference between damaged data and undamaged data is near 9 kHz. There are mainly four extreme peaks for the healthy data and five extreme peaks for the damaged data.

The PSD results of 100 damaged data observations and 100 undamaged observations by using Welch’s method are plotted in the Figure 3. There are 5001 dimensions for this data, which means it is important to select valuable features. Therefore, the point of highest difference, which has been chosen as the main feature of the data, is selected as well as the five points before and after that point. This means that the matrix of 11 dimensions × 200 observations (100 damaged state observations and 100 undamaged state observations) is extracted as features to be utilised in the feature discrimination methods.

### 3.1.2 Results of the PCA methods.

The PCA method is to reduce to two dimensions of data in this case. For an example of one damaged state observation and one healthy state observation, the Figure 4 presents the results of using the PCA method to decrease the dimension of data from the 11 dimension features extracted by Welch’s method. The features are extracted from 2 dimensions × 200 observations (100 damaged state observations and 100 undamaged state observations) for the feature discrimination. There are two different sorts of points. The red points are damaged data observations, and the blue points are undamaged data observations. The characteristics for PCA results are described as follows: The healthy state observations are normally lower than the damaged state observations for each principal component; The difference between healthy state observations and unhealthy state observations is massive; All the health state observations are clustering and similarly for the damaged state observations.

![PCA results of 100 damaged state observations and 100 healthy state observations](image)

**3.2 Feature discrimination results**

#### 3.2.1 Results of outlier analysis.

There are two feature extraction methods used for outlier analysis to classify the data in this paper. Firstly, the results of the outlier analysis using the Welch’s method are showed in the Figure 5. The red line is the value of threshold calculated by the Monte Carlo method. There are 200 points in this figure, each point being an observation. Specifically, observations 1 to 100 observations are healthy state observations, and observations 101 to 200 observations are the damaged state observations. The data from 40-90 observations of healthy data are selected as training data. The sample interval is 1, and the confidence interval is 99%.

![Outlier analysis for using Welch’s method extracting the 11-dimension features](image)
Figure 6 Outlier analysis using PCA method (2-dimension) features extracting by Welch’s method.

From Figure 5, it is obvious that the outlier analysis can classify the damaged data and undamaged data successfully (training data 40-90 observations). The healthy state observations are all under the threshold line and the damaged state observations are all above the threshold line. Secondly, the Figure 6 illustrates the outlier analysis in terms of PCA method. The 11-dimension data from Welch’s method are reduced to two-dimension principal components. Similarly, the red line means the value of the threshold, and there are 100 sets of undamaged data as well as 100 sets of damaged data. The training data is from 40 to 90. The sample interval and confidence interval are 1 and 99%, respectively. The results in Figure 6 show that the outlier analysis using the PCA method can successfully discriminate the damaged data and the undamaged data without errors.

3.2.2 Results of MLP and RBF.

There are 11 dimensions in the features extracted by Welch’s method. The first and second dimension of these 11 dimensions are randomly selected as input feature data in the MLP method. In this case, the two-layer MLPs are used. The activation function is “logistic function”. The main parameters of MLPs are set out in the Table 2.

![Table 2: Parameters of MLPs and RBF settings](image)

The results of Welch’s method used in the Two-layer MLP are plotted in Figure 7, which is the yellow line. This shows that the damaged data observations and the undamaged data observations are successfully classified by using the MLP method. The horizontal axis is the first dimension of the rate of power over frequency, and the vertical coordinate is the second dimension of the rate of power over frequency. The blue points are healthy data observations, whereas the red points are the damaged data observations. The yellow line, which is seen to be linear, is calculated by the MLP method. The healthy state observations are distributed near a line and the difference between each value is small, as is also the case for the damaged state observations.

![Figure 7 Welch’s method classified by MLP and RBF](image)
In comparison, the output activation functions in RBF are different. The “tps” function is selected as the activation function. The parameters for setting RBF using the 11 dimension features are also showed in the Table 2. The green curve in Figure 7 is the result of RBF method using 11 dimension features which are calculated by Welch’s method. Although these lines are different, they can classify the observations.

The two dimension features extracted by the PCA method are applied in the MLP and RBF. The parameters are the same as Table 2, which are used in the case of PCA case. The results are presented in the Figure 8. The yellow curve shows that MLP using PCA extracted features can discriminate the damaged state observations from the undamaged state observations. The green line shows the results of RBF methods to discriminate the two dimension features from the PCA method. It is similar to the PCA applied in the MLP method. However, the green line is nearly linear in RBF case, while the yellow curve is non-linear in MLP case. The blue points are healthy state observations, while the red points are damaged state observations. The horizontal and vertical coordinates for are the first principal component and second principal component, respectively. The data are divided by the red curve, which is calculated by the MLP.

![Figure 8 PCA method applied in MLP and RBF method](image)

### 3.3 Discussion

The signals can be addressed and extracted the useful features which can be used to discriminate damaged data from undamaged data by Welch’s method. The variance of the periodogram can be reduced by increasing the bias. The type and the length of window have a significant influence on the performance of signals processing.

PCA is also a useful method to reduce the dimension of data; however, there are no physical meanings for the principal components. This will lead to the lack of information of faults location. Furthermore, the high dimension can be decreased to relatively low dimensions which may lead to loss of information. For example, in this case, if the two dimensions are not based on the 11 dimensions, which is the main difference from the Welch’s method, the figure may be more complex.

Outlier analysis method can be applied to detect the damage successfully, but the selected training data have an important impact on the accuracy of the detection. The threshold for the same data has different value in some degree, because this value is based on the Monte Carlo method which is randomly spawned from a zero-mean, unit standard deviation Gaussian distribution. However, if 95% of confidence bounds are on the same settings, there will be an error in the healthy testing data. The 38th observation in the healthy data is over the threshold, which is not correct. This indicates that 95% of confidence limit to detect the damaged data.

Compared with MLP, the RBF method is more complex, because the activation function in RBF is an accumulation of local processing neurons which are sensitive to the changes of training vector. Nevertheless, it may not be necessary for the RBF to require a full non-linear optimisation of all parameters, which means the speed of RBF algorithms is faster than MLP. The basis in the functions is unconditional density of the input data. Hence, choosing the basis function parameters is used in
an unsupervised learning procedure, and the optimising the output layer weights for a supervised method.

It is general practice to use logistic output activation functions for the MLP used to classify, and the corresponding cross-entropy error function ensure the outputs value to one and all in the interval 0 to 1. In addition, it will not take extra time for train an MLP, because of the linear outputs. However, there is no quadratic error surface for the RBF method with logistic outputs in the output layer. This means that without quadratic error surface, it will slow down the algorithms speed, even lower than MLP method. On the other hand, the spaces distant from the training data will be generated by the MLP, which cannot be achieved by the RBF method. This is because the MLP computes a global approximation to a pattern mapping.

With regarding to the Hessian matrix, which is a square matrix of second order partial derivatives of a scalar-valued function, the computation of this matrix for RBF is more complex than for MLP. This is due to different roles in the overall function for each type of weight.

In terms of the best number of hidden layers and hidden units, factors should be considered as follows: The numbers of input and output units; The noise in the training data; The number and distribution of observations; The type of activation functions.

In this case, the linear distribution of data classified by MLP will plot a linear line, while the RBF will use a non-linear curve for classification. If the hidden units are not enough this will produce generalisation errors, because of under-fitting. For example, in the RBF results, the hidden units may not be suitable; therefore, the data may sometimes not be classified. However, a large number of hidden units will be also generated over-fitting, with poor results. For example, there are 20 numbers of hidden units analysing in the 11-dimension case, which is showed in Figure 9. In this figure, the data cannot be discriminated clearly, and it is over-fitting.

![Figure 9: The 20 numbers of hidden units analysing in the 11-dimension case.](image)

4. Conclusion And Further Work

The focus of this thesis is to classify the damaged data and the healthy data from the wind turbine gearboxes by using vibration-based condition monitoring methods. The wind turbine gearboxes are under varying loading and speed, which is difficult to detect the damage. Therefore, the advanced signal processing methods (Welch’s method and PCA method) are used to extract features. For the feature discrimination part, the outlier analysis method, MLP and RBF method are used to analyse.

Each method has its own limitations. Welch’s method can address the signals successfully, but the time-frequency methods may be better to analyse the non-linear and non-stationary signals. The PCA method can reduce the dimension of data. However, this method is lack of meaning for the principal components and the loss information for the relatively to reduce high dimension. In terms of machine learning methods, the accuracy of Outlier analysis method highly depends on the training data and the confidence boundary. The numbers of hidden layers, hidden units, input and output units, observations and the type of activation functions mainly affect the performance of MLP and RBF method.
In the future work, there are more difficult situations in practice. These used machine learning methods are affected by the training data we have. This can be limiting since we might not always have the training data we need to train our algorithms and create our models. Maybe using an adaptive approach can be a solution to this problem.

REFERENCES