Automated Fault Detection of Wind Turbine Gearbox using Data-Driven Approach

Hemanth Mithun Praveen[1], Tejas[2], Sabareesh G R[3]

[1,2,3]*Department of Mechanical Engineering, BITS-Pilani Hyderabad Campus, 500078, India*

hemanth.mithun.praveen@gmail.com
tejas.anilkumar28@gmail.com
sabareesh@hyderabad.bits-pilani.ac.in

ABSTRACT

Wind turbine manufacturers have adopted condition monitoring systems to monitor and report a turbine's health and operating parameters to ensure that the system operates within its design specifications. While the present systems use specialized condition monitoring hardware to detect abnormal acoustic or vibration signals, it is not capable of pinpointing the exact location of the fault apart from isolating the system from which the signal originated. This drawback can be attributed to the requirement of powerful signal processors in order to decode the signal and efforts to train a system to identify the signal emitted by a faulty component. In the light of recent advancement of data-driven approaches and signal processing, these drawbacks can be overcome with increased computation power and sophisticated algorithms that foray into every integrated system. This paper reports such an investigation conducted on a miniature wind turbine planetary gearbox subjected to multi-component failures. The vibration signals were acquired using two accelerometers placed inside the gearbox. The speed of the gearbox was varied according to a simulated wind flow pattern. The primary goal of the study was to investigate the practicality of implementing data-driven approaches to categorise multi-component faults from a composite non-stationary signal. Short time Fourier transforms (STFT) coefficients were used as attributes by a set of data-driven algorithms to build machine learning models. Each model built was tested with a randomised set of instances which was reserved from the main dataset and tested multiple times by means of cross validation. The novelty in the paper entails a methodology which has been devised to classify faults using a randomised vibration dataset with little human intervention by means of machine learning algorithms. The authors propose that this methodology can also be used for real-time fault detection and classification for various machinery and components.

1. INTRODUCTION

Wind energy, one of the popular non-conventional energy harvesting methods has been around for a while and new technological innovations for wind turbines on its components have contributed to steady improvements regarding its reliability. These turbines, during their service life need to be maintained and monitored to ensure longevity and effective utilization. A study by the NREL has suggested that majority of the faults (about 65%) arising in wind turbines occur within the bearings, predominantly the high-speed shaft bearing of the wind turbine gear box (Sheng, 2015).The other major component in which the fault is likely to occur is a gear. These components can be considered as the critical components and any further reference to critical components in the paper will refer to either bearings or gears. A thorough analysis and monitoring of these critical components will ensure its safety and assure longevity of the wind turbine. Vibration signal monitoring and analysis is an effective way to pinpoint the occurrence and location of the fault.

Figure 1 shows the schematic of a wind turbine. A wind turbine comprises of a rotor assembly mounted on a hub which houses the mechanism for controlling the rotor pitch. This allows the turbine to vary the blades angle of attack to maintain the speed of the low speed shaft. The low-speed shaft connects the turbine to the gearbox. Most modern turbines use a planetary gearbox. A planetary gearbox is mostly a multistage gearbox having a gear ratio as high as 1:98 and is normally used for a megawatt category turbine (Sheng, 2015). This gearbox is capable of providing high transmission ratios within a small volumetric space when compared to a fixed axis gearbox. It comprises of a ring gear, planet gear(s), planet carrier and a sun gear. The ring gear is bolted to the transmission casing which provides a reaction force for the gear assembly to work. The planet carrier acts as the input and the sun gear acts as the output. The output shaft of the gearbox is the high-speed shaft. The high-speed shaft is coupled to a generator for power generation.
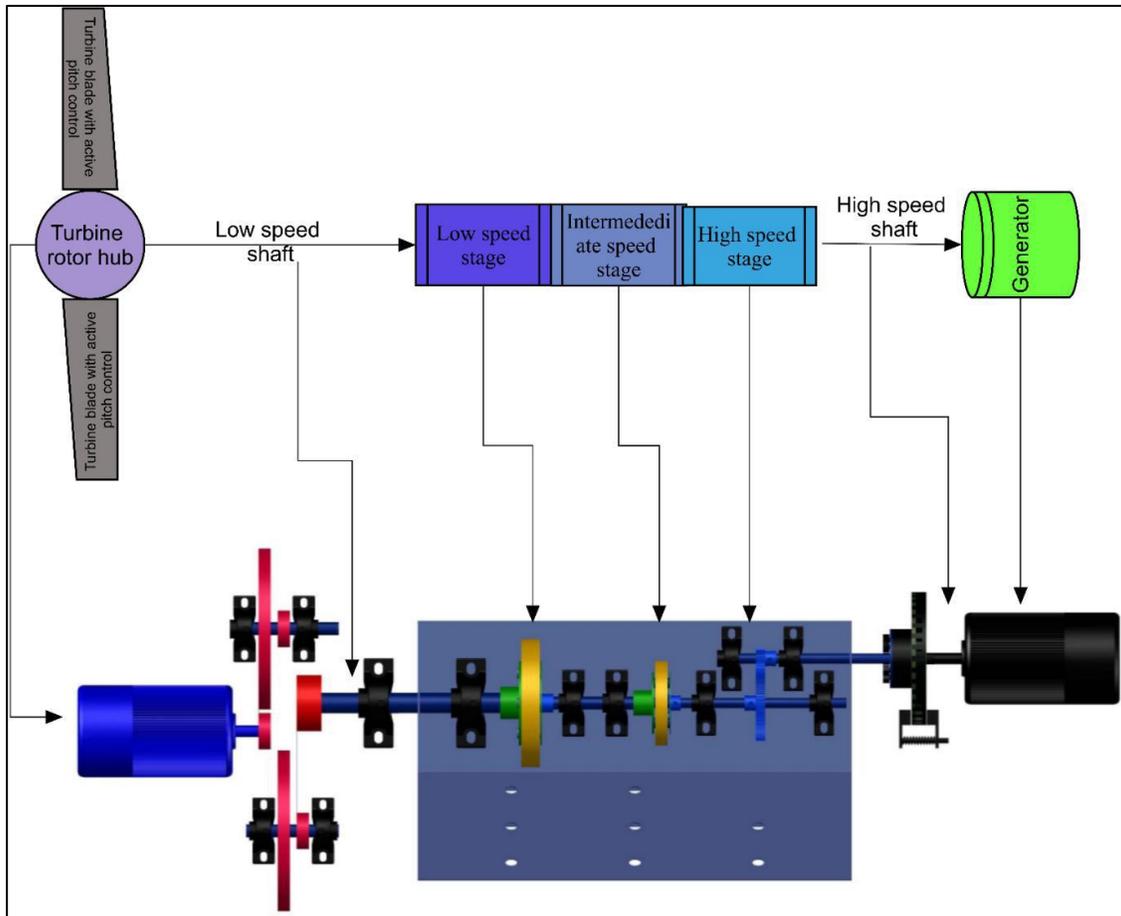
Figure 1. Schematic of Wind Turbine System Used

The design life of a wind turbine is about 20 years (Evans, 2012). However, reports have shown that they tend to breakdown prior to the end of the designed life cycle (Shuangwen Sheng and Paul Veers, 2011). Recent studies have revealed that most common system to fail is the gearbox. Gearbox failures can be attributed to bad lubrication, cooling, misalignment, shock loads and overloads (Shuangwen Sheng and Paul Veers, 2011). Overloads can be caused by external factors outside the gearbox such as a bad coupling, wind gust and misalignment during installation (Haastrup, Hansen, and Ebbesen, 2011). These can cause internal components to wear prematurely leading to internal overloads. Recent studies have attributed the primary cause of gearbox failures to bearing failures particularly, the high-speed shaft bearings and the planet bearings. The most commonly observed bearing failures are bearing spall and thermal distortion (Lu, Li, Wu, & Yang, 2009). The next prominent component to fail is the gear (Sheng, 2015). The high-speed shaft gear and planet gear tends to fail before the low speed stage gears. The most common faults noted in gears are pitting and scoring. A scheduled maintenance usually covers minor defects and preventive maintenance may not always be a viable option as it tends to be expensive and, in most cases, prevents utilising the full design life of the component. Hence, condition-based maintenance would contribute greatly to utilising the design life of the component and preventing unnecessary maintenance (Villa, Reñones, Perán, and De Miguel, 2012). Fault diagnosis entails methods to detect and isolate faults within a system to prevent catastrophic failures and be deemed as the crux of condition based maintenance (Yang, 2013). Fault detection systems monitor critical parameters such as vibration, temperature and acoustic levels in order to predict a possible failure. Such systems require three basic components; a sensor, a data logger and an output device which processes the data and generates a meaningful human readable result.

The heart of any monitoring equipment is the signal processing algorithm. Sophisticated devices use signal processing methods such as Fast Fourier transforms to plot frequency spectrums of the input signal, thereby providing an in-depth information of the health of the machine. However, when multiple components fail, the signals tend to be complex (Feng, Lin, & Zuo, 2016). Moreover, the signal acquired may not always be stationary in nature as the speed of the machine may vary. This adds to the

complexity and makes the signal non-stationary (Villa et al., 2012). Simple methods such as statistical parameter-based feature extraction may not always provide meaningful results as the signal can vary significantly and corrupting the statistical data. Also, analysing the frequency spectrum of the acquired data alone may lead to inaccurate results as the frequency emitted by the components during operations are linked to their rotational speed. This means that a change in the operating speed could cause shifting of the frequency peaks. This can lead to cases where the frequency peaks of the component being monitored may interfere with the frequency of other components in its proximity (Bisoi & Haldar, 2014). This may deter the machine learning algorithms from building accurate models which reflect the vibration patterns emitted by the machine or component being monitored. Alternatively, short time Fourier transform (STFT) can provide valuable insight into a signal as both the time and frequency information are available for correlation(Moosavian, Najafi, Ghobadian, and Mirsalim, 2017).

Many past researchers have investigated on various methods to deal with acquired signals so as to decode the component health encoded within the signal and have tried to address the challenges with non-stationary loading. Gryllias, Andre, Leclere, and Antoni (2017) proposed that when Cyclo-Non-Stationary Indicators are used along with a multi-order probabilistic methodology, it could pave way to angular speed tracking instantaneously for the condition-based monitoring of rotary machines operating under varying conditions. Cyclo-stationarity is a close approximation of a non-stationary case, to avoid the errors that arise due to this assumption, a study considering a case of non-stationarity without assumptions is needed which has been discussed in the present study. Zimroz, Bartelmus, Barszcz, and Urbanek (2014) proposed a procedure for the estimation of instantaneous speed and analysis of vibrations of planetary gearboxes under non-stationary operating conditions. This work focused on feature extraction by only statistical measures and signal processing methods were not implemented. Zhang, Han, and Deng (2017) proposed that feature extraction can be carried out by using a scaled spectrogram and subjecting it to a support vector machine (SVM) algorithm for automatic heart sound classification. Spectrogram features being extracted using a variance-based approach and machine learning techniques being used for speech recognition was demonstrated by Xie, Mcloughlin, Zhang, Song, and Xiao (2016).Ozer, Ozer, and Findik (2017) proposed a method to resize spectrogram images using the Lanczos kernel and used the same for automatic sound recognition and compared it using deep neural networks. Moosavian et al., (2017) investigated the effect of piston scratch on the vibration behaviour of an internal combustion engine using STFT and CWT. This does not discuss fault classification as it targets only a single type of fault obtained from a piston scratch. Muralidharan & Sugumaran, (2013)

proposed feature extraction from multiple datasets using wavelet transforms and classification using a decision tree-based machine learning algorithm, the J48 for the fault detection of a mono-block centrifugal pump. Again, this consisted of only a single fault and multi-component faults arising needs to be addressed and evaluated. Wang, Wang, and Wang (2018) proposed a method in which a generative adversarial learning deep neural networks can be used to analyse the vibration signature patterns of a planetary gearbox by combining Generative Adversarial Networks and Stacked Denoising Autoencoders. Singh & Parey, (2017) proposed applying angular vibration technique to both vibration and sound signatures emitted by a gearbox to estimate possible faults. Rajeswari, Sathiyabhama, Devendiran, and Manivannan (2014) proposed gear fault identification using wavelets and a variety of classification algorithms. Elangovan, Sugumaran, Ramachandran, and Ravikumar (2011) proposed SVM for the classification of vibration signatures of a single point cutting tool. Lei & Zuo, (2009) proposed using the algorithm pertaining to the weighted K nearest neighbour classification to recognise the level of gear crack in a system.

To the best of author's knowledge, past research on multi-component fault classification when subjected to non-stationary loading is rare. Despite efficient methods for fault classification such as using STFT, CWT, a robust methodology for automated classification of multi-component faults from a non-stationary dataset with little human intervention is needed for applications in real-time fault detection and classification. This paper focuses on realising and overcoming these drawbacks and aims to incorporate a unique methodology for a non-stationary dataset.

2. EXPERIMENTAL SETUP AND DATA ACQUISITION

A custom miniature wind turbine gearbox (see figure 1 for schematic 2 for gearbox) was constructed to facilitate the study. The details of the gearbox are shown in table 1. The planetary gearbox used in the present study was designed with a gear ratio of 1: 100. The design consists of three stages. Two planetary and one parallel stage. Both the planetary stages had a gear to speed ratio of 1:5 and the parallel stage had a gear to speed ratio of 1:4. The Low Speed Shaft was the input shaft. The imitation of the wind rotating the turbine blades and in turn the rotor was achieved by a series of belt and chain drives coupled to a motor, controlled by a Variable Frequency Drive (VFD). The speed reduction ratio of the section between the motor and the gearbox was 100:1. Since the experiment was tailored to mimic an actual wind turbine The Low Speed Shaft (LSS) was operated between 12 to 18 RPM to simulate an actual wind turbine in operation thereby limiting the high-speed shaft (HSS) to an operation range of 1200 to 1800 RPM.
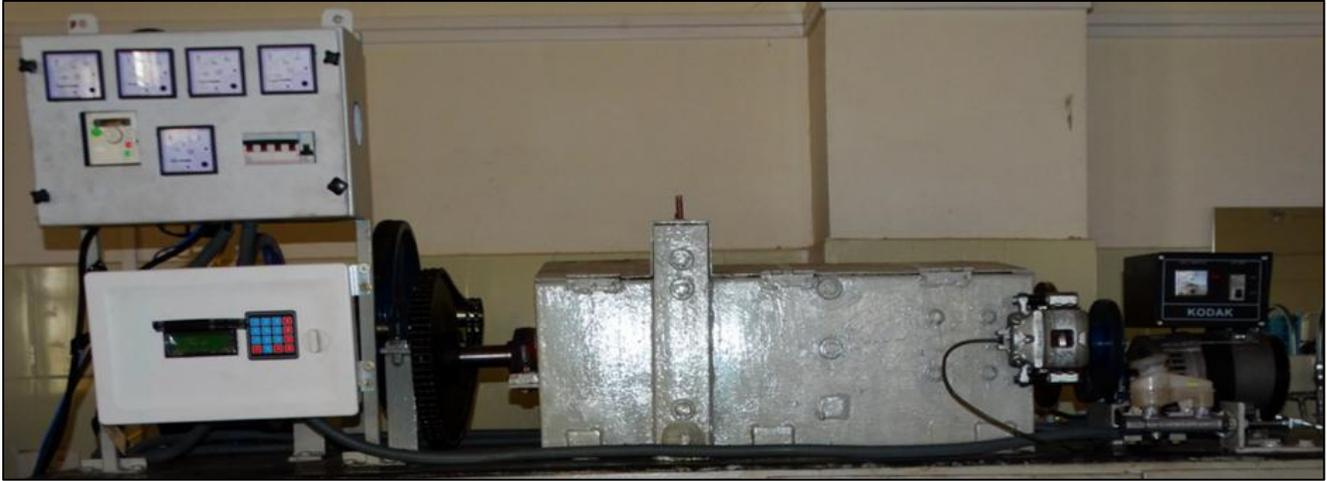
Figure 2. wind turbine planetary gearbox test rig

A synchronous generator (230V, 50Hz) was coupled to the high-speed shaft section to facilitate mechanical loading on the gearbox. This was achieved by electrically loading the generator using resistive loads which generated a mechanical resistance in the armature coil proportional to the electrical load (counter torque). The electrical load was in the form of 100W bulbs coupled to the generator assembly. Electrical loading was set to 100W which was approximately 17% of the generator's capacity. However, the AC synchronous generator was required to be maintained at a speed between 2500 RPM - 3500 RPM to ensure stable operation. As such a belt and pulley arrangement with a speed ratio of 1:2 was used for the generator. Two tri-axial accelerometers were placed at two distinct points on the gearbox Figure 3 depicts the raw waveform acquired from the intermediate-speed stage while being subjected to non-stationary loads.NI USB 4432 data acquisition system was used to acquire data from the two tri-axial accelerometers (PCB 356A43) mounted as shown in Figure 4. The X and Y axis information from both the accelerometers were collected simultaneously. They were designated as low-speed stage X axis (LSS X-axis), low-speed stage Y axis (LSS Y-axis), Intermediate-speed stage X axis (ISS X-axis) and intermediate-speed stage Y axis (ISS Y-axis).
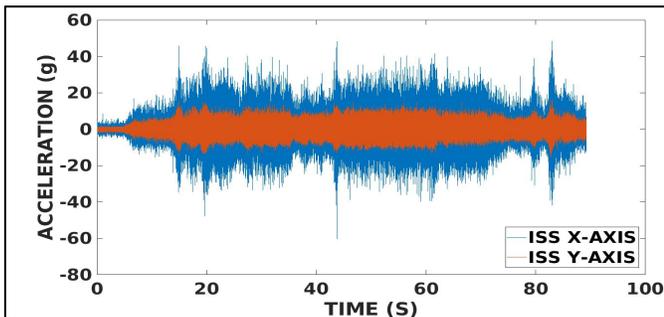

Figure 4. Planetary gearbox showing sensor placement

Property	Value
Rated power	1.5kW
Rated speed	3000RPM (high speed shaft)
Gear ratio	1:100
Gear Stage 1	Ratio 1:5 Module-3 mm
Gear Stage 2	Ratio 1:5 Module-2 mm
Gear Stage 3	Ratio 1:4 Module-1.5 mm
Lubrication type	Splash lubrication

Table 1. Planetary gearbox specifications

Figure 3. Raw vibration data depicting acceleration amplitude vs time

Figure 5. Bearing with single fault

Figure 6. HSS pinion with single fault

In order to study the behaviour of different faulty components, defects were simulated using wire cut EDM. The diameter of the wire was 0.2 mm. The components chosen for the study were the high-speed shaft bearing (model SYJ20TF) and a powder coated mild steel straight cut gear of 24 teeth having a module of 1.5mm which was used as the high-speed shaft pinion. Three bearing faults and one gear fault were simulated. Seven cases of faults which include multiple faults in combination were simulated. Figure 5 shows the bearing with a single faults. Figure 6 shows the pinion with a single fault and figure 7 shows the bearing with multiple fault. Components with more than one fault were treated as a multicomponent fault for the study. Table 2 shows the faults simulated for the study.
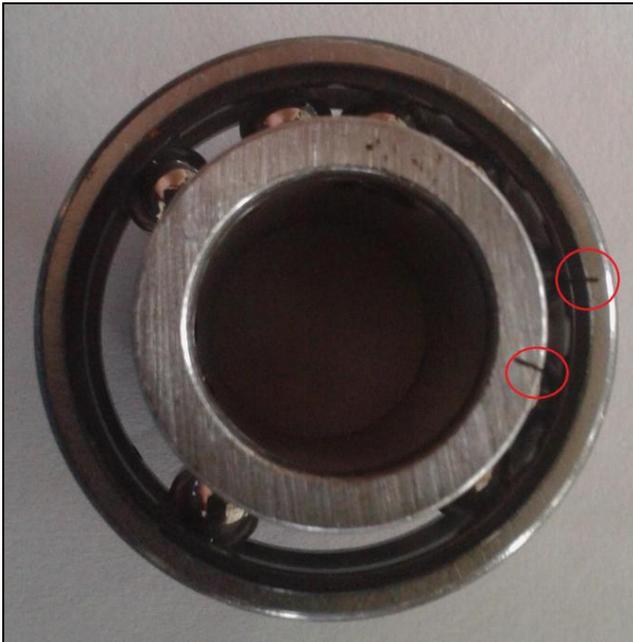

Figure 7. Bearing with multiple faults

Class	Fault	Code
1	Root crack	HRC
2	Inner race fault	IR
3	Outer race fault	OR
4	Inner + outer race fault	IO
5	Root crack + Inner race fault	IRHR
6	Root crack + Outer race fault	ORHR
7	Root crack + Inner + Outer race fault	IOHR
8	Healthy	Healthy

Table 2. Simulated faults

2.1. Simulation of Wind Speed

The motor speed was varied to replicate different wind speeds encountered by the wind turbine. The gearbox was run at varying speeds ranging from 1200RPM to 1800RPM at the high-speed shaft (HSS), this being the rated speed of an actual wind turbine (Shuangwen Sheng and Paul Veers, 2011). A simulated wind speed profile was used to set the motor speeds. This was done by generating a reference voltage for the VFD. A 0V is interpreted as 0% power and a 10V is interpreted as 100% power. The speeds were obtained using a random number generator which gave random values between 6 to 10 volts which corresponds to 60% and 100% of the VFD power respectively. The values in the speed profile were converted into a set of analogue voltages by means of a dedicated Micro-control unit (MCU), where 1200 RPM was about 60% of the VFD power and 1800 RPM was the maximum allowable VFD power. This speed profile generated can be considered equivalent to the one induced by non-stationary wind loading. This randomisation of the wind speed was done to ensure that no human intervention was needed and to facilitate the mimicking of an actual wind turbine scenario. The same wind speed profile was used for healthy as well as faulty cases. A sampling frequency of 22kHz was used so as to capture every intricate information embedded in the signal. The total time for which the vibration data collected from the accelerometers were for eight seconds with a sample length of 131072 data points which guaranteed that at least one full rotation was acquired by the DAQ. Fifteen such files were collected for the total duration of the wind load profile so as to ensure the repeatability of the data. Raw vibration signals acquired were uncompressed and stored as acceleration values with time information. In-order to extract useful information from the signal, feature extraction must be performed on the raw data. The vibration data acquired for the aforementioned cases (in table 2) was analysed using Short Time Fourier transform (STFT).

2.2. Data Processing using Short Time Fourier Transform

Short time Fourier transform (STFT) is employed as the feature extraction method due to its advantage of providing time as well as frequency information.
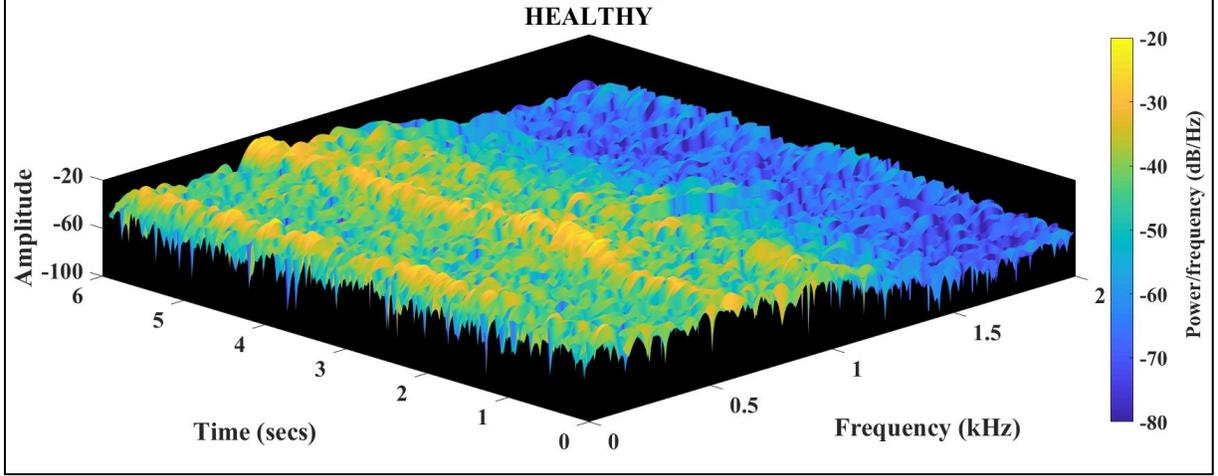
Figure 8. Spectrogram for healthy condition

The STFT is an assemblage of moving two-dimensional frames of Fast Fourier transforms (FFT), where each frame corresponds to an instant of time. The FFT computes an array of complex numbers for different frequencies of the signal. This is then computed multiple times at moving time instances, based on the sampling frequency, giving rise to a three-dimensional array of complex numbers possessing both time and frequency information. The plot obtained from these three axes is often referred to as a spectrogram. The time information in the spectrogram is essential as the location and occurrence of the fault can be predicted enabling an accurate classification of the fault using the extracted coefficients. Eq 1 Describes the general equation for STFT. Eq 2 computes STFT for a windowed signal.

$$S(t_1, p) = X(t_1, w)\big|_{w=\frac{2\Pi}{L}p} \tag{1}$$

$$S(n, w) = \sum_{-\infty}^{\infty} x[m]W[t_1 - m]e^{-jwt_1} \tag{2}$$

Terms: S- signal, W- window, w- frequency, m- magnitude, L- Number of frequencies, j- jump discontinuity, t_1-time, p-position.

Key parameters influencing the spectrogram are the window length, the number of FFT points (NFFT) and sampling frequency (Teng, Ding, Zhang, Liu, and Ma, 2016). Relatively low window lengths offer a high resolution along the time axis whereas a high resolution in the frequency axis is obtained at very high window lengths. In the present investigation an optimal value for the window length was arrived at; considering the above trade-offs.

The choice of the window length has been discussed in the next section. NFFT was chosen by the number of data points in each file (131072) to avoid any loss of information.
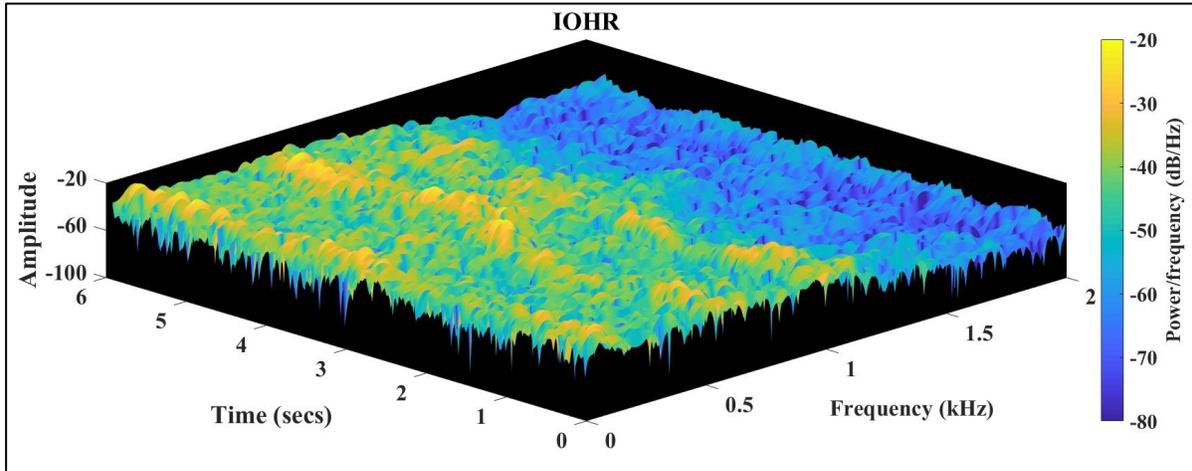

Figure 9. Spectrogram for IOHR condition

Sampling frequency as mentioned in the earlier section, was maintained at 22kHz. Initially, a window length of 64 was chosen and incremented in steps of 64 till 2048. Spectrograms were plotted for all the aforementioned window lengths and for all the four axes namely LSS-X, LSS-Y, ISS-X AND ISS-Y axes for which vibration data was collected. This was repeated for all fifteen files corresponding to each experiment and then further repeated for all eight experimented cases (see table 2).

Figure 8 and 9 are sample spectrogram plots of two datasets namely healthy and a multi-component fault, IOHR (see table 2). The input wind speed profile used for all the cases were the same, as such a comparison between the two spectrogram plots can be made. It can be observed that there are variations in the amplitude between the healthy and faulty cases.Spectrograms were used in earlier studies where the loading/ operating speed was stationary, to visually identify specific features between the healthy and faulty conditions. In the present investigation, under the non-stationary nature of loading, spectrogram plots of healthy when compared with 'Inner and Outer Race fault with pinion root crack (IOHR)' showed differences in amplitudes between frequencies of 0.5-1 kHz, corresponding to the time interval of 3-4 sec. However, when there are multi-component faults in the system, visual examination of the spectrogram plots may alone be insufficient to provide reliable information in order to distinguish between healthy and faulty conditions or between different faulty conditions. Hence, the use of machine learning algorithms is adopted so as to facilitate an automated process for fault classification. For classification of these faults the extracted features are given as an input to data-driven algorithms. The time information obtained in the dataset was neglected for training the algorithms so as to ensure that the algorithm uses only the extracted statistical features from the STFT coefficients as attributes for classification.

2.3. Data-driven approach Implementation

STFT was able to provide the information about the interested range of frequencies occurring at a particular instance of time. This signature arising from the spectrum of frequencies was found to be unique for each case of the induced fault and the extracted coefficients contained this information. To condense the vast dataset of points, descriptive statistics parameters, max, min, mean, median, mode, standard deviation, sample variance, range, kurtosis, RMS value, range, sum and skewness were taken for the frequency coefficients (vectors) at a particular time instance to minimise the size of the dataset. The statistics information derived from each vector of the extracted coefficient matrix was marked with its respective fault generating a matrix comprising of statistical data of every vector of the extracted STFT coefficients.
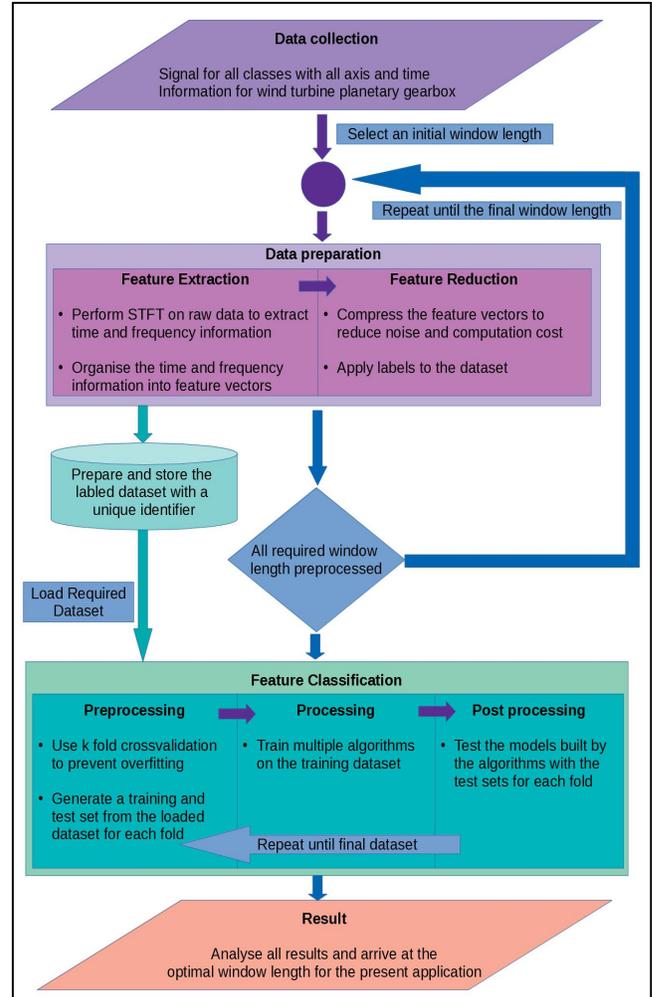

Figure 10. Methodology

This was culminated in a single file for every fault case. This dataset when subjected to a machine learning algorithm could be used to predict the type of fault. Figure 10 shows the methodology followed for the current study. The statistical parameters described above, computed from STFT coefficients was subjected as an input to machine learning algorithms. WEKA (Frank, Hall, and Witten, 2016), an open source machine learning toolbox was used to train and test the classification accuracy of the obtained datasets. Four decision tree algorithms– J48, Random forest, REP tree, and Random tree were chosen, as decision trees take minimal time for classification. The J48 algorithm is the C4.5 algorithm implemented in java and was developed by Ross Quinlan (Quinlan, 1996). The C4.5 algorithm is capable of building decision trees using a set of data available for training by the implementation of the concept of information entropy. The attributes are chosen by the algorithm in such a way that at each node subsets can be formed effectively. The splitting criterion is based on the normalized information gain, a corollary of difference in

7

entropy. The attribute having the largest difference in entropy shall be chosen to make the decision. The random tree algorithm evaluates multiple random features and builds a decision tree. It randomises the data several times and builds multiple trees. Random forest makes use of an ensemble of multiple random trees and can be considered the finest of the lot. REP tree uses the concept of information gain and reduced error pruning. Pruning, minimises the size of the tree by deleting attributes which are of little use for classification and hence minimises the classification time. REP tree minimises the errors arising from variance by generating multiple trees with every iteration and choosing the best of the lot (Frank et al., 2016). The number of training instances varied from 24,576 (window length of 2048) to 7,86,432 (window length of 64). Such a large number of instances coupled with K fold cross validation was pursued as it would minimise the chances of overfitting thereby constructing a valid machine learning model which would cover the entier operating range of the gearbox. Overfitting is a condition where an algorithm builds a model and achieves a very high accuracy for one supplied dataset but the model collapses when tested with another dataset (Frank et al., 2016). This usually occurs if cross fold validation is not used or if the dataset is very small. A fold can be defined as a dataset generated by shuffling the supplied data randomly by the algorithm. This is an automated process where the algorithm randomly reserves a portion of the dataset for training and another for testing. A fold comprises of both training and testing instances. The dataset consisted of all eight classes mentioned in table 2 and were given as input to the algorithm at the same time. Before training, the dataset was tagged with the respective fault. The dataset was initially subjected to a five-fold cross validation for all algorithms. Five-fold indicates that the dataset was randomised five times and validated using cross-validation.

3. RESULTS

The notable observations are presented in section 2.1. Detailed discussion by choosing a suitable window length and a suitable data-driven model is presented in section 2.2.

3.1. Observations

As mentioned earlier, window length was varied from 64 to 2048 in steps of 64 to obtain the coefficients of the spectrogram. Among the algorithms used for classification of these coefficients, the highest classification was obtained at the lowest window lengths which is 64 for all four algorithms (see figure 11). As the window length was increased, the classification accuracy marginally dropped. Among the four algorithms, it was observed that random forest had the highest classification accuracy and the J48 algorithm performed slightly better than the random tree algorithm in terms of classification accuracy. It can also be observed that the classification accuracy of the REP tree at

higher window lengths dropped substantially. An accuracy of 99.8% correctly classified instances was obtained at a window length of 64 for all four algorithms and the lowest was 93.6%, 75.2% ,75.7%, 62.4% for random forest, J48, random tree and REP tree algorithms respectively at a window length of 2048. As smaller window lengths possess a very high resolution along the time axis, the identification of fault location becomes more distinct and hence we obtain a high classification accuracy. Smaller window lengths also have a larger number of data points amounting to a higher computational time. Despite obtaining very high accuracy at smaller window lengths, the accuracy only marginally drops for increase in window lengths. Hence a compromise should be made between computation time and classification accuracy to obtain an optimal window length.
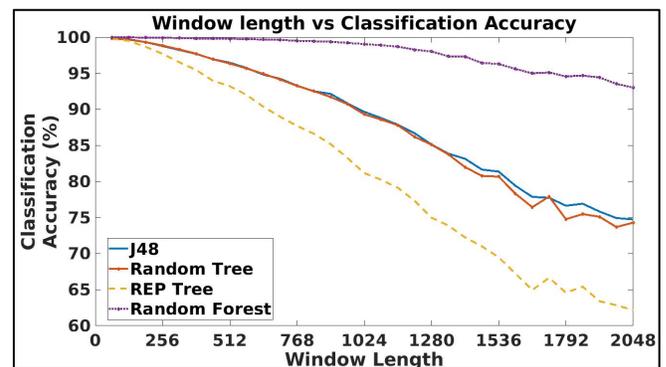

Figure 11. Window length vs classification accuracy for multiple algorithms
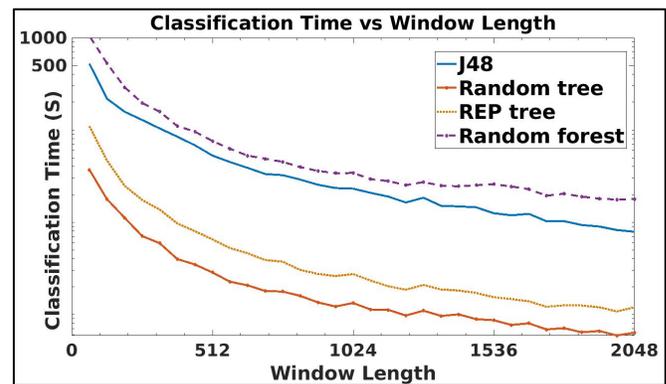

Figure 12. Classification time vs Window Length for multiple algorithms

From figure 11 it can be observed that the accuracy of classification of the random forest algorithm is the highest but the time taken to build the model is significantly higher than the rest (figure 12). Hence, if a larger dataset needs to be used to train the model, it can become computationally intensive. It can also be observed that the classification accuracy drops significantly for the REP tree algorithm at higher window lengths hence making it fairly inconsistent. Thus, it can be found that, for a window length of 576 and

for the random tree algorithm, the classification accuracy and the time for classification are optimal. The detailed study in the next section shall correspond to feature extraction using STFT at a window length of 576 and classification using the random tree algorithm subjected to an eight-fold cross validation. The algorithm shall be discussed in detail in the following section.

3.2. Discussions

Property	Value
Number of folds for cross-validation	8
Correctly Classified Instances	118856
Incorrectly Classified Instances	4024
Percentage of Correctly Classified Instances	96.725 %
Percentage of Incorrectly Classified Instances	3.275 %
Kappa statistic	0.9626
Mean absolute error	0.0082
Root mean squared error	0.0905
Relative absolute error	3.7426%
Root relative squared error	27.3589%
Total Number of Instances	122880

Table 3. Classification summary for Random tree

As chosen, the random tree algorithm was used to classify the STFT coefficients at a window length of 576. Table 3 shows the summary of the classifier after training and testing with eight-fold cross validation for STFT. A total of 122880 instances were taken out of which 118856 were correctly classified. Thus, a very high classification accuracy of 96.72% was achieved. The random tree algorithm follows a set of default rules in order to handle incompatible datasets.

- If it is seen that data in a particular list can be grouped in the same class, the algorithm creates a leaf node and the decision tree chooses that particular class.

- If no information gain is provided by any feature, the features are most likely indistinguishable. Should such a condition occur, the algorithm will create a decision node higher up the tree using an approximate value of the class.

- If an instance of previously-unseen class is encountered, the algorithm creates a decision higher up the tree using an approximate value.

Table 4 shows the detailed accuracy by class which can be used to identify potential classes which could cause inaccuracies during classification by increasing misclassification for its own class or other classes. The TP rate or true positive rate must approach 1 and the FP rate known as the false positive must approach 0 (Frank et al., 2016). Precision recall and F measure can be termed as performance parameters of the algorithm. When it comes to pattern recognition, retrieval of information and classification, precision is the ratio of relevant instances to the retrieved instances, whereas recall is the ratio of relevant retrieved instances to the total number of relevant instances.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.968	0.005	0.968	0.968	0.968	0.964	0.982	0.941	**ORHR**
0.964	0.005	0.965	0.964	0.965	0.960	0.979	0.935	**OR**
0.968	0.005	0.967	0.968	0.967	0.963	0.981	0.940	**IOHR**
0.969	0.005	0.968	0.969	0.968	0.964	0.982	0.941	**IO**
0.964	0.005	0.964	0.964	0.964	0.959	0.980	0.934	**IRHR**
0.973	0.004	0.969	0.973	0.971	0.967	0.984	0.947	**IR**
0.963	0.005	0.965	0.963	0.964	0.959	0.979	0.934	**HR**
0.969	0.004	0.971	0.969	0.970	0.966	0.982	0.945	**Healthy**

Table 4. detailed accuracy by class

ORHR	OR	IOHR	IO	IRHR	IR	HR	Healthy	<-- classified as
14873	80	67	73	82	62	61	62	**ORHR**
66	14805	75	84	96	81	87	66	**OR**
80	71	14862	74	65	60	90	58	**IOHR**
74	68	72	14877	67	71	67	64	**IO**
78	90	64	74	14811	79	81	83	**IRHR**
57	63	59	54	68	14948	70	41	**IR**
76	76	100	68	108	61	14798	73	**HR**
62	83	66	67	60	61	79	14882	**Healthy**

Table 5. Confusion Matrix for Random Tree

The quality of the classification can be quantified by MCC or Matthews's correlation coefficient. The MCC produces a value between −1 and +1. A coefficient of +1 represents a perfectly correct classification, 0 which is nothing but a random classification and −1 indicates a total disagreement between classification and observation. A receiver operating characteristic curve, i.e. ROC curve, is a plot that depicts the ability to diagnose the classifier system.

The ROC curve is obtained by plotting the true positive values against the false positive values whose corresponding values have been shown in table 4. Both ROC and PRC (precession recall rate) must be above 0.5 to ensure that the classification process is not occurring randomly.

From Table 4, it can be noted that the TP and FP rate had marginal changes which clearly indicates the faults did not interfere with one another. An interesting observation was that though the components involved in IOHR were located close to each other, with the bearing fault being the weak signature, the fault signature was not masked by the more dominant HSS pinion signal. Table 5 shows the confusion matrix provided by the classifier. The confusion matrix is also known as an error matrix and is used to visualise the performance of an algorithm.

In the confusion matrix the diagonal elements are the correctly classified instances and the other elements are the misclassified elements. Both the confusion matrix and the detailed accuracy by class can provide valuable information on classifier performance and possible misclassification due to complex signals within a class.

4. CONCLUSION

The present study investigated the practicality of implementing data-driven algorithms to categorize multi-component faults from a composite non-stationary signal which was achieved by extracting features using a popular signal processing method, STFT.

- Coefficients of the STFT were extracted for multiple window lengths and an optimum window length of 576 was arrived at as the classification accuracy and the classification time was optimal.

- A detailed analysis of the dataset was performed using STFT for a window length of 576 and classification was performed using the random tree algorithm.

- An eight-fold cross validation was performed for the same to ensure repeatability and avoid over fitting.

- Random tree algorithm scored an accuracy of 96.72% and it was observed that the compound faults did not interfere with each other.

The window length of 576 will be valid for a gearbox which produces a vibration response with an identical frequency spectrum. The response frequency primarily depends on the operating speed but may also vary with other parameters such as gear mesh frequency and material damping. Furthermore, the window length will be valid only if the sample rate is maintained at 22KHz. As such, the conclusions are application specific and the corresponding window length was arrived at as it gave the best classification accuracy.

However, the automated fault detection process employed here for detecting multi-component faults from a wind turbine gear box subjected to non-stationary loading can be extended to any gear box subjected to non-stationary loading.

REFERENCES

Bisoi, S., & Haldar, S. (2014). Dynamic analysis of offshore wind turbine in clay considering soil-monopile-tower interaction. *Soil Dynamics and Earthquake Engineering*, *63*, 19–35. https://doi.org/10.1016/j.soil dyn.2014.03.006

Elangovan, M., Sugumaran, V., Ramachandran, K. I., & Ravikumar, S. (2011). Effect of SVM kernel functions on classification of vibration signals of a single point cutting tool. *Expert Systems with Applications*, *38*(12), 15202–15207. https://doi.org/10.1016/j.eswa.2011.05. 081

Evans, M.-H. (2012). White structure flaking (WSF) in wind turbine gearbox bearings: effects of 'butterflies' and white etching cracks (WECs). *Materials Science and Technology*, *28*(1), 3–22. https://doi.org/10.1179 /026708311X13135950699254

Feng, Z., Lin, X., & Zuo, M. J. (2016). Joint amplitude and frequency demodulation analysis based on intrinsic time-scale decomposition for planetary gearbox fault diagnosis. *Mechanical Systems and Signal Processing*, *72–73*, 223–240. https://doi.org/10.1016/j.ymssp. 2015.11.024

Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA Workbench. *Morgan Kaufmann, Fourth Edition*, 553– 571. https://doi.org/10.1016/B978-0-12-804291-5.000 24-6

Gryllias, K., Andre, H., Leclere, Q., & Antoni, J. (2017). Condition monitoring of rotating machinery under varying operating conditions based on Cyclo-Non-Stationary Indicators and a multi-order probabilistic approach for Instantaneous Angular Speed tracking. *IFAC-PapersOnLine*, *50*(1), 4708–4713. https://do i.org/10.1016/j.ifacol.2017.08.857

Haastrup, M., Hansen, M. R., & Ebbesen, M. K. (2011). Modeling of wind turbine gearbox mounting. *Modeling, Identification and Control*, *32*(4), 141–149. https://doi.org/10.4173/mic.2011.4.2

Lei, Y., & Zuo, M. J. (2009). Gear crack level identification based on weighted K nearest neighbor classification algorithm. *Mechanical Systems and Signal Processing*, *23*(5), 1535–1547. https://doi.org/10.1016/j.ymssp. 2009.01.009

Lu, B., Li, Y., Wu, X., & Yang, Z. (2009). A review of recent advances in wind turbine condition monitoring and fault diagnosis. *Electronics and Machines in Wind*, 1–7. https://doi.org/10.1109/PEMWA.2009.5208325

Moosavian, A., Najafi, G., Ghobadian, B., & Mirsalim, M. (2017). The effect of piston scratching fault on the vibration behavior of an IC engine. *Applied Acoustics*, *126*, 91–100. https://doi.org/10.1016/j.apacoust. 2017.05.017

Muralidharan, V., & Sugumaran, V. (2013). Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. *Measurement*, *46*(1), 353– 359.https://doi.org/10.1016/j.measurement.2012.07.00 7

Ozer, I., Ozer, Z., & Findik, O. (2017). Lanczos kernel based spectrogram image features for sound classification. *Procedia Computer Science*, *111*(2015), 137–144. https://doi.org/10.1016/j.procs.2017.06.020

Rajeswari, C., Sathiyabhama, B., Devendiran, S., & Manivannan, K. (2014). A Gear fault identification using wavelet transform, rough set based GA, ANN and C4.5 algorithm. *Procedia Engineering*, *97*, 1831– 1841. https://doi.org/10.1016/j.proeng.2014.12.337

.Quinlan, J. R(1996). Improved Use of Continuous Attributes in C4 . 5, *Journal of Artificial Intelligence Research, 4, 77-90.* https://doi.org/10.1613/jair.279

Sheng, S. (2015). Wind Turbine Gearbox Reliability Database, Condition Monitoring, and O&M Research Update. *PR-5000-63868, National Renewable Energy Laboratory (NREL), Golden, CO (US),*.

Shuangwen Sheng and Paul Veers. (2011). Wind Turbine Drivetrain Condition Monitoring - An Overview. *Machinery Failure Prevention Technology (MFPT): The Applied Systems Health Management Conference 2011*, (October).

Singh, A., & Parey, A. (2017). Gearbox fault diagnosis under non-stationary conditions with independent angular re-sampling technique applied to vibration and sound emission signals. *Applied Acoustics*. https://doi.org/10.1016/j.apacoust.2017.04.015

Teng, W., Ding, X., Zhang, X., Liu, Y., & Ma, Z. (2016). Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform. *Renewable Energy*, *93*, 591–598. https://doi.org/ 10.1016/j.renene.2016.03.025

Villa, L. F., Reñones, A., Perán, J. R., & De Miguel, L. J. (2012). Statistical fault diagnosis based on vibration analysis for gear test-bench under non-stationary conditions of speed and load. *Mechanical Systems and Signal Processing, , 29*, 436–446. https://doi.org/10. 1016/j.ymssp.2011.12.013

Wang, Z., Wang, J., & Wang, Y. (2018). An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition. *Neurocomputing*. https://doi.org/10.1016/j.neucom. 2018.05.024

Xie, Z., Mcloughlin, I., Zhang, H., Song, Y., & Xiao, W. (2016). A new variance-based approach for discriminative feature extraction in machine hearing classification using spectrogram features. *Digital Signal Processing: A Review Journal*, *54*, 119–128. https://doi.org/10.1016/j.dsp.2016.04.005

Yang, W. (2013). Condition Monitoring the Drive Train of a Direct Drive Permanent Magnet Wind Turbine Using Generator Electrical Signals. *Journal of Solar Energy Engineering*, *136*(2), 021008. https://doi.org/10.1115/1.4024983

Zhang, W., Han, J., & Deng, S. (2017). Heart sound classification based on scaled spectrogram and tensor decomposition. *Expert Systems with Applications*, *84*, 220–231. https://doi.org/10.1016/j.eswa.2017.05.014

Zimroz, R., Bartelmus, W., Barszcz, T., & Urbanek, J. (2014). Diagnostics of bearings in presence of strong operating conditions non-stationarity - A procedure of load-dependent features processing with application to wind turbine bearings. *Mechanical Systems and Signal Processing*, *46*(1), 16–27. https://doi.org/10.1016/j.ymssp.2013.09.010