

Data-driven monitoring of the gearbox using multifractal analysis and machine learning methods

Andrzej Puchalski^{1*}, and Iwona Komorska¹

¹University of Technology and Humanities, Faculty of Mechanical Engineering, 26-600 Radom, Poland

Abstract. Data-driven diagnostic methods allow to obtain a statistical model of time series and to identify deviations of recorded data from the pattern of the monitored system. Statistical analysis of time series of mechanical vibrations creates a new quality in the monitoring of rotating machines. Most real vibration signals exhibit nonlinear properties well described by scaling exponents. Multifractal analysis, which relies mainly on assessing local singularity exponents, has become a popular tool for statistical analysis of empirical data. There are many methods to study time series in terms of their fractality. Comparing computational complexity, a wavelet leaders algorithm was chosen. Using Wavelet Leaders Multifractal Formalism, multifractal parameters were estimated, taking them as diagnostic features in the pattern recognition procedure, using machine learning methods. The classification was performed using neural network, k-nearest neighbours' algorithm and support vector machine. The article presents the results of vibration acceleration tests in a demonstration transmission system that allows simulations of assembly errors and teeth wear.

1 Introduction

Monitoring the operation of a machine, we assume that detection, isolation and identification of a damage requires a comparison of the current state to a state taken as a reference. Diagnosis can therefore be considered a process in line with the paradigm of pattern recognition. Properly defined purpose, data collection, data cleaning and normalisation, data transformation and selection of relevant diagnostic features, as well as classification based on the chosen exploration technique are required. In view of the intrinsic dynamic nature and external excitations, most of the real signals coming from the monitoring of mechanical vibrations of rotating machines has nonlinear, non-stationary and multiscale properties both in time and space [1]. Traditional methods of signal analysis in engineering applications are based on the assumption of stationarity. Such methods can not reveal local features in time and frequency domains.

One of more frequently used nonlinear methods is the study of point regularity of time series, represented by the Holder scaling exponent. The estimation of scaling exponents leads to determination of a multifractal spectrum, which is a statistical model of the monitored signal [2]. This fact was used to determine the diagnostic features of the test gear being tested and to identify assembly errors and wear. Small fluctuations of shaft rotational speed occurred during the test, but the proposed method does not require synchronisation and resampling of the data, as in the case of e.g. Fast Fourier Transform (FFT). In the studies of vibration signals in terms of their multifractality, methods of multifractal version of the detrending fluctuation analysis (MF-DFA) [3 – 9] and

various methods based on wavelet analysis [10 - 12] are applied.

The article analyses the possibility of using wavelet leaders to select features and neural network, k-nearest neighbours and support vector machines in the process of fault classification of recorded data from the pattern of the monitored system and inferring the occurrence of damage.

2 Wavelet leaders multifractal formalism

This paragraph presents the idea of quantities called wavelet leaders, which are determined on the basis of wavelet coefficients and are representative of Holder exponents [13 – 15]. Multifractal analysis based on leaders is not dependent on wavelet selection, provided that the number of zero wavelet moments is greater than the largest Holder exponent of the signal.

The wavelet leaders were calculated based on the discrete wavelet transform (DWT). The point Holder exponent of the function/signal $x(t)$ in point t denotes the number h defined as the supremum of all exponents satisfying, for a certain $C > 0$, the condition:

$$|x(t) - P_n(t - t_0)| \leq C|t - t_0|^h \quad (1)$$

where $P_n(t - t_0)$ is a polynomial of $n < h$ degree.

Information on variability of regularity $x(t)$ along t can be described by the multifractal spectrum, estimated using wavelet leaders multifractal formalism.

Let us define discrete wavelet transform of the function $x(t)$ and mother wavelet $\psi_0(t)$ with the compact support using its coefficients:

* Corresponding author: andrzej.puchalski@uthrad.pl

$$d_x(j, k) = \int_R x(t)2^{-j}\psi_0(2^{-j}t - k)dt \quad (2)$$

Moreover, let us introduce the dyadic intervals:

$$\lambda = \lambda_{j,k} = [k2^j, (k+1)2^j] \quad (3)$$

and the union of such neighbour intervals:

$$3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1} \quad (4)$$

The quantities:

$$L_x(j, k) = \sup_{\lambda' \in 3\lambda} |d_{\lambda'}| \quad (5)$$

illustrated in Figure 1 are referred to as wavelet leaders.

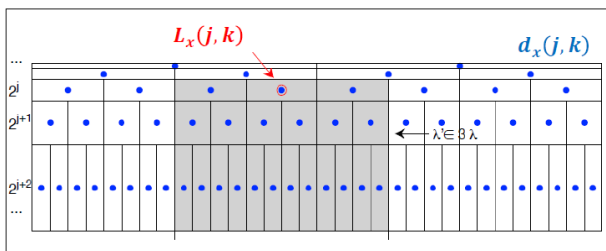


Fig. 1. Wavelet leader interpretation.

If $x(t)$ has a Holder exponent h at a point $t_0 = 2^j k$:

$$h = h(t_0) = \lim_{j \rightarrow 0} \frac{\log L_x(j, k)}{\log 2^j} \quad (6)$$

then the corresponding wavelet leaders decay as power laws of scales:

$$L_x(j, k) \sim 2^{jh} \quad (7)$$

The structure functions defined as sample mean estimators for the ensemble averages of q -power of wavelet leaders exhibit power law behaviours with respect to the analysis scale:

$$Z_L(q, j) = \frac{1}{n_j} \sum_{k=1}^{n_j} L_x(j, k)^q = \mathbb{E} L_x(j, k)^q \sim 2^{j\zeta(q)} \quad (8)$$

The scaling exponents $\zeta(q)$ can be expanded using cumulants c_p :

$$\begin{aligned} \zeta(q) &= \lim_{j \rightarrow 0} \frac{\log_2 Z_L(q, j)}{j} = \\ &= \sum_{p=1}^{\infty} c_p \frac{q^p}{p!} = c_1 q + c_2 \frac{q^2}{2} + c_3 \frac{q^3}{6} + \dots \end{aligned} \quad (9)$$

$\forall p \geq 1$ c_p satisfy the relation:

$$C^L(j, p) = c_p \ln 2^j \quad (10)$$

where $C^L(j, p)$ are the $p \geq 1$ order cumulants of the $\ln L_x(j, \cdot)$.

Moreover c_1 describes the location of the multifractal spectrum maxima and c_2, c_3 describe the level of multifractality: spectrum width and spectrum asymmetry, respectively.

This leads to a Wavelet Leader based Multifractal Formalism

$$D(h) = 1 + qh - \zeta(q) \quad (11)$$

Under some mild regularity conditions $D(h)$ acts as an upper bound for multifractal spectrum of analysed signal.

3 Methods of fault classification

3.1 Neural Network NN

Artificial neural networks are relatively crude electronic networks of neurons based on the neural structure of the brain. They process records one at a time and learn by comparing their classification of the record (i.e., largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network and used to modify the networks algorithm for further iterations.

In the training phase, the correct class for each record is known (termed supervised training), and output nodes can be assigned correct values. The typical back-propagation network has an input layer, an output layer, and at least one hidden layer. There is no theoretical limit on the number of hidden layers but typically there are just one or two [16,17].

3.2 k-Nearest Neighbours k-NN

The k-NN algorithm is among the simplest of all machine learning algorithms. In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour. The neighbours are taken from a set of objects for which the class is known. This can be thought of as the training set for the algorithm, though no explicit training step is required [18, 19].

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples.

In the classification phase, k is a user-defined constant, and an unlabelled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point.

A commonly used distance metric for continuous variables is Euclidean distance. In the Euclidean plane, if $p = (p_1, p_2)$ and $q = (q_1, q_2)$ then the distance is given by

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \quad (12)$$

Various metrics can be used to determine the distance, for example Mahalanobis, city block, Minkowski, Chebychev, cosine, correlation, Hamming, Jaccard, Spearman distance.

3.3 Support Vector Machines SVM

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyse data used for classification. Given a set of training examples, each marked as belonging to one or the other two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier [20, 21].

A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalisation error of the classifier.

Given a training vector $\{\mathbf{x}|\mathbf{x}_i \in \mathbf{R}^p\}_{i=1}^n$ and an indication vector $\{\mathbf{y}|\mathbf{y}_i \in \{-1,1\}\}_{i=1}^n$, here the training vector contains two different categories and the element \mathbf{y}_i in the indication vector serves to indicate which category the sample \mathbf{x}_i in the training vector belongs to. Then, an SVM is trained to find a maximum-margin hyperplane for distinguishing the samples marked with $\mathbf{y}_i = -1$ from those marked with $\mathbf{y}_i = 1$. If there is no hyperplane which can separate these two categories, a soft margin method can be employed to determine a hyperplane which can separate these two categories, still keeping the most maximum distance to the closest clearly separated points.

The original SVM can be used only for solving a binary classification problem. Nonetheless, the multi-class classification is a frequently faced problem in the real world. To solve this problem, “one-against-one” and “one-against-all” approaches can be proposed [22].

4 Gearbox fault diagnosis

An experimental system, which photo is in Figure 2, makes up of the drive motor, one-stage planetary gearbox, and hydraulic brake. The main measurement path included B&K DeltaTron type 4519-002 acceleration sensor with a frequency range of 1 Hz-20 kHz, screwed to the bearing housing and DBK04 dynamic signal acquisition card equipped with a set of antialiasing filters cooperating with the main IQtech DaqBook 2005 acquisition module, which transmits data using an Ethernet connection to the computer. Data were registered with sampling frequency 10kHz during about 100 seconds (about 1 million samples).

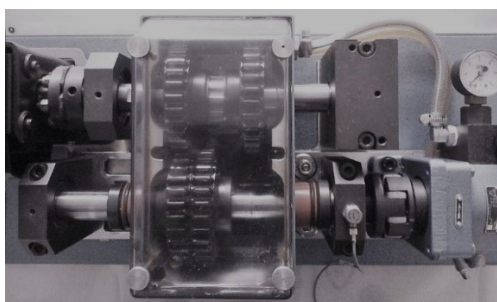


Fig. 2. Experimental setup.

Three types of faults were simulated on the test bench: misalignment 1/3°, enlarged clearance 0,4mm and worn gears. Figure 3 displays the vibration waveforms for fault-free and faulty states of the gearbox.

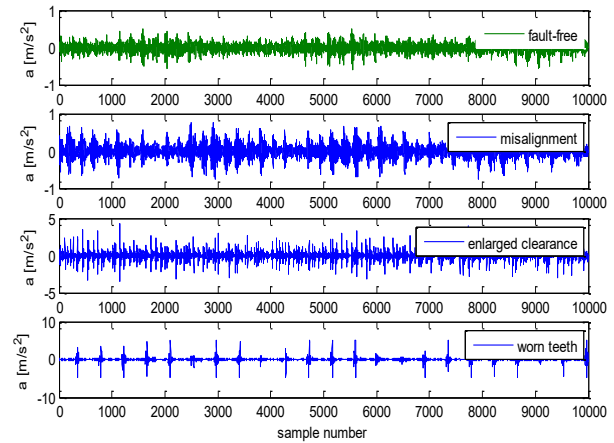


Fig. 3. Vibration acceleration for four states of gearbox under the motor running speed 1365 rpm and the 15% of max load during 1second.

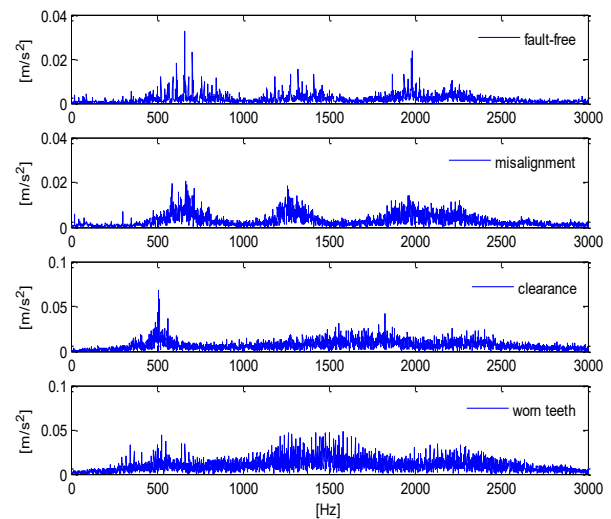


Fig. 4. Vibration FFT spectra for four states of gearbox under the motor running speed 1365 rpm and the 15% of max load during 1second.

The motor rpm and load were slightly changing during the test, so the signal is not stationary. So, the Fast Fourier Transform (FFT) for signal analysis without resampling is not a good solution. In turn, resampling does not provide sufficient accuracy in on-line analysis. FFT spectra of raw vibration signal (without resampling) for different states of gearbox are shown in Figure 4. Slightly changing rotational speed causes the frequency spectrum to be modulated by variable frequency and difficult to compare.

Then for each gearbox condition, 100 pieces of data were gathered, each of the length of 8192 points. The Multifractal Wavelet Leaders method was adopted to analyse these gearbox vibration data. The multifractal spectra of sample data are demonstrated in Figure 5.

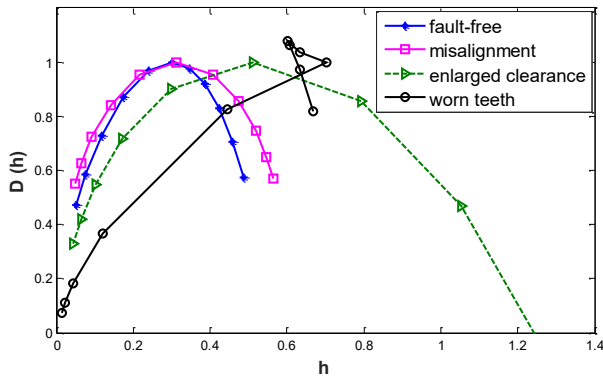


Fig. 5. Multifractal spectra of different gearbox condition under motor running speed 1365 rpm and the 15% of max load during 1second.

Obtained spectra can be treated as multifractal dimensions $D(h)$ associated with singularities h representing local scaling of the measure at various places in the time series. The description of the dynamic properties of the system based on the multifractal spectrum of time series is permitted by: the singularity with the largest dimension, i.e. the most often met the time series singularity and dimensions span of the singularities subsets. These parameters are represented by first and second cumulants, respectively.

Next, for each piece of four gearbox condition, the cumulants were calculated. For classification purpose first two cumulants were chosen. Figure 6 shows a scatter plot second cumulant versus first cumulant.

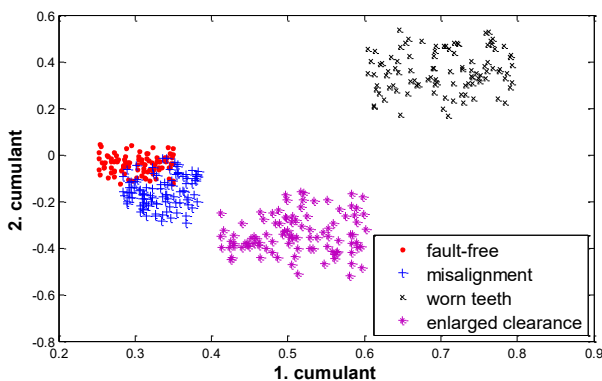


Fig. 6. Second versus first cumulant for four gearbox states and 100 pieces of the vibration signal.

Since the scratch teeth and enlarged clearance states can be easily separated, separation of fault-free and misalignment states may be more complicated. Three methods of classification were compared for separation of the states: neural networks (NN), nearest neighbours (kNN) and Support Vector Machine (SVM) using Matlab environment.

Each of four sets registered for different fault states was divided into a training set (70% of samples) and a test set (30% of samples) and then subjected to machine learning. Using NN method a two-layer, feed-forward network with 10 sigmoid hidden neurons and linear output neurons was performed. The network was trained with Levenberg-Marquardt back propagation algorithm. In the k-NN method, $k = 3$ was assumed. In SVM method the

“one-against-one” approach was employed to solve the problem of classification. The results of the classification are summarised in Table 1.

Table 1. The results of the correct classification of the system state for test sets.

Classification method	NN	k-NN	SVM
Fault type	Classification accuracy [%]		
fault-free	90	87	90
misalignment	93	93	87
enlarged clearance	100	100	100
worn teeth	100	100	100

The states: worn teeth and enlarged clearance were separated perfectly. Some of the fault-free sets were recognised as the misalignment sets and vice versa. This can be explained by the fact that during dynamic systems work, the assembly errors change and may temporarily give incorrect classification results. When analysing the time course of the vibration signal (see Figure 3), it can be noticed that some cycles of fault-free and misalignment states have a similar character. However, these are occasional errors occurring once in a dozen or so cases. If there is a permanent change in the system state, the mean value of the instant cumulants will move (see Figure 6).

5 Conclusions

The paper proposed Multifractal Wavelet Leader method for the automated recognition of the condition of the gearbox based on the monitored vibration signal. Using the wavelet leaders two first cumulants as the signal features were estimated. Then the methods of machine learning were employed for classification of the gearbox state: neural network, nearest neighbours and support vector machines. Regardless of the classification method, satisfactory results were obtained.

The best classification results were obtained for the NN method. The weakest results of the classification of sets: fault-free and misalignment are due to their overlap. This can be explained by the similarity of some fragments of the vibration signal for two states.

Despite the complicated mathematical notation, the wavelet leaders method is characterised by low computational complexity and can be performed on-line.

References

1. W. Batko, Z. Dąbrowski, J. Kiciński, *Nonlinear effects in technical diagnostics* (ITE-PIB Radom, 2008)
2. A. Puchalski, *J. of Vibroeng.* **17**(7) 3586–3593 (2015)
3. I.W. Kantelhardt, *Math. of Compl. and Dyn. Sys.* 463–487 (2012)
4. I.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, H.E. Stanley, *Phys. A* **316**, 87-114 (2002)

5. E.P. Moura, A.P. Vieira, A.S. Irmao, A. A. Silva, *Mech. Sys. & Sign. Proc.* **23**, 682-689 (2009)
6. J. Lin, Q. Chen, *Mech. Sys. & Sign. Proc.* **38**, 515-533 (2013)
7. H. Liu, X. Wang, C. Lu, *Mech. Sys. & Sign. Proc.* **60-61**, 273-288 (2015)
8. A. Puchalski, I. Komorska, *Appl. Condition Monitoring* (Springer Int. Pub. AG 9, 91-101 2018)
9. A. Puchalski, I. Komorska, *J. of Vibroeng.* **20(4)**, 1667-1675 (2018)
10. W. Du, J. Tao, Y. Li, C. Liu, *Mech. Sys. & Sign. Proc.* **43**, 57-75 (2014)
11. R. Yan, R.X. Gao, X. Chen, *Sign. Proc.* **96**, 1-15 (2014)
12. H. Bendjama, M.S. Boucherit, *J. of Theor. and Appl. Mech.* **54(2)**, 659-670 (2016)
13. H. Wendt, S.G. Roux, S. Jaffard, P. Abry, *Sign. Proc.* **89(6)**, 1100-1114 (2009)
14. S. Jaffard, B. Lashermes, P. Abry. *Appl. & Num. Harm. Analysis* (Birkhauser Verlag, 201-246, 2006)
15. E. Serrano , A. Figliola, *Physica A* **388**, 2793-2805 (2009)
16. S. Grossberg, *Neural Neyworks* **1**, 17-61 (1988)
17. S. Haykin, *Neural Networks: A comprehensive foundation* (Prentice Hall, New Jersey 2004)
18. N.S. Altman, *The American Statistician*, **46** (3), 175-185 (1992)
19. P. Hall, B.U. Park, R.J. Samworth *Annals of Statistics*, **36** (5), 2135-2152 (2008)
20. N. Christianini, J.C. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (Cambridge University Press, Cambridge UK 2000)
21. B. Scholkopf, A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond, Adaptive Computation and Machine Learning* (The MIT Press, Cambridge, MA 2002)
22. J. Lin, C. Dou, Q. Wang, *Sound&Vibration*, **4**, 11-15 (2018)