

Fatigue Feature Clustering Algorithm Using the Morlet Wavelet

T. E. PUTRA¹, S. ABDULLAH², M. Z. NUAWI²

Department of Mechanical Engineering

Universitas Syiah Kuala

Darussalam Banda Aceh 23111

INDONESIA

edi_unsyiah@yahoo.co.id

Department of Mechanical and Materials Engineering

Universiti Kebangsaan Malaysia

43600 UKM Bangi Selangor

MALAYSIA

Abstract: - This paper presents clustering of fatigue features resulted from the segmentation of the SAESUS time series data. The segmentation process was based on the Morlet wavelet coefficient amplitude level which produces 49 segments that each has an overall fatigue damage. Observation of the fatigue damage and the wavelet coefficient was made on each segment. In the end of the process, the segments were clustered into three clusters to identify any improvements in the data scattering for fatigue data clustering prospects. This algorithm produced a more reliable and suitable method of segment by segment analysis for fatigue strain signal segmentation. According to the analysis findings, the higher Morlet wavelet coefficient presented damaging segment, otherwise, it was undamaging segment. It indicated that the relationship between the Morlet wavelet coefficient and the fatigue damage was strong and parallel.

Key-Words: - Fatigue strain signal, segmentation, fatigue damage, Morlet wavelet coefficient, clustering

1 Introduction

Occasionally, fatigue signal measured from critical automotive parts has variable amplitude pattern with mean value of the data is change with time (each pattern has different statistical value). It contains large percentage of small amplitude cycles and the fatigue damage for these cycles can be small. For this reason, in many cases, the signal was edited by removing these cycles in order to produce representative and meaningful yet economical testing [1].

Segmentation is aimed to remove lower or minimal damaging features of an original signal. It is performed by identifying and extracting segments which contributes to the more fatigue damaging events to a metallic material. On the other hand, segments containing lower amplitude are omitted, since these data type is not damaging. The goal of the removal of those parts from the original signal is to generate a new shortened edited signal so reduce the testing time and cost for fatigue testing purpose [2]. This method is also known as the fatigue feature extraction.

This paper discussed on the clustering of fatigue data (represented as time series) by evaluating the fatigue damage and the Morlet wavelet coefficient of each segment, resulting data scattering, and clustering the data. It is hypothesized that the fatigue damage and the Morlet wavelet coefficient have a strong correlation. The segmentation always was being used for classified

data in order to analyse discrete data in time domain in vibration and fatigue data analyses.

2 Literature Background

2.1 The Palmgren-Miner's Linear Cumulative Damaging Rule

The cumulative fatigue damaging approach presented in this research was based on the Morrow's strain-life relationship. In a case of the loading being predominantly compressive, particularly for wholly compressive cycles, the Morrow's model provides more realistic live estimates. The mean stress correction effect seems to work reasonably well for steels. The model is mathematically defined as the following expression [3]:

$$\varepsilon_a = \frac{\sigma_f'}{E} \left(1 - \frac{\sigma_m}{\sigma_f'} \right) (2N_f)^b + \varepsilon_f' (2N_f)^c \quad (1)$$

where ε_a is the true strain amplitude, σ_f' is the fatigue strength coefficient, E is the material modulus of elasticity, σ_m is the mean stress, N_f is the numbers of cycle to failure for a particular stress range and mean, b is the fatigue strength exponent, ε_f' is the fatigue ductility coefficient, and c is the fatigue ductility exponent.

For strain - based fatigue live prediction, current industrial practice uses the Palmgren-Miner’s linear cumulative damaging rule normally associated with the established strain-life fatigue damaging models. The fatigue damage caused by each cycle of repeated loading is calculated by reference to material life curves, such as $S-N$ or $\epsilon-N$ curves. The fatigue damage D for one cycle and the total fatigue damage ΣD caused by cycles are expressed respectively as [2]:

$$D = \frac{1}{N_f} \quad (2)$$

$$\Sigma D = \Sigma \frac{N_i}{N_f} \quad (3)$$

where N_i is the numbers of cycle within a particular stress range and mean.

2.2 The Morlet Wavelet Coefficient

The Wavelet Transform (WT) approach is probably the most recent solution to overcome the nonstationary signals. This time-frequency technique is applied by cutting time domain signal into various frequency components through the compromise between time and frequency - based views of the signal. It presents information in both time and frequency domain in a more useful form [4-6].

The analysis is started with a basic function (called the mother wavelet) scaled and translated to represent the signal being analyzed [7]. The transform shifts a window along the signal and calculates the spectrum for every position. The process is repeated many times with a slightly shorter (or longer) window for every new cycle. In the end, the result will be a collection of time-frequency representations of the signal with different resolutions. The WT provides information on when and at what frequency the signal occurs [4].

The Morlet wavelet is one of functions which are the most generally used in the Continuous Wavelet Transform (CWT) analysis [8]. The name of the wavelet family is written “morl”. The wavelet decomposition calculates a resemblance index between signal being analyzed and the wavelet, called coefficient. It is a result of a regression of an original signal produced at different scales and different sections on the wavelet. It represents correlation between the wavelet and a section of the signal. If the index is large, the resemblance is strong, otherwise it is slight [9].

The WT of any time-varying signal $f(t)$ is defined as the sum of all of the signal time multiplied by a scaled and shifted version of the wavelet function $\psi(t)$ [10]. The CWT is expressed by the following integral:

$$CWT_{(a,b)} = \int_{-\infty}^{+\infty} f(t)\psi_{a,b}(t)dt \quad (4)$$

The parameter a represents the scale factor which is a reciprocal of frequency, the parameter b indicates the time shifting or translation factor, and t is time.

$\Psi_{a,b}(t)$ denotes the mother wavelet, i.e [11]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \quad a, b \in R; a \neq 0 \quad (5)$$

$$CWT_{(a,b)} = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt \quad (6)$$

In addition, the wavelet coefficient indicates how energy in the signal is distributed in the time-frequency plane [12]. The energy spectrum (the energy density over frequency) is plotted in order to observe the signal behaviour and its content gives significant information about the random signal pattern.

2.3 Data Clustering

Running fatigue damaging window, running global signal statistical window, analysis of the wavelet transform, the integrated kurtosis - based algorithm for Z-filter (I-kaz), and data correlation are required parameter for fatigue data classification. Fatigue data analysis technique based on signal classifying has to develop in detail to ensure higher amplitude signal can be determined and be extracted from original signal, and at the same time retain ability of fatigue damage. Product from this analysis will be used to develop an algorithm which can cluster and classify fatigue signal that gives the fatigue damage to automotive components. For this purpose, artificial intelligence concept can be used from the Morlet wavelet coefficient analysis for optimizing fatigue data analyses. With that, a mapping classification of fatigue history in signal can be generated.

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait - often proximity according to some defined distance measure. The purpose of clustering is to identify natural / intrinsic groupings of data from a large unlabeled data set to produce a concise representation of a system behaviour.

Fuzzy C-Means (FCM) is a data clustering technique wherein each data point belongs to a cluster to some degree that is specified by a membership grade. This technique was originally introduced by Jim Bezdek in 1981 as an improvement on earlier clustering methods. It provides a method that shows how to group data points that populate some multidimensional space into a

specific number of different clusters. This technique starts with an initial guess for the centers, which are intended to mark the mean location of each cluster. The initial guess for these cluster centers is most likely incorrect. Additionally, it assigns every data point a membership grade for each cluster. By iteratively updating the cluster centers and the membership grades for each data point, FCM iteratively moves the cluster centers to the right location within a data set. This iteration is based on minimizing an objective function that represents the distance from any given data point to a cluster center weighted by that data point membership grade [13].

3 Materials and Methods

The strain signal selected for the simulation purpose was from the database of Society of Automotive Engineers (SAE) profiles, named the SAESUS. The signal (in the unit of microstrain) was collected from a suspension component of a car and it was assumed to be sampled at

204.8 Hz for 25,061 data points. It gave the total record length of the signal of 122.4 seconds, as illustrated in Fig. 1.

For the calculation of the fatigue damage, the selected material for the simulation purpose was the SAE1045 carbon steel shaft. This material was chosen because it was commonly used in automotive industries for fabricate a vehicle lower suspension arm structure [14]. The material properties and their definitions are given in Table 1 [15].

For the purpose of this study, segmentation on the signal was done by implementing a fatigue feature extraction algorithm defined as an algorithm that inputs the signal and produces retained segments. As the algorithm was run, lower wavelet coefficient would gradually be removed until a stopping criterion is met. The criteria were set to $\pm 10\%$ difference of the fatigue damage. The segments were scattered and then clustered in order to develop a correlation between the fatigue damage and the Morlet wavelet coefficient. The flowchart is schematically illustrated in Fig. 2.

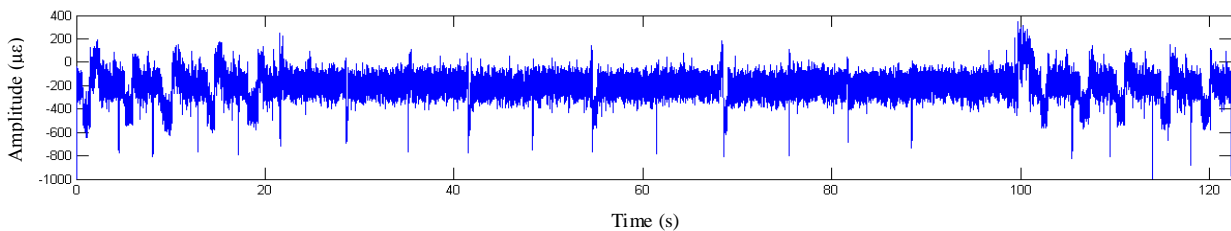


Fig. 1 Time histories of the SAESUS strain signal

Table 1 The mechanical properties of the SAE1045 carbon steel shaft

Properties	values
Ultimate tensile strength, S_u (MPa)	621
Modulus of elasticity, E (GPa)	204
Fatigue strength coefficient, σ'_f (MPa)	948
Fatigue strength exponent, b	-0.092
Fatigue ductility exponent, c	-0.445
Fatigue ductility coefficient, ϵ'_f	0.26

4 Results and Discussions

4.1 Segmentation process

It was found that solely using usual method to segment the fatigue time series data, in the end, it produced scatter plots that contain certain data that was anomalous to what was projected. These uncharacteristic data were mostly outliers that made it difficult for any observable pattern of data scattering to be identified. Therefore this scatter plot was deemed unsuitable and unreliable for further use in data classification and clustering [16].

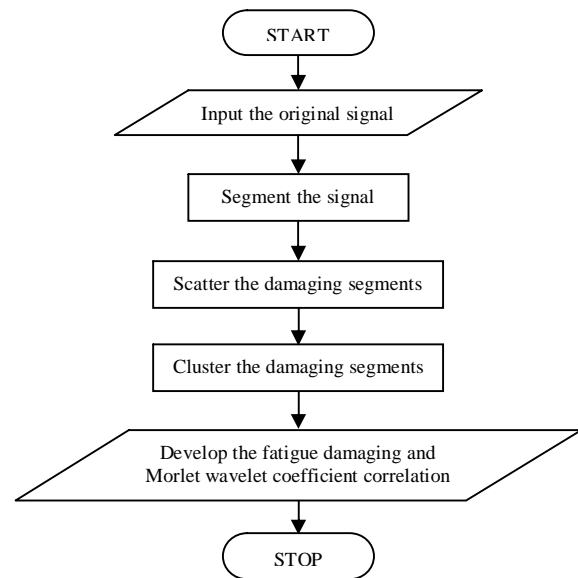


Fig. 2 Simplified flowchart of the study

By introducing the wavelet coefficient segmentation, the segmental wavelet coefficient analysis can be made more accurately since every segment contained only an overall peak. Therefore, it is more reasonable and practical to perform the wavelet coefficient analysis on data segments so that the wavelet coefficient measurement is a better representative of the segmental peakedness of the time series.

For the purpose of simplicity and criteria acceptability, this algorithm was used for segmenting the signal into 49 segments. These segments were not uniform in size; their lengths varied from one segment to another. This was because the algorithm segmented the wavelet coefficient time series so that each segment and

its corresponding linear representation would have the least amount of error. The segmentation was used to ensure that like features in the time series data would be isolated and be grouped into the same segments. Further analyses of each segment would help to determine which parts of the data gave significant contributions to the overall fatigue damage.

The segmentation process resulted 101 second edited signal. With respect to the fatigue damage, the shortened signal contained at least 90.8 % of the original fatigue damage (below than 10 %). It means that the algorithm preserved the originality of the fatigue damage and the signal behaviour. The Morlet wavelet coefficient and segmental data can be seen in Fig. 3.

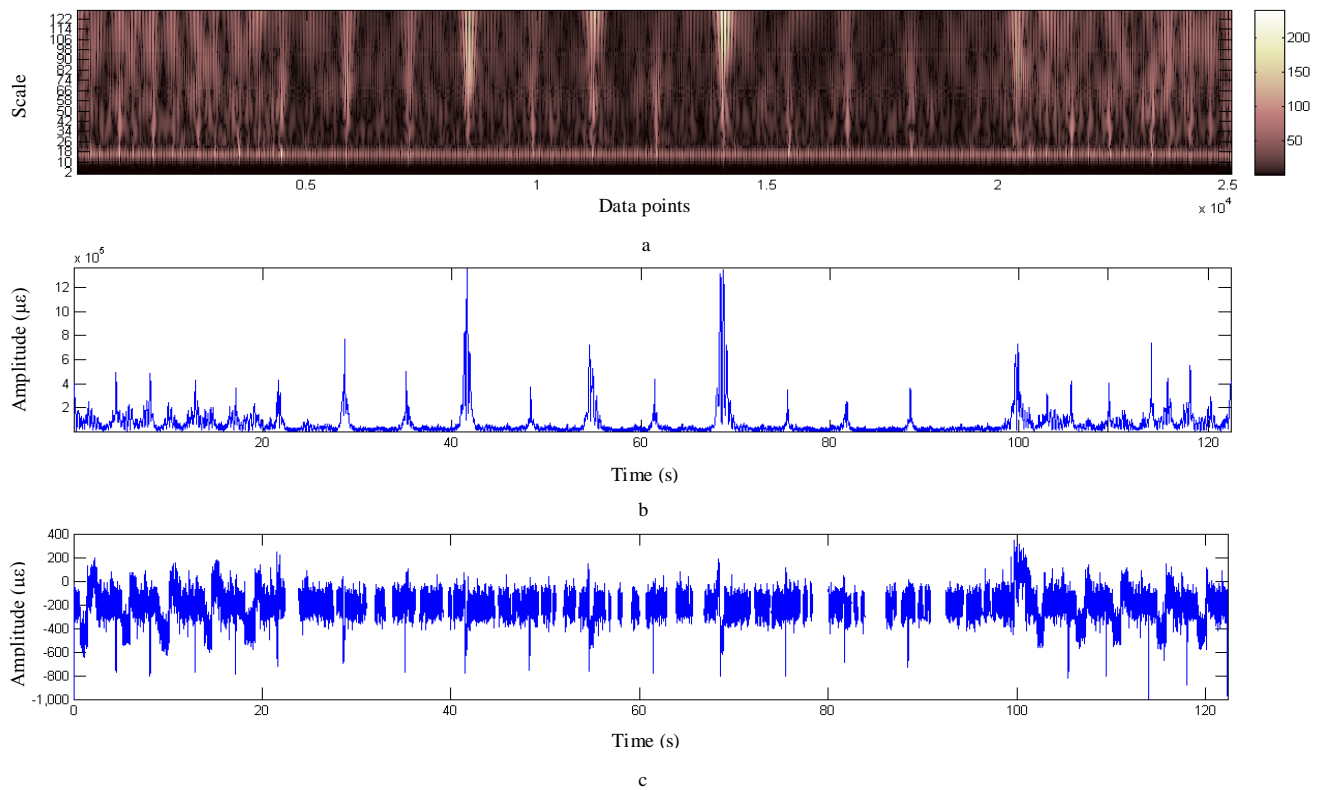


Fig. 3 The segmentation process: (a) the Morlet wavelet coefficient in time-frequency representation, (b) the Morlet wavelet coefficient in time representation, (c) 49 damaged segments

4.2 Clustering of the fatigue damage and the Morlet wavelet coefficient

The fatigue damage of these segments then was calculated using a specific commercial software package. The segments were also run through other software that calculated the Morlet wavelet coefficient of each segment. Another algorithm generated scatter plot of the fatigue damage against the wavelet coefficient values. Based on this scatter plot, patterns of data scattering, if any, were identified and noted.

Fig. 4 shows scatter plot of the fatigue damage over

the Morlet wavelet coefficient. In the figure, after including the wavelet coefficient segmentation, it is evident that no significant outliers were present to affect the overall look of the scatter. Thus, it can be seen certain patterns of scattering concentrated at the same areas of the scatter plot, although the points are widely distributed. The variations are mainly due to the randomness of the data and the variety in size of each segment.

In this case, it can be seen clearly that as a result of scatter pattern, where small fatigue damaging corresponds to small wavelet coefficient values and vice

versa. The higher wavelet coefficient points were only presented in the higher fatigue damaging range. Shorter segments with higher amplitude usually result in higher wavelet coefficient values, whereas longer segments

with lower amplitude would result in lower wavelet coefficient values. Therefore, for the higher fatigue damage, the wavelet coefficient values are theoretically higher.

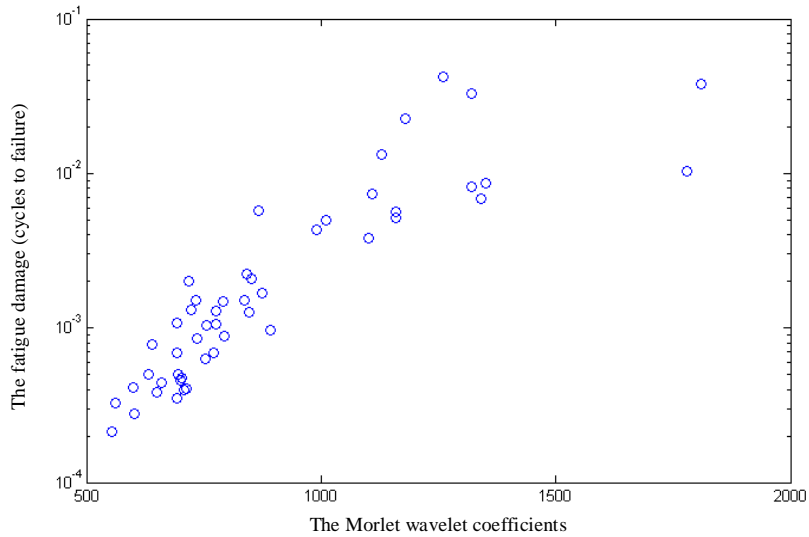


Fig. 4 Scatter plot of the fatigue damage and the Morlet wavelet coefficient

For further research work, scatter plot of the segmental fatigue damage versus the segmental wavelet coefficient as shown in Fig. 4 could be utilized for fatigue data classification and clustering. The utilization of the wavelet coefficient segmentation has resulted in the production of reliable data scatter for the described purpose. The clustering plot is shown in Fig. 5.

The clustering process divided the data in three clusters, they are lower damaging, damaging, and higher damaging segments. In the first cluster (lower

damaging segments), it can be seen that all the data points concentrated at 0.0003 - 0.008 cycles to failure. At the points, the wavelet coefficients were between 550 - 900. For the second cluster, the most of the data points seemed to be concentrated at 0.005 - 0.06. The wavelet coefficient values for this cluster were at range 1000 - 1400. The last cluster was higher damaging segments. This cluster only had two data points at the wavelet coefficient of 1800. The fatigue damaging values for these data points were 0.01 and 0.04 cycles to failures.

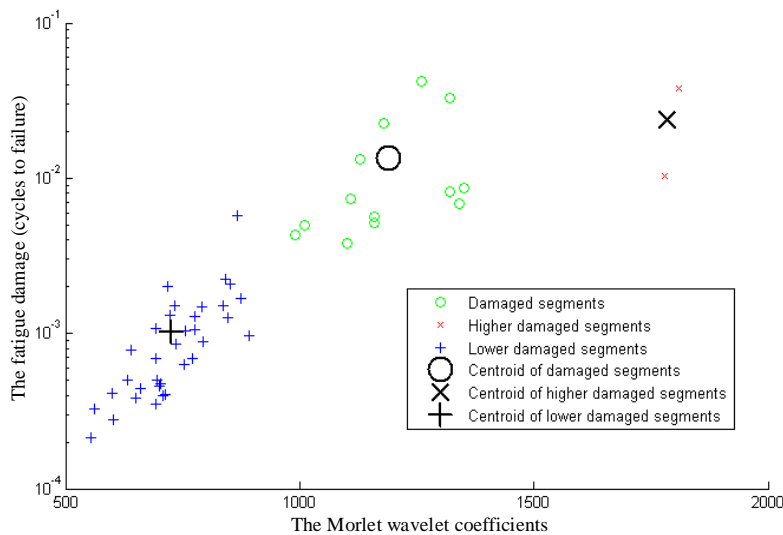


Fig. 5 Clustering of the fatigue damage and the Morlet wavelet coefficient

Furthermore, each cluster has centroid point. If these points in the plot were linked, it showed clearly correlation between the fatigue damage and the wavelet coefficient. At lower Morlet wavelet coefficient, the fatigue damage was decrease (not damaging). Whereas at higher Morlet wavelet coefficient, the fatigue damage was increase (more fatigue damaging). It indicated that the higher Morlet wavelet coefficient presented damaging segment, otherwise, it was undamaging segment. It means that the relationship between the Morlet wavelet coefficient and the fatigue damage was strong and parallel. Therefore it was expected that clustering plot of the fatigue damage versus the wavelet coefficient should reflect this trend in some manner.

As for the figure, since the segmentation algorithm had been incorporated with the wavelet coefficient, it can observe a certain trend of proportional relation between the fatigue damage and the Morlet wavelet coefficient. The pattern shows that the higher fatigue damage generally translates to the higher wavelet coefficient, as expected prior to the analysis. This means that this scatter plot truthfully reflects the hypothesized relationship between the fatigue damage and the wavelet coefficient.

5 Conclusion

The study has demonstrated the use of the Morlet wavelet coefficient segmentation for fatigue data clustering. As our main focus in this study, we suggested that the implementation of the Morlet wavelet coefficient segmentation algorithm would produce significantly reliable and accurate scatter plot for fatigue data clustering prospects. Finally, as a possible future work, after identifying and clustering the data in the signal, fatigue data editing through the elimination of certain noncontributory or insignificant segments of the signal may help in reducing the length and complexity of the data and may thus speed up the process of fatigue testing of metal components of mechanical systems or any similar application.

Acknowledgements

The authors would like to express their gratitude to Universitas Syiah Kuala and Universiti Kebangsaan Malaysia for supporting the research.

References:

- [1] Stephens, R. I., Dindinger, P. M., and Gunger, J. E., Fatigue damage editing for accelerated durability testing using strain range and SWT parameter criteria, *International Journal of Fatigue*, Vol. 19, 1997, pp. 599-606.
- [2] Abdullah, S., Wavelet Bump Extraction (WBE) for editing variable amplitude fatigue loadings, Ph.D, Thesis, The University of Sheffield, July 2005.
- [3] nCode, *The nCode book of fatigue theory*, Sheffield, UK: nCode International Ltd.
- [4] Valens, C., A really friendly guide to wavelets, <http://pagesperso-orange.fr/polvalens/clemens/wavelets/wavelets.html>, 1999.
- [5] Addison, P. S., *The illustrated wavelet transform handbook*, UK: Institute of Physics Publishing, 2002.
- [6] Percival, D. B. and Walden, A. T., *Wavelet methods for time series analysis*, UK: Cambridge University Press, 2000.
- [7] Berry, S., Practical wavelet signal processing for automated testing, IEEE 0-7803-5432-X/99, 1999, pp. 653-659.
- [8] Gao, J. H., Wu, R. S., and Wang, B. J., A new type of analyzing wavelet and its applications for extraction of instantaneous spectrum bandwidth, *SEG International Exposition and Annual Meeting*, San Antonio, Texas, 2001.
- [9] Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. M., *Matlab user's guide: wavelet toolbox™ 4*, MA, USA: the Math Works Inc., 2008.
- [10] Kim, B. S., Lee, S. H., Lee, M. G., Ni, J., Song, J. Y., and Lee, C. W., A comparative study on damage detection in speed-up and coast-down process of grinding spindle-typed rotor-bearing system, *Journal of Materials Processing Technology*, Vol. 187-188, 2007, pp. 30-36.
- [11] Purushotham, V., Narayanan, S., and Prasad, S. A. N., Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition, *NDT & E International*, Vol. 38, 2005, pp. 654-664.
- [12] Darpe, A. K., A novel way to detect transverse surface crack in a rotating shaft, *Journal of Sound and Vibration*, Vol. 305, 2007, pp. 151-171.
- [13] Matlab, *Matlab user's guide: fuzzy logic toolbox™ 2*, MA, USA: the Math Works Inc., 2008.
- [14] Khalil, M. and Topper, T. H., Prediction of crack-opening stress levels for 1045 as-received steel under service loading spectra, *International Journal of Fatigue*, Vol. 25, 2003, pp.149-157.
- [15] nCode, *ICE-flow: GlyphWorks 4.0 tutorials*, Sheffield, UK: nCode International Ltd., 2005.
- [16] Nopiah, Z. M., Khairir, M. I., Abdullah, S., Nizwan, C. K. E., and Baharin, M. N., Peak-valley segmentation algorithm for Kurtosis analysis and classification of fatigue time series data, *European Journal of Scientific Research*, ISSN 1450-216X Vol. 29 No. 1, 2009, pp. 113-125.