



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Development of Integrated Kurtosis-Based Algorithm for Z-Filter Technique

M.Z. Nuawi, M.J.M. Nor, N. Jamaludin, S. Abdullah, F. Lamin and C.K.E. Nizwan
Department of Mechanical and Materials Engineering,
Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia

Abstract: This study presents the development of an alternative statistical analysis method called Integrated Kurtosis-based Algorithm for Z-filter (I-kaz). The I-kaz analysis was developed in order to measure the degree of data scattering with respect to the data centroid for a dynamic signal analysis. The I-kaz parameter of a particular signal was determined by the value of I-kaz coefficient. The algorithm of I-kaz analysis provides a three dimensional graphic display of the magnitude distribution, which each axis represents the data distribution of the signal at three different frequency ranges. The I-kaz method which is a graphic-based technique is a proactive signal analysis tools because the relevant information can be extracted and the I-kaz space of scattering can be directly interpreted. The reliability of the I-kaz method has been evaluated by observing the I-kaz coefficient for a multivariate signal analysis and the trend of the coefficient values were compared to the existing statistical parameters, which are the standard deviation, variance, kurtosis and root mean square (rms). Finally, the applicability of the I-kaz method was accordingly verified.

Key words: Statistical analysis, digital signal processing, kurtosis, digital signal filtering, pattern recognition

DEVELOPMENT OF STATISTICAL SIGNAL ANALYSIS

Statistics is a mathematical science pertaining to the collection, analysis, interpretation or explanation and presentation of data. Statistical methods are applied in a wide variety of occupations and help people to identify, study and solve many complex problems. In the engineering applications, these methods enable decision makers to make informed and better decisions about uncertain situations. The statistical method plays a very important role in data analysis. The statistical data analysis focuses on the interpretation of the output to make inferences and predictions. Effective tools serve in two capacities: to summarize the data and to assist in interpretation.

The objectives of interpretive aids are to reveal the data at several levels of detail (Chatfield and Collins, 1980). A database may view as a domain that requires probes and tools to extract relevant information. As in the measurement process itself, appropriate instruments of reasoning must be applied to the data interpretation task. Statistical methods which summarise or describe a collection of data either numerically or graphically is called descriptive statistics while the inferential statistics model the pattern in the data in a way that accounts for randomness and uncertainty in the observations. The pattern recognition is important to draw inferences about

the process or population being studied. Pattern recognition aims to classify data based on either a prior knowledge or on statistical information extracted from the patterns (Hand *et al.*, 2001; Thuraisingham, 1998; Westphal and Blaxton, 1998).

Global signal statistics are frequently used to classify random signals. The most commonly used statistical parameters are the mean value, the standard deviation value, the root mean square (rms) value, the skewness and the kurtosis (Abdullah, 2005). For a signal with n-number of data points, the standard deviation value s, is given by:

$$s = \left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right)^{\frac{1}{2}} \quad (1)$$

where, x_i is the value of data point and μ is the mean of the data. The standard deviation value measures the spread of the data about the mean value. The square of the standard deviation value gives the variance, σ as in Eq. 2.

$$\sigma = s^2 \quad (2)$$

The r-th order of moment, M_r for the discrete signal in the frequency band can be written as:

$$\text{r-th order of moment, } M_r = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^r \quad (3)$$

where, n is the number of data and r is the order of moment. The root mean square (rms) value, which is the 2nd order of moment, is used to quantify the overall energy content of the signal. For discrete data sets the rms value is defined as in Eq. 4.

$$\text{r.m.s.} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (4)$$

Kurtosis K , which is the signal 4th statistical moment, is a global signal statistic which is highly sensitive to the spikiness of the data. For discrete data sets the kurtosis value is defined as in Eq. 5.

$$K = \frac{1}{n\sigma^4} \sum_{i=1}^n (x_i - \mu)^4 \quad (5)$$

For a Gaussian distribution the kurtosis value is approximately 3.0. Higher kurtosis values indicate the presence of more extreme values than should be found in a Gaussian distribution. Kurtosis is used in engineering for detection of fault symptoms because of its sensitivity to high amplitude events (Abdullah, 2005).

Various ways of analysing the statistical properties of signals have been applied, but in condition monitoring application it is prior to the flexibility of the chosen method. Signal classifying on the basis of mean and variance was not compatible because the real-life signals often contain outliers that can bring a noticeable shift in the actual value of both mean and variance (Pontuale *et al.*, 2003). The statistical features described above reflect the pattern characteristic of the time domain signal that only contains the typical frequency band (He *et al.*, 2007). However, these features are insufficient as the machine signals usually indicate non-stationary. Research has shown that some typical frequency bands could reflect the development of faults when analyzing the behavior of measured signals (Zhu *et al.*, 2005). Therefore, the spectral distribution in frequency domain of the machine signals can also reflect the symptoms that are sensitive to fault severity (He *et al.*, 2007). Such health monitoring is often complex, intricate and a highly demanding task. Thus, the modeling technique should possess certain required qualities such as adaptive, accurate, lucid and generalized, flexible, has noise tolerant and extrapolative to be applicable in monitoring applications (Kothamasu *et al.*, 2005).

In view of this, it is useful to develop an alternative statistical approach as a supplement to the existing statistical method. This study presents an alternative statistical analysis method called Integrated kurtosis-based algorithm for Z-filter (I-kaz) technique and its

viability in forming an effective condition monitoring system. Both descriptive and inferential statistics were utilised in the I-kaz method. The numerical descriptor of the I-kaz method is the I-kaz coefficient, Z^∞ and the value of the coefficient is synchronized by the three dimensional graphical summarizations of frequency distribution. The I-kaz display is used to model patterns in the data, accounting for randomness and drawing inferences about the larger population, which is classified as the inferential statistics. These inferences are very useful for estimation and forecasting of future observations.

The I-kaz method can be used to identify key variables or groups of variables that control the system under study. The resulting graphical representation permits the dimension reduction so that significant relationships among observations or samples can be identified. A common goal for a statistical research project is to investigate causality and in particular to draw a conclusion on the effect of changes in the values of predictors or independent variables on response or dependent variables. This can be achieved by applying the I-kaz method.

When a function is evaluated by numerical procedures, it is always necessary to sample the function in some manner, because digital computers cannot deal with analogue, continuous functions except by sampling the signal. Examples of where sampled time domain data is used in engineering include simple and complex vibration analysis of machinery, as well as measurements of other variables such as boiler pressures, temperatures, flow rates and turbine speeds and many other machine parameters. It is also used in areas where computers and microcontrollers are used to automate processes and react to input data from the processes.

If the signal has significant variation, then the sampling period must be small enough to provide an accurate approximation of the signal. Significant signal variation usually implies that high frequency components are present in the signal. It could therefore be inferred that the higher the frequency of the components present in the signal, the higher the sampling rate should be. If the sampling rate is not high enough to sample the signal correctly then a phenomenon called aliasing occurs.

The term aliasing refers to the distortion that occurs when a continuous time signal has frequencies larger than half of the sampling rate (Figliola and Beasley, 2000). The process of aliasing describes the phenomenon in which components of the signal at high frequencies are mistaken for components at lower frequencies. In statistics, signal processing, computer graphics and related disciplines, aliasing refers to an effect that causes different

continuous signals to become indistinguishable when sampled. It also refers to the distortion or artifact, which results when a signal is sampled and reconstructed as an alias of the original signal. The Nyquist Sampling Theorem states that to avoid aliasing occurring in the sampling of a signal the sampling rate should be greater than or equal to twice the highest frequency present in the signal. This is referred to the Nyquist sampling rate. The next sections explain on the development of the I-kaz method by considering the aliasing effect.

DEVELOPMENT OF THE INTEGRATED KURTOSIS-BASED ALGORITHM FOR Z-FILTER (I-KAZ)

Typical application areas for dynamic models are control, prediction, planning, fault detection and diagnosis. The aim is a better understanding by visualization in 3 dimensional spaces and to generalize the ideas for higher dimensions. Thus, the I-kaz method was developed based on the concept of data scattering about its centroid. The sampling frequency of the raw signal is using 2.56 Nyquist number. The number was selected because most researchers in signal processing are comfortable with the number. Thus, to avoid aliasing effect, the maximum frequency span will be as follows:

$$f_{max} = \frac{fs}{2.56} \tag{6}$$

Consider a dynamic signal as illustrated in Fig. 1. The time domain signal is decomposes into three frequency ranges, which are:

- x-axis: Low frequency (LF) range of $0-0.25 f_{max}$
- y-axis: High frequency (HF) range of $0.25 f_{max}-0.5 f_{max}$
- z-axis: Very high frequency (VF) range of $0.5 f_{max}$

The selection of the $0.25 f_{max}$ and the $0.5 f_{max}$ as the low and high frequency limit respectively were done by considering the 2nd order of the Daubechies concept in signal decomposition process (Daubechies, 1992).

Based on kurtosis, the I-kaz method provides a three dimensional graphical representation of the measured signal frequency distribution. In order to measure the scattering of a data distribution, variance σ^2 for each frequency band which are σ_L^2 , σ_H^2 and σ_V^2 is calculated as in Eq. 7. The variance obtains the average magnitude deviation of the instantaneous points with respect to the mean value as illustrated in Fig. 2.

$$\sigma_L^2 = \frac{\sum_{i=1}^N (x_i^L - \mu_L)^2}{n}; \sigma_H^2 = \frac{\sum_{i=1}^N (x_i^H - \mu_H)^2}{n}; \sigma_V^2 = \frac{\sum_{i=1}^N (x_i^V - \mu_V)^2}{n} \tag{7}$$

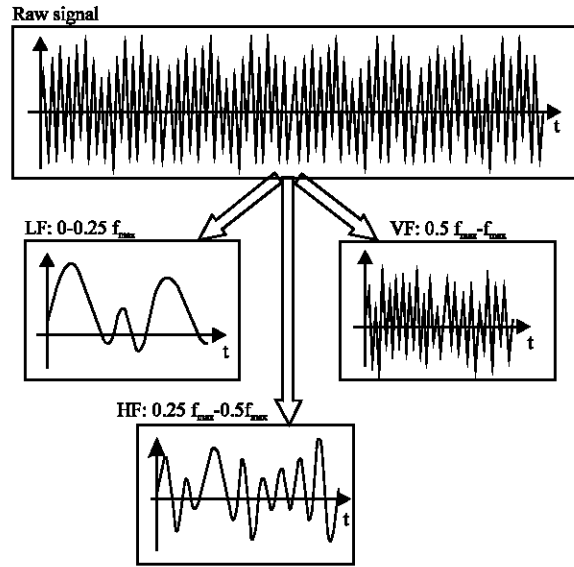


Fig. 1: The decomposition process of a signal in I-kaz procedure

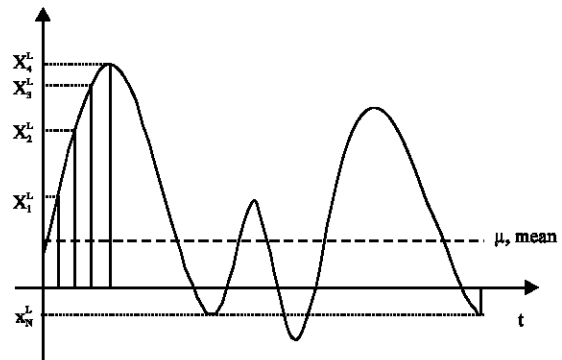


Fig. 2: An example of low frequency signal showing instantaneous points for I-kaz calculation

Since the I-kaz method was developed based on the concept of data scattering about its centroid, the I-kaz coefficient, Z^∞ can be written as in Eq. 8.

$$Z^\infty = \sqrt{\frac{\sum_{i=1}^N (x_i^L - \mu_L)^4}{n^2} + \frac{\sum_{i=1}^N (x_i^H - \mu_H)^4}{n^2} + \frac{\sum_{i=1}^N (x_i^V - \mu_V)^4}{n^2}} \tag{8}$$

where, x_i , y_i , z_i are the value of discrete data in LF, HF and VF range, respectively at the i -sample of time, μ_L , μ_H and μ_V are mean of each frequency band and N is the number of data. The I-kaz coefficient, Z^∞ can be simplified in terms of variance, σ as in Eq. 9.

$$\text{I-kaz coefficient, } Z_\sigma = \sqrt{(\sigma_L^2)^2 + (\sigma_H^2)^2 + (\sigma_V^2)^2} \tag{9}$$

From the Eq. 5, the 4th moment can then be derived based on the kurtosis, K and standard deviation, s parameters such as in Eq. 10.

$$Ks^4 = \sum_{i=1}^n (x_i - \bar{x})^4 \tag{10}$$

$$Ks^4 = M_4 \times n$$

$$M_4 = \frac{Ks^4}{n}$$

where, M_4 is the 4-th order of moment, K is the kurtosis, s is the standard deviation and n is the number of data. The I-kaz coefficient can be derived in terms of 4th order of moment, M_4 as shown in Eq. 11.

$$Z^\infty = \sqrt{\frac{\sum_{i=1}^N (x_i^L - \mu_L)^4}{n^2} + \frac{\sum_{i=1}^N (x_i^H - \mu_H)^4}{n^2} + \frac{\sum_{i=1}^N (x_i^V - \mu_V)^4}{n^2}}$$

$$Z^\infty = \sqrt{\frac{1}{n}(M_4^L) + \frac{1}{n}(M_4^H) + \frac{1}{n}(M_4^V)} \tag{11}$$

where, n is the number of data and M_4^L , M_4^H and M_4^V are the 4th order moment of signal in LF, HF and VF range respectively. From the Eq. 10 and 11, the I-kaz coefficient can then be written in terms of kurtosis, K and standard deviation, s as in Eq. 12:

$$Z_k = \sqrt{\frac{K_L s_L^4}{n^2} + \frac{K_H s_H^4}{n^2} + \frac{K_V s_V^4}{n^2}}$$

$$Z_k = \frac{1}{n} \sqrt{K_L s_L^4 + K_H s_H^4 + K_V s_V^4} \tag{12}$$

where, n is the number of data, K_L , K_H , K_V are the kurtosis of signal in LF, HF and VF range and s_L , s_H and s_V are the standard deviation of signal in LF, HF and VF range, respectively. The I-kaz algorithm was summarized as presented in Fig. 3.

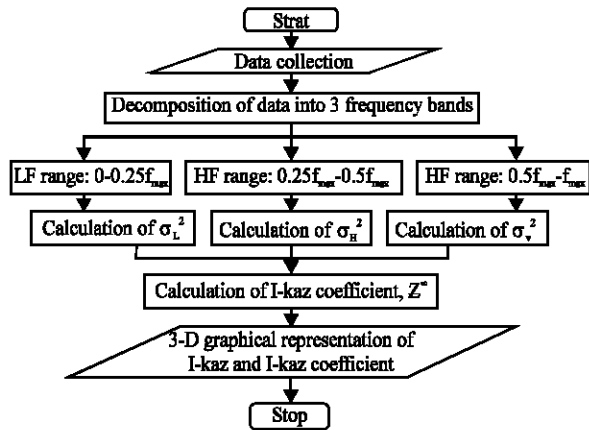


Fig. 3: Flowchart of the I-kaz method

RESULTS OF I-KAZ SIMULATION

Case study I: Signal with different amplitude at a constant frequency: Two test signals, named as A1 and A2 (Fig. 4) were defined with 512 data points and sampled at 1000 Hz. Both signals A1 and A2 were consist of a combination of 50 Hz and 170 Hz sinusoidal and it was intentionally defined to be different in the amplitude of 0.1 and 0.5, respectively. The logic of creating the A1 and A2 with the different amplitude while keeping the frequency constant (Fig. 5) were to observe the effect of amplitude

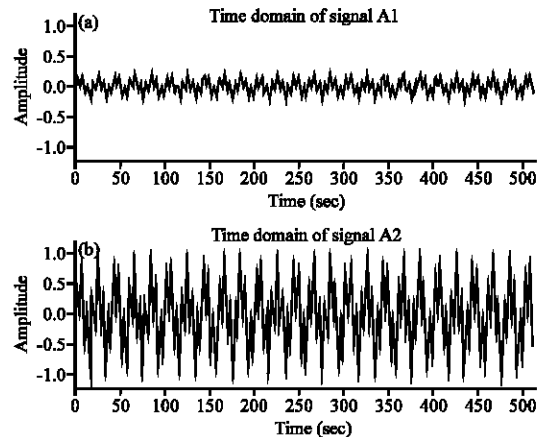


Fig. 4: The time domain representation of test signals: (a) A1, (b) A2

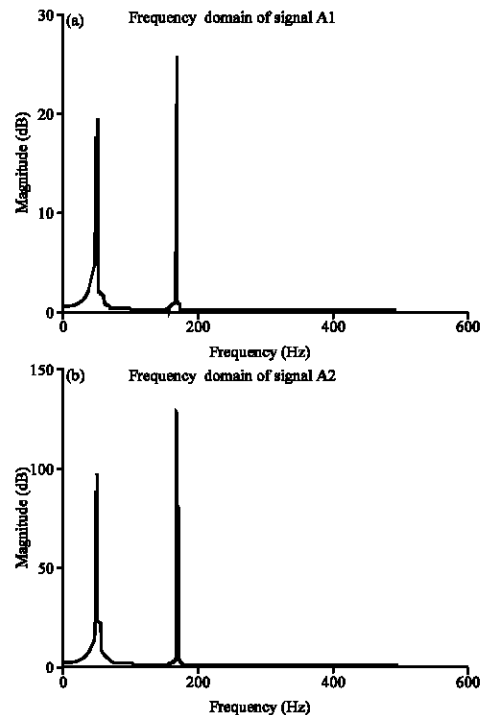


Fig. 5: The frequency domain representation of test signals: (a) A1, (b) A2

Table 1: Statistical parameter for signal A1 and A2

| Statistical parameter | Signal A1 | Signal A2 | Deviation(%) |
|-----------------------------|-----------|-----------|--------------|
| I-kaz coefficient (Z^*) | 1.78E-05 | 4.24E-04 | 95.8 |
| Standard deviation (s) | 0.1001 | 0.5005 | 80.0 |
| Variance (σ) | 0.0100 | 0.2505 | 96.0 |
| Kurtosis (K) | 2.2406 | 2.2406 | 0.0 |
| rms | 0.1000 | 0.5000 | 80.0 |

changes on the I-kaz representation synchronising with the I-kaz coefficient.

Using Eq. 5, the maximum frequency span, f_{max} of the test signal is 390.6 Hz while the low and high frequency limit for the signal are 97.7 Hz ($0.25 f_{max}$) and 195.3 Hz ($0.5 f_{max}$), respectively. Therefore, the frequency range of each axis of the I-kaz representation was summarized as follows:

- x-axis: 0-97.7 Hz-low frequency (LF) range
- y-axis: 97.7-195.3 Hz-high frequency (HF) range
- z-axis: 195.3-390.6 Hz-very high frequency (VF) range

The characteristic of each signal was determined by statistical analysis i.e. standard deviation, variance, kurtosis, rms and I-kaz coefficient. Using the Eq. 1-2, 4-5 and 8 the statistical analysis results for both test signal are shown in Table 1.

The value of I-kaz coefficient for A1 signal (1.78E-05) is lower than the A2 signal (4.24E-4) as the A1 signal was consist of lower amplitude signal compared to the A2 signal. This trend of the result was parallel with the trend of results obtained by using standard deviation, variance and rms Nevertheless, the Z^* which has 95.8% deviation was more efficient compared to the standard deviation and rms, which both parameters have 80% deviation. In contrast, the value of kurtosis for A1 was equal to the A2 signal because even both signal obviously have the different amplitude, the trend of signal's magnitude with respect to its mean was similar for both signals. Since the kurtosis method was unable to show the differences between A1 and A2 signals, thus it was unable to be as a tool to detect the amplitude changes.

The I-kaz representation for A1 and A2 signal was shown in Fig. 6. In spite of using standard deviation, variance and rms method, the I-kaz method provides a better illustration in characterising a signal since it generates a 3-D graphical representation besides of it numerical indicator, I-kaz coefficient Z^* . The higher value of Z^* refers to the bigger space of scattering of the I-kaz representation.

Case study II: Signal with different frequency at a constant amplitude: In order to observed the ability of I-kaz method in classifying signals, which have differences in frequency, two test signals named as B1

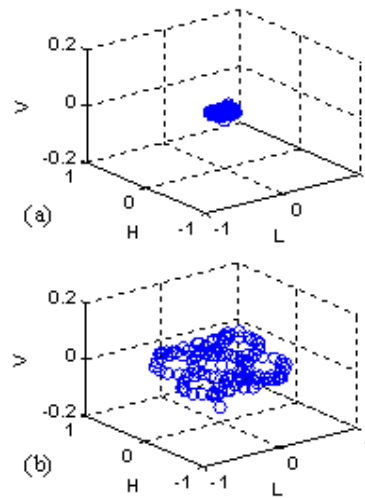


Fig. 6: The I-kaz representation of test signals: (a) A1, (b) A2

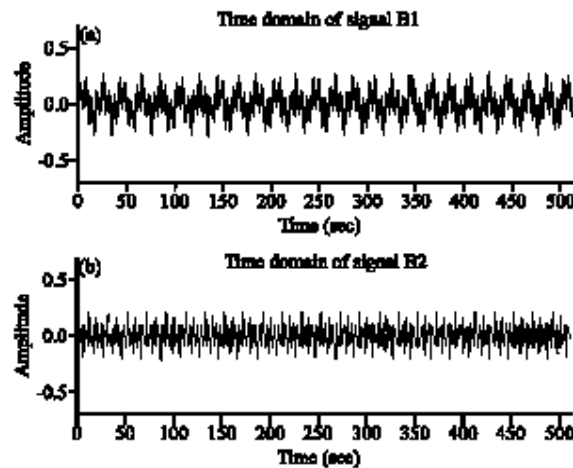


Fig. 7: (a) The time domain representation of test signals: (a) B1, (b) B2

and B2 (Fig. 7) were defined with 512 data points and sampled at 1000 Hz. Signal B1 was consist of a combination of 50 and 170 Hz sinusoidal while the signal B2 was consist of 250 and 400 Hz of frequency combination (Fig. 8). The amplitude of the B2 signal was purposely defined to be equal to the amplitude of B1 signal. The characteristic of each signal was accordingly determined by the previous procedure of statistical analysis for amplitude-based signal (case study I) by using the Eq. 1-2, 4-5 and 8. Table 2 represents the statistical parameters obtained for both B1 and B2 signals.

Since the sampling frequency was 1000 Hz, using Eq. 5, the maximum frequency span, f_{max} of the test signal is 390.6 Hz while the low and high frequency limit for the signal are 97.7 Hz ($0.25 f_{max}$) and 195.3 Hz ($0.5 f_{max}$),

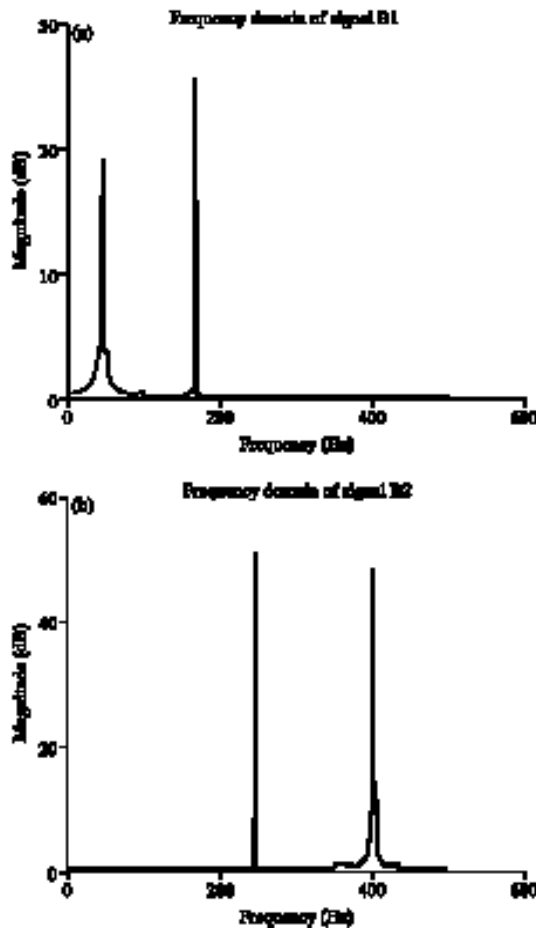


Fig. 8: (a) The frequency domain representation of test signals: (a) B1, (b) B2

Table 2: Statistical parameter for signal B1 and B2

| Statistical parameter | Signal B1 | Signal B2 | Deviation (%) |
|---------------------------------------|-----------|-----------|---------------|
| I-kaz coefficient (\mathcal{Z}^*) | 1.78E-05 | 2.29E-05 | 28.7 |
| Standard deviation (s) | 0.1001 | 0.1000 | 0.0 |
| Variance (σ) | 0.0100 | 0.0100 | 0.0 |
| Kurtosis (K) | 2.2404 | 2.3828 | 4.3 |
| rms | 0.1000 | 0.1000 | 0.0 |

respectively. Therefore, the frequency range of each axis of I-kaz representation was summarised as follows:

- x-axis: 0-97.7 Hz-low frequency (LF) range
- y-axis: 97.7-195.3 Hz-high frequency (HF) range
- z-axis: 195.3-390.6 Hz-very high frequency (VF) range

The statistical analysis result obtained for the frequency-based signal (case study II) was in contradiction with the amplitude-based signal (case study I), where except for kurtosis, the standard deviation, variance and rms methods resulting the similar value. Meanwhile, the I-kaz coefficient still shows its ability to

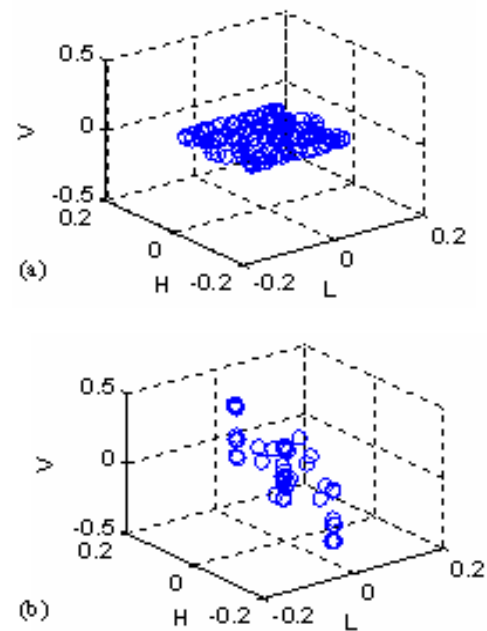


Fig. 9: (a) The I-kaz representation of test signals: (a) B1, (b) B2

differentiate the frequency difference between signals. The I-kaz coefficient, \mathcal{Z}^* obtained for B2 was 2.29E-05 was comparatively higher than B1 (1.78 E-05). The higher value of I-kaz coefficient, \mathcal{Z}^* explains that the frequencies of the B2 signal (250 and 400 Hz) were higher than the B1 signal (50 and 170 Hz). The \mathcal{Z}^* gives 28.7% deviation which has raised it to be more efficient compared to the kurtosis that only gives 6.3% deviation. In addition, the I-kaz coefficient value was supported by the I-kaz representation (Fig. 9), where the higher value of the I-kaz coefficient was illustrated by the bigger space of scattering of the I-kaz display.

CONCLUSION

This study discussed on the development of an alternative statistical analysis method, known as integrated kurtosis-based algorithm, I-kaz. The I-kaz method was adaptive in general and detects very well any changes of the measured signal. Unlike the existing statistical analysis; standard deviation, variance and kurtosis, I-kaz method was capable to indicate both amplitude and frequency difference by simultaneously obtaining the I-kaz representation and the I-kaz coefficient, \mathcal{Z}^* .

For the first case study, the \mathcal{Z}^* obtained was from 1.78E-05 to 4.24E-04 for the signal with an increased in the amplitude value. It could therefore be inferred that the

higher the amplitude present in the signal, the higher value of the I-kaz coefficient will be obtained. The condition of this increasing amplitude can be simultaneously monitored using the I-kaz representation. By observing the space of scattering of the frequency distribution, the bigger space of scattering illustrates that the amplitude of the signal is comparatively going higher.

In addition, for the second case study the Z° obtained for the signal with an increased in frequency value was from 1.78E-05 to 2.29E-05. To conclude, the higher value of the Z° indicates the higher frequency of signal. Again, the condition of the increasing in frequency can be simultaneously monitored using the I-kaz representation. By observing the space of scattering for the frequency distribution, the bigger space of scattering illustrates that the frequency of the signal is comparatively going higher.

For the first case study the analysis using variance gave 96% of deviation, which was higher compared to 95.8% deviation of the I-kaz method. Nonetheless, the variance parameter was unable to detect both amplitude and frequency changes. Thus, the I-kaz method was reliable especially for monitoring purpose where the observation on the changes of the signal amplitude and frequency were commonly required.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to University Kebangsaan Malaysia and Ministry of Science, Technology and Innovation, through the fund of 03-01-02-SF0303, for supporting this research.

REFERENCES

Abdullah, S., 2005. Wavelet bump extraction (Wbe) for editing variable amplitude fatigue loading. Ph.D Thesis, University of Sheffield, United Kingdom.

- Chatfield, C. and A.J. Collins, 1980. An Introduction to Multivariate Analysis. Chapman and Hall, London.
- Daubechies, I., 1992. Ten Lectures On Wavelets. SIAM, Philadelphia.
- Figliola, R.S. and D.E. Beasley, 2000. Theory and Design for Mechanical Measurements. 3rd Edn. John Wiley and Sons. Hoboken, USA.
- Hand, D., H. Mannila and P. Smyth, 2001. Principles of Data Mining. MIT Press, Cambridge, MA.
- He, Q., F. Kong and R. Yan, 2007. Subspace-based gearbox condition monitoring by kernel principal component analysis. Mech. Syst. Sig. Proc., 21 (4): 1755-1772.
- Kothamasu, R., S.H. Huang and W.H. Verduin, 2005. Comparison of computational intelligence and statistical methods in condition monitoring for hard turning. Int. J. Prod. Res., 43 (3): 597-610.
- Pontuale, G., F.A. Farelly, A. Petri and L. Pitolli, 2003. A statistical analysis of acoustic emission signals for tool condition monitoring (TCM). ARLO-Acoustics Rev. Lett. On-line, 4 (1): 13-18.
- Thuraisingham, B., 1998. Data Mining: Technologies, Techniques, Tools and Trends. CRC Press, Inc., Boca Raton, FL., USA.
- Westphal, Ch. and T. Blaxton, 1998. Data Mining Solutions: Methods and Tools for Solving Real-World Problems. John Wiley and Sons.
- Zhu, Z.K., Z.H. Feng and F.R. Kong, 2005. Cyclostationarity analysis for gearbox condition monitoring: Approaches and effectiveness. Mech. Syst. Sig. Proc., 19: 467-482.