Research Journal of Applied Sciences 2 (5): 600-605, 2007
© Medwell Journals, 2007

Application of Moving Average Approaches in Analysing Fatigue Road Loading

S. Abdullah, M.D. Ibrahim, A. Zaharim and I. Ahmad
Faculty of Engineering, Universiti Kebangsaan Malaysia,
43600 UKM Bangi, Selangor, Malaysia

Abstract: This study describes the analysis of fatigue road loading using the Moving Average approaches. Thus, techniques for preserving data associated to the underlying probabilistic properties were introduced. Fatigue damage cumulating is a random variable in essence. The randomness comes from the loading process and fatigue resistance of material. Seldom have models dealing with fatigue damage cumulating considered the co-influence of the two aspects of randomness at the same time. This article has established a probabilistic distribution model of moving average fatigue damage. In the model, the moving average trend can be estimated by smoothing the data to reduce the random variation and randomness at fatigue resistance of material is described by introducing a random variable of a variable amplitude loading sampled at 200Hz. This model can calculate the fatigue damage cumulating distribution after moving.

Key words: Fatigue, moving average, random data, time histories, variable amplitude loading

INTRODUCTION

A signal is a series of numbers that come from measurement, typically obtained using some recording method as a function of time (Meyer, 1993). In the study of fatigue analysis, the signal consists of a measurement of the cyclic loads, i.e., force, strain and stress against time. A time series typically consists of a set of observations of a variable taken at equally spaced intervals of time (Harvey, 1981). Today, most experimental measurements, or data samples, are performed digitally. And it is also known as a discrete time series, which is formed as a function of time. The objective of time series analysis is to determine the statistical characteristics of the original function by manipulating the series of discrete numbers (New Land, 1993).

Many data mining applications deal with privacy-sensitive data. The best means of obtaining unpredictable random numbers is by measuring physical phenomena such as fatigue damage, radioactive decay, thermal noise in semiconductors and even digitized images of a lava lamp. However few computers (or users) have access to the kind of specialized hardware required for these sources and must rely on other means of obtaining random data (Yan *et al.*, 2001).

The objective of this study is to observe a technique for preserving data by randomly perturbing the data associated to the underlying probabilistic properties. This has fostered the development of a class of data algorithms (Agrawal and Aggawal, 2001) that try to extract the data pattern without directly accessing the original data and guarantees that the process. This approach tries to preserve data privacy from random noise (Estivill and Brankovic, 1999). Typically, these data are the used with curve-fitting techniques to develop the average fatigue behavior of the material over an appropriate range of stress levels.

LITERATURE BACKGROUND

Many signals in nature exhibit random or nondeterministic characteristics which provide a challenge to analysis using signal processing techniques (Natrella, 1966). A signal representing a random physical phenomenon cannot be described in a point by point manner by means of a deterministic mathematical equation. A signal representing a random phenomenon can be characterised as either stationary or nonstationary.

A stationary signal is characterised by values of the global signal statistical parameters, such as the mean, variance and root-mean-square, which are unchanged across the signal length. Stationary random processes can further be categorised as being ergodic or nonergodic. If the random process is stationary and the mean value and the autocorrelation function do not differ when computed over different sample segments measured for the process, the random process is defined as ergodic. In the study of nonstationary signals the

Corresponding Author: S. Abdullah, Faculty of Engineering, Universiti Kebangsaan Malaysia,
43600 UKM Bangi, Selangor, Malaysia

global signal statistical values are dependent on the time of measurement (Tacer and Loughlin, 1998). Nonstationary signals can be dividedinto 2 categories: Mildly nonstationary and heavily non stationary.

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 (r.m.s.) value, the skewness, the kurtosis and the crest factor (Bendat and Piersol, 1986).

For a signal with a number *n* of data points, the mean value and is given by

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{1}$$

The Standard Deviation (SD) is mathematically defined as

$$SD = \left\{\frac{1}{n}\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2\right\}^{1/2} \tag{2}$$

for the samples more than 30 (Bendat and Piersol, 1986). The standard deviation value measures the spread of the data about the mean value.

The r.m.s. value, which is the 2nd statistical moment, is used to quantify the overall energy content of the signal. For a zero-mean signal the r.m.s. value is equal to the SD value. For discrete data sets the r.m.s. value is defined as

$$r.m.s. = \left\{\frac{1}{n}\sum_{j=1}^{n} x_j^2\right\}^{1/2} \tag{3}$$

The skewness, which is the signal 3rd statistical moment, is a measure of the symmetry of the distribution of the data points about the mean value. The skewness for a symmetrical distribution such as a sinusoid or a Gaussian random signal is zero. Negative skewness values indicate probability distributions that are skewed to the left, while a positive skewness values indicate probability distributions that are skewed to the right, with respect to the mean value. The skewness of a signal is given by

$$S = \frac{1}{n(SD)^3}\sum_{j=1}^{n}(x_j - \bar{x})^3 \tag{4}$$

Kurtosis, 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

$$K = \frac{1}{n(r.m.s.)^4}\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^4 \tag{5}$$

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 (Hinton, 1995).

The crest factor, which is commonly encountered in engineering applications, is defined as the ratio between the maximum value in the time history and the r.m.s. value:

$$CF = \left|\frac{x_{jmax}}{r.m.s.}\right| \tag{6}$$

The crest factor value for sinusoidal time histories is 1.41 and the value approaches 4.00 in the case of a Gaussian random signal of infinite length.

Since nonstationary data exhibits the random pattern, the moving average is introduced as one of the approaches used to smoothing the time series data. It can be estimated by smoothing the data to reduce the random variation (Qu and He, 1986). A range of smoothers is available, but it begin with the simplest and oldest smoother, a moving average. Moving averages have several methods such as the simple moving averages, the double moving averages and the weighted moving averages (Makridakis *et al.*, 1998). Moving average model is also a dependence relationship to set up among the successive error terms. The model which use past errors as explanatory variable:

$$Y_t = \theta_o + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots\dots\dots + \theta_q e_{t-q} + e_t \tag{7}$$

Where Y_t is the actual value of the data, θ_o is the constant, θ_1, θ_2,…., θ_{t-q} are the linear regression coefficients and e_t is the error coefficient.

The idea of applying the simple moving averages concept is related to the data observations which are likely to be closed in value. By taking an average of the points near an observation, it provides a reasonable estimation of the data. Thus, it eliminates the randomness of the data and producing a smooth trend with respect to the original nonstationary data pattern (Holt *et al.*, 1960).

The simple moving average method required an odd number of observations to be included in each average (Kendall *et al.*, 1983). The purpose of this requirement is to ensure the average (T_t) was centered at the middle of the data values being averaged, which is shown in the following expression, i.e.,

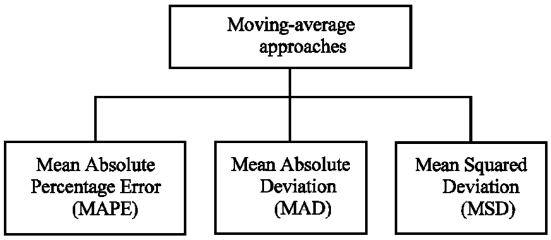

Fig. 1: Criteria to determine the best moving average model

$$T_t = \frac{1}{k}\sum_{j=-m}^{m} Y_{t+j} \tag{8}$$

Where k is an odd integer and m = (k -1)/2. The double moving averages can itself be smoothed by another moving average. In fact, any combination of moving average can be used together to form a double moving average.

The weighted moving averages is a special case of the weights which are set to 1/k where a weighted k-point moving average is written as

$$T_t = \sum_{j=-m}^{m} a_j Y_{t+j} \tag{9}$$

Where a_j is the weight factor of the moving average method.

MATERIALS AND METHODS

The moving average method had been applied to the fatigue data in order to determine the best model based on three criteria and it is illustrated in a diagram of Fig. 1. From this figure, Mean Absolute Percentage Error (MAPE) is the mean or average of the sum of all of the percentage errors for a given data set taken without regard to sign. It is one measure of accuracy commonly used in quantitative methods (Qu and He, 1986).

PE_t is the relative or percentage error is the discrepancy between an exact value and some approximation that need to be applied in the analysis. PE_t for MAPE is mathematically defined as following:

$$PE_t = \left(\frac{Y_t - F_t}{Y_t}\right) \times 100 \tag{10}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} |PE_t| \tag{11}$$

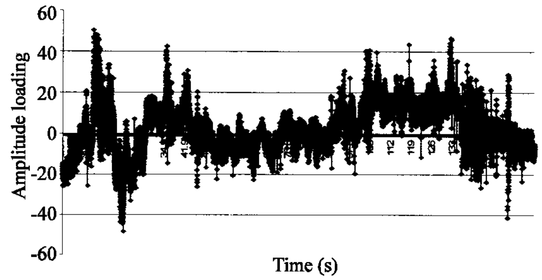

Fig. 2: A variable amplitude fatigue strain time histories

The Mean Absolute Deviation (MAD) is the average absolute deviation from the mean and is a common measure of forecast error in analysis. The term mean deviation is used as a synonym for mean absolute deviation, to be precise it is not the same; in its strict interpretation (namely, omitting the absolute value operation), the mean deviation of any data from its mean always zero (Qu and He, 1986; Makridakis *et al*., 1998; Holt *et al*., 1960; Kendall *et al*., 1960). The MAD mathematical expression is shown as following:

$$MAD = \frac{1}{n}\sum \left|Y_i - \overline{Y}\right| \tag{12}$$

The Mean Squared Deviation (MSD) is the variance measure, which is defined as the sum of squared deviations divided by one less than the total number of observation (Qu and He, 1986).

$$MSD = \frac{1}{n}\sum \left(Y_i - \overline{Y}\right)^2 \tag{13}$$

Autocorrelation Function (ACF) is used to identify the seasonality which is present in the given specific situations and to determine if data are stationary. In addition, the ACF is also being used to recognise appropriate models for nonstationarity of the random data. Partial Autocorrelation Function (PACF) is also used in order to identify the extent of relationship between current values of variable with earlier of that same variable while holding the effects of all other constant. Thus, it is completely analogous to partial correlation but refers to a single variable.

RESULTS AND DISCUSSION

The data are which are plotted in Fig. 2 was measured on the front left lower suspension arm of an automobile which was traveling on public road surface (mixture of smooth and irregular asphalt). It is

Table 1: Global statistical parameter obtained for the analysed time series

Descriptive		Statistical values
Mean [με]		2.337
Standard deviation [με]		25.466
Root-mean-square [με]		0.104
Minimum [με]		-138.72
Maximum [με]		167.13
Skewness	Statistics	1.869
	Standard error	0.010
Kurtosis	Statistics	13.745
	Standard error	0.020

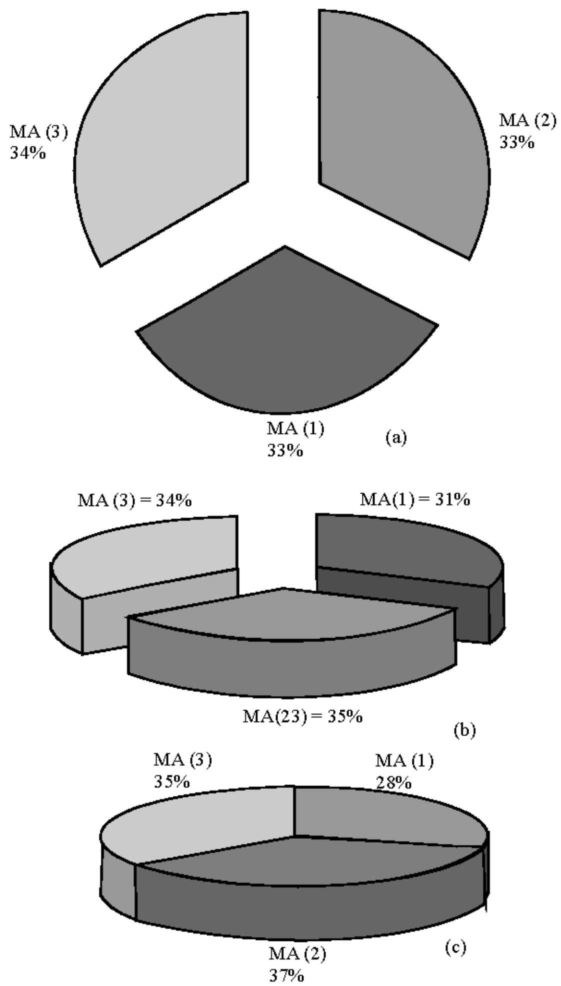

Fig. 3: The pie charts showing the value of moving-average according to the following approaches: (a) MAPE, (b) MAD, (c) MSD

sampled at 200 Hz for 45,000 data points, hence to produce 225 sec of the record length.

The global statistical parameter values were calculated and the results were tabulated in Table 1. Referring to these results in Table 1, lower vibrational energy of the signal has been found based on the root-mean-square value, i.e, 0.104 microstrain. The kurtosis value of this data has been calculated at 13.743, hence, the higher kurtosis value indicated that this data has heavily nonstationary behaviour. In addition, the loading showed the minimum point at the value of -138.72 microstrain and the maximum point was found at the value of 167.13 microstrain.

Equation 14-16 were then used in order to produce the stationary behaviour of this variable amplitude fatigue loadings. These equations were produced by applying the this data in the SPSS software. Then, the data set need to be applied with the moving average approach in order to produce the stationarity pattern.

$$Y_1 = 2.336 + 0.935e_{t-1} + e_t \tag{14}$$

$$Y_2 = 2.335 + 0.498e_{t-1} + 0.920e_{t-2} + e_t \tag{15}$$

$$Y_3 = 2.335 + 0.325e_{t-1} + 0.564e_{t-2} + 0.801e_{t-3} + e_t \tag{16}$$

Using the Mean Absolute Percentage Error (MAPE) criteria, it showed that MA(1) gave the lowest error compared to MA(2) and MA(3), i.e, at the level 32.9%. This result can be referred in Fig. 3a. With the application of the method of Mean Absolute Deviation (MAD), as shown in Fig. 3b, it was found that the MA(1) approach gave the lowest error compared to the MA(2) and MA(3), i.e, 31%. The data was finally analysed using the Mean Squared Deviation (MSD) and it was showed the MA(1) approach again gave the lowest error compared to MA(2) and MA(3). From this analysis, it was found that the error calculated using the MA(1) was 28%, which is shown in Fig. 3c. Based on the results, it is found that the best moving average model is MA(1), which gave the lower error value. The MA(1) is lower than MA (2) and MA (3) because the data for MA (2) and MA (3) is cumulated each other. It means that the best model is MA(1) which is $Y_1 = 2.336 + 0.935e_{t-1} + e_t$

Figure 4 shows the theoretical ACF (Fig. 4a) and PACF (Fig. 4b) for the MA1 model with θ = -0.935. It shows that correlation between every two points is 93.5%. Finally, the distribution exhibited an exponential decay pattern and there were many non-zero partial autocorrelation can also be found in this distribution plot. Thus, it is suggested this data set should be performed with more than 15 moving average segments in order to produce a stable and stationary behaviour of the data.

The plots of Fig. 5a and b showed the theoretical ACF and PACF for the MA(2) models, respectively. For both models, the θ value was found to be at -0.920. The plot also gave a finding in the correlation coefficient

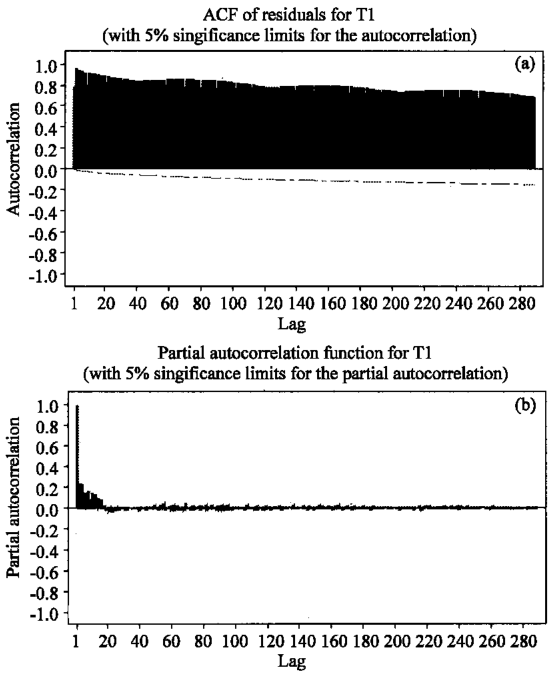

Fig. 4: (a) ACF for MA(1), (b) PACF for MA(1)

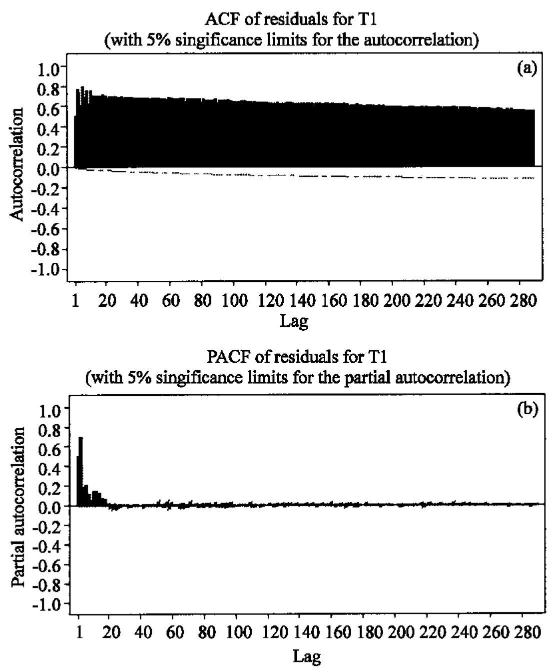

Fig. 5: (a) ACF for MA(2), (b) PACF for MA(2)

for every 2 points, i.e., at the value of 92%. Finally, the distribution exhibited an exponential decay pattern and there were many non-zero partial autocorrelation can also be found in this distribution plot. Thus, it is suggested

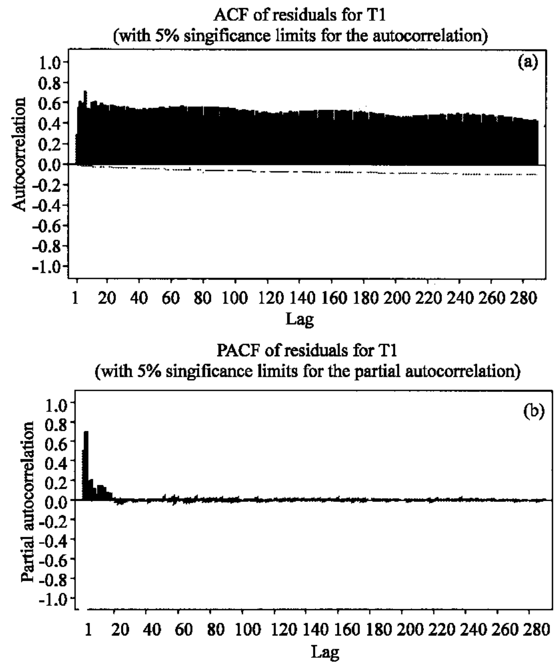

Fig. 6: (a) ACF for MA(3), (b) PACF for MA(3)

this data set should be performed with more than 15 moving average segments in order to produce a stable and stationary data behaviour.

Figure 6 shows the theoretical ACF (Fig. 6a) and PACF (Fig. 6b) for an MA(3) model withθ = -0.801. It shows that correlation between two points is 80.1%. Finally, the distribution exhibited an exponential decay pattern and there were many non-zero partial autocorrelation can also be found in this distribution plot. Thus, it is suggested this data set should be performed with more than 19 moving average segments in order to produce a stable and stationary behaviour of the data.

Based on the autocorrelation analysis of the moving average approaches, it is observed that the MA(1) show the highest correlation with each points than MA(2) and MA(3). The autocorrelation values for MA(1), MA(2) and MA(3) are 0.935, 0.92 and 0.801, respectively. Therefore, it is suggested that that the MA(1) model is found to be more correlated compared to MA(2) and MA(3).

CONCLUSION

The random number generator described in this paper has proven to be relatively portable across different systems, provide a good source of practically strong random data on most systems and can be set up to function independently of special hardware or the need for user or programmer input, which is often not available. From the analysis, finally, it is suggested that MA 1 is the best model because it produced lower error value compared to the values produced from MA 2 and MA 3.

REFERENCES

Agrawal, D. and C.C. Aggawal, 2001. On the Design and Quantify of Privacy Preserving Data Mining Algorithms. Proc. of the 20th ACM SIMOD Symposium on Principles of Database Systems, Santa Barbara, pp: 247-255.

Agrawal, R. and R. Srikant, 2000. Privacy-preserving Data Mining. Proc. of the ACM SIGMOD Conference on Management of Data, Texas, USA, pp: 439-450.

Bendat, J.S. and A.G. Piersol, 1986. Random Data: Analysis and Measurement Procedures, New York: Wiley-Interscience.

Estivill-Castro, V. and L. Brankovic, 1999. Data Swaping: Balancing Privacy Against Precision in Mining for Logic Rules. Proc. of the 1st Conference on Data Warehousing and Knowledge Discovery (DaWaK), Florence, Italy, pp: 389-398.

Harvey, A.C., 1981. Time Series Model. Oxford: Philip Allan Publishers Limited.

Hinton, P.R., 1995. Statistics Explained: A Guide for Social Science Students, London: Routledge.

Holt, C.C., F. Modigliani, J.F. Muth and H.A. Simon, 1960. Planning Production Inventories and Work Force. New Jersey: Prentice Hall.

Kendall, M.G., A. Stuart and K. Ord, 1983. The Advanced Theory of Statistics. London: Charles Griffin.

Makridakis, S., S.C. Wheelwright and R.J. Hyndman, 1998. Forecasting Methods and Applications. (3rd Edn.), New York: John Wiley and Sons, Inc.

Meyer, Y., 1993. Wavelets: Algorithm and Applications. Philadelphia: SIAM.

Natrella, M.G., 1966. Experimental Statistics, Handbook 91, Washington: NIST.

Newland, D.E., 1993. An Introduction to Random Vibrations Spectral and Wavelet Analysis. (3rd Edn.), New York: Longman Scientific and Technical.

Qu, L. and Z. He, 1986. Mechanical Diagnostics, Shanghai: Shanghai Science and Technology Press.

Tacer, B. and P.J. Loughlin, 1998. Nonstationary signal classification using the joint moments of time-frequency distributions, Pattern Recognition, 31: 1635-1641.

Yan, J.H., X.L. Zheng and K. Zhao, 2001. Experimental investigation on the small-load-omitting criterion. Int. J. Fatigue, 23: 403-415.