

Satisfying Statistical Constraints in Preparing Edited Variable Amplitude Loading History Using Genetic Algorithm

Zulkifli Mohd Nopiah¹, Mohd Haniff Osman¹, Shahrum Abdullah^{1,2}, Mohd Noor Baharin Che Kamarudin¹ & Izamarlina Asshari¹

¹ Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, UKM Bangi, Selangor, Malaysia

² Centre for Automotive Research, Universiti Kebangsaan Malaysia, UKM Bangi, Selangor, Malaysia

Correspondence: Mohd Haniff Osman, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, UKM Bangi 43600, Selangor, Malaysia. Tel: 60-3-8921-7043. E-mail: haniff68@eng.ukm.my

Received: July 23, 2012

Accepted: August 18, 2012

Online Published: August 29, 2012

doi:10.5539/mas.v6n9p68

URL: <http://dx.doi.org/10.5539/mas.v6n9p68>

Abstract

A major concern that surfaces when performing the segment-based fatigue data editing technique is to certify that the values of two global statistics (root mean square and kurtosis) of the edited load history are within an acceptance interval whilst maximizing the data reduction rate and minimizing the loss in damage. The root mean square (rms) quantifies an overall energy underlying the history whilst kurtosis is important to identify impulsive character. In this paper, the stochastic Genetic Algorithm (GA) is employed as a post processing tool that helps the edited history satisfy the statistical requirements with minimum cost i.e. small decrement in the initial reduction rate. Consider the initial version of edited history being composed of high fatigue damage segments resulted from the non-overlapping segmentation method. In a case that the history does not comply with the statistical requirements, then importing a subset of low segments into the present edited history might reverse the outcome. Thus, the GA aims to search for the smallest subset that turns the history into fulfilling the rms and kurtosis needs without affecting the reduction rate too much. Experimental results show the capability of the proposed method in making the edited history fit the statistical constraints without imposing harm on the overall fatigue damage value.

Keywords: fatigue data editing, combination optimization problem, genetic algorithm, fatigue segment, root mean square, kurtosis

1. Introduction

Time history of load (load history) has great influence in durability testing which is associated with fatigue failure analysis. Fatigue failure may be described as a product of oscillatory actions under varying loads at one or several stress concentration points. Prior to commencing the testing, it is recommended to edit load history in making a cost-effective environment e.g. short operation time and save money (Abdullah, Choi, Giacomini, & Yates, 2006; Petracconi, Ferreira, & Palma, 2010). The editing refers to the effort of simplifying a load history and it can be done by removing small amplitude cycles that make up a large percentage of generated cycles in the history. To date, those cycles are found individually as appeared in (Stephens, Dindinger, & Gunger, 1997) or based on the segment-to-segment analysis.

Segment-based fatigue data editing technique can be seen as a series of analysis involving segmentation, labelling and selection. A proper segmentation method partitions the loading history into number of meaningful segments. Once completion of labelling the segments, low (or high) damage segments are removed (or kept) to shorten the history. In depth segment-to-segment analysis provided details that assist editing techniques under the approach preserving cumulative fatigue damage and retaining load sequence in edited history. However, it is found that most of the techniques mentioned have not dealt with satisfying the statistical constraints- those of which are other criteria that edited history must meet in such detail. For example, authors in (Putra, Abdullah, Nuawi, & Nopiah, 2010) simply apply a try-and-error approach in their respective algorithm to ensure the values of rms and kurtosis of edited history that appear within the acceptance interval i.e. within 10% deviation. The approach could be seen as a negative approach compared to its predecessor, where an incremental step correction method is applied (Abdullah et al., 2006). Meanwhile, techniques in (Abdullah, Nidzwan, & Nuawi, 2009;

Nopiah, Baharin, Abdullah, Khairir, & Ariffin, 2010) do not explain how the edited history fits the constraints i.e. conclusion is merely depending on the comparative results. It could be summed up that the authors overlook the fact that the data point in the data segment contributes to the statistical calculation. Thus, it might more meaningful and becoming less complex if the statistical constraints could be confronted separately i.e. not in the same level with other objectives. A simple way to perform is by identifying and removing only undamaging segments that contribute less changes in the history's statistics. In order to minimise risk when performing segment selection thus it is reformulated as a combinatorial optimization problem.

For the combinatorial problem, particularly the combination without repetitions, an optimization tool search for the best combination consists of r objects out of $n!/r!(n-r)!$ finite possible outcomes that minimize or maximize certain objective functions defined on some domains. However, solving a combinatorial problem i.e. finding optimal solution(s) is NP-hard, and it becomes worse as the value of parameter r is unknown i.e. significantly expanding the search space (Ahmed, 2010). Therefore, a heuristic algorithm e.g. the GA is often used as an approximation algorithm to produce good solutions (so-called approximate solutions) (Shiqiong et al., 2008). The GA implements a searching strategy based on natural evolution principle to seek an optimal solution. Besides no requirement to examine the data structure and to compile it with any auxiliary knowledge, the ability to perform solution exploration in many directions in large search space simultaneously have made GA a popular choice for solving combination optimization problems (Fung, Kwong, Siu, & Yu, 2012; Konak, Colt, & Smith, 2006).

In this study, a time series non-overlapping window segmentation method with fixed window length is applied to loading history; the easiest way to get meaningful and inarguable non-overlapped fatigue segments. Then, each of them is classified according to the damage level. Once the segments are located and labelled, then preliminary edited history is built by joining together all high damage segments. Here, we assume that the above strategy cannot guarantee satisfaction on statistical constraints. One of the possible ways to counter the problem is by retuning a subset of low fatigue segments into a sequence of history. Undoubtedly, this action will reduce the current data reduction rate due to more fatigue segments participating in the edited version; however, those extra segments contribute to total damage even if the amount is small (Wu, Liou, & Tse, 1997). To avoid unnecessary loss in reduction rate, the GA is applied to a collection of low damage fatigue segments. The GA aims to search for the best low segment combination to join a set of high fatigue segments in order to build applicable edited history at minimum cost. A combination cost i.e. fitness function is evaluated based on the size of segment subset weighted by a unique penalty function. Subset i.e. chromosome with smaller fitness value represents the subset that is worthier to be sent back inline. Comparison results on strain-stress cycle properties exhibits the effectiveness of the proposed GA-based segment selection.

2. Research Methods

2.1 Overview on Genetic Algorithm

Genetic algorithm is a stochastic search algorithm inspired from Darwin's evolution theory. The search uses an evolutionary strategy i.e. survival of the fittest among the string environment to solve a problem e.g. optimization problem. GA starts with zero knowledge about the true solution and entirely depends on the genetic operators such as selection, crossover and mutation to expand the initial population to reach optimal solution. In GA terminology, a population consists of large number of possible solutions called chromosome or individual. Each chromosome is evaluated by the fitness function, where in most optimization cases, it is directly defined by objectives of the search. Figure 1 illustrates the core operation of GA (basic mode) called reproduction cycle. On top of the cycle, s chromosomes are randomly selected and compared. A winner which is the fittest chromosome is passed into a mating pool while all the remaining is moved back to the recent population. Then the population is reshuffled and the selection process continues until the pool is fully-loaded. The size of pool is usually half of the population size, m . In the second phase, the use of a common single-point crossover is able to cross-join two random pool members. At this point, exchanging tails between two chromosomes automatically breeds two offspring where they are probably fitter than their parents. In some occasions, an offspring is produced via mutation, not from the crossover operation. By mutating one of the parent's genes, a new chromosome is obtained. The second phase is over after $m/2$ repetitions. Finally, both populations are combined and sorted in the descending order. Only top m chromosomes are later moved to substitute the previous population. The whole process is repeated over t times but it sometimes stops once it has met other predefined stopping criteria.

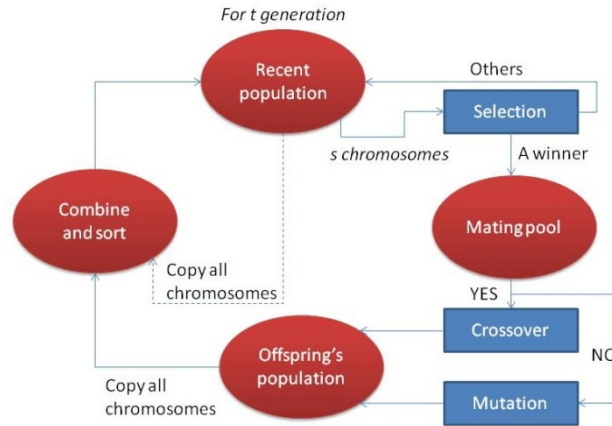


Figure 1. Basic process of GA

2.2 Fatigue Segment Combination Problem

In this study, loading history H is a time series of N points measured in the unit of microstrain ($\mu\epsilon$), $H = \{x_j, j = 1, \dots, N\}$. The history is transformed into a row of non-overlapping (disjoint) data segments where each consists of 500 points. The figure corresponds to sampling rate used in the data acquisition. This is a conservative but reliable way to split the time series into M meaningful (informative) segments ("How fast should a signal be sampled?"). Now, the history is defined to be $H = S_1 \cup S_2 \dots \cup \dots S_M$ where the first segment $S_1 = \{x_1, x_2, \dots, x_{500}\}$ while for the subsequent i th segment $S_{i+1} = \{x_{500i+1}, x_{500i+2}, \dots, x_{500(i+1)} : 1 \leq i \leq M - 1\}$. The segments were then classified either to low or high damage. The fatigue segment combination (FSCo) problem can be described as follows:

$$\text{minimize } |S| \tag{1}$$

subject to

$$(1 - a)RMS(H) \leq RMS(Edt) \leq (1 + a)RMS(H) \tag{2}$$

$$(1 - a)Kurtosis(H) \leq Kurtosis(Edt) \leq (1 + a)Kurtosis(H) \tag{3}$$

where S is a subset of a collection of low damage segments L i.e. $S \subset L$ and an edited history Edt constitutes all high fatigue segments associated with the S . $RMS(X)$ returns rms value of time series X . Similarly, $kurtosis(X)$ returns kurtosis value of time series X . The objective is to find the optimal set S^* that satisfies both constraints (refers Eqs. 2-3) where each of which is an interval that corresponds to lower and upper limits of relative statistical parameter at the degree of tolerance α equals 0.1 (Nopiah, Abdullah, et al., 2010), with minimum cost i.e. small size of S .

2.3 Genetic Algorithm for the FSCo

Having appropriate fitness function associated to the problem domain is very important in the GA formulation to avoid distortion in the fitness evaluation of chromosome i.e. solution candidate. For the FSCo problem, Eq. 1 cannot be used directly as a fitness function because it does not have the ability to distinguish whether the participation of S causes the edited history to fit the constraints or not. The easiest way to overcome this weakness is by assigning such kinds of flag to Eq. 1 and we therefore propose the following form for FSCo fitness function:

$$\text{minimize } |S| + p \tag{4}$$

where a penalty function p returns $|H|$ if the chromosome violates any of the constraints, otherwise 0. This forces the irrelevant combination far out from the neighbourhood of optimal point. The following subsection briefly explains central GA parameters, such as chromosome representation, crossover and mutation operator, etc.

2.3.1 Chromosome Encoding

The S^* separates the collection of low fatigue segment into two groups; 1) chosen to participate in the edited history and 2) permanently removed. Thus, it is suitable to use standard binary-coded chromosome to encode the solution candidate. Here, all chromosomes have fixed N genes where N depends on the number of low segments resulted from the labelling task. A gene with the binary value of 0 representing a segment does not appear in the corresponding S ; if the value is 1, it does appear. Consider $N = 10$, a chromosome with bits strings of 0110110000 that indicates that only the 2nd, 3rd, 5th and 6th segments are subscribed by S .

2.3.2 Genetic Operators

The settings of GA parameters used in the FSCo optimization problem are given in Table 1.

Table 1. Configuration of GA parameters for the FSCo problem

Component	Configuration	Description
Selection	Tournament selection	At default value of two, two random chromosomes are selected and their fitness value determines the winner.
Crossover	Single-point crossover	Structure of two chromosomes is mutually cross-exchanged where it occurs at probability of 0.9. A joint point is determined in a random way.
Mutation	Flip-coined	The corresponding binary value of selected gene is simply flipped.
Population size	Percentage of total population size	Initial population is populated with 30,000 chromosomes.
Generation size	Fixed	Set to 1,000.

2.4 Experimental Design

Two automotive variable amplitude loading histories employed in literature (Nopiah, Baharin, et al., 2010) labelled as S1 and S2 were considered in the performance evaluation of the GA-based fatigue segment combination. For each history, the following activities were applied to generate inputs to the proposed system.

- i. For scaling purposes, observation points of history were initially normalized to a range of -999 to 999. The corresponding rms and kurtosis value set the constraints.
- ii. The history was then split into $M = |H|/s$ non-overlapping segments where s is a sampling rate.
- iii. Next, the Palmgren-Miner's linear damage rule (Baek, Cho, & Joo, 2008) was used to calculate damage value for obtained segments meanwhile the strain-life Morrow model (Downing, 2004) involved in fatigue life estimation. For labelling purpose, a fatigue segment is conveniently labelled as low damage (high cycle) if damage value is less than 10^{-3} otherwise it will be labelled as high damage (low cycle) (Abdullah et al., 2010).
- iv. Low damage segments were pulled out from the history and grouped into set L while the remaining segments were joined together as an initial edited history.

After the FSCo, the edited history suggested by the GA and the full length history were cycle-counted using the rainflow counting algorithm. Comparison results on the distribution of cycle properties e.g. mean and amplitude measures the performance of our method. Since GA is stochastic, then the experiment has been repeated for 10 times, each of which was populated randomly. The best result for each history is shown in the next section.

3. Results and Discussion

Some exact sixty 500-point non-overlapping segments were located in each of S1 and S2. The corresponding damage level of those segments is graphically shown in Figure 2. The plots show imbalance in class distribution, where a number of low damage is about 2-4 times more than the high damage. This finding corresponds to the nature of the loading history in which the majority of cycles come in a small amplitude cycle (Stephens et al., 1997). Using these figures the chromosome length is set to 47 and 50 for S1 and S2, respectively. This indicates that it provides extremely large space for the GA to search for the S^* . It can be estimated that there would be

$2^{47} - 1$ possible low segment combinations for S1 and that for S2 is $2^{50} - 1$. Due to the enormous size of search area, the solution that GA provided for FSCo probably is a nearly optimal not the true optimal.

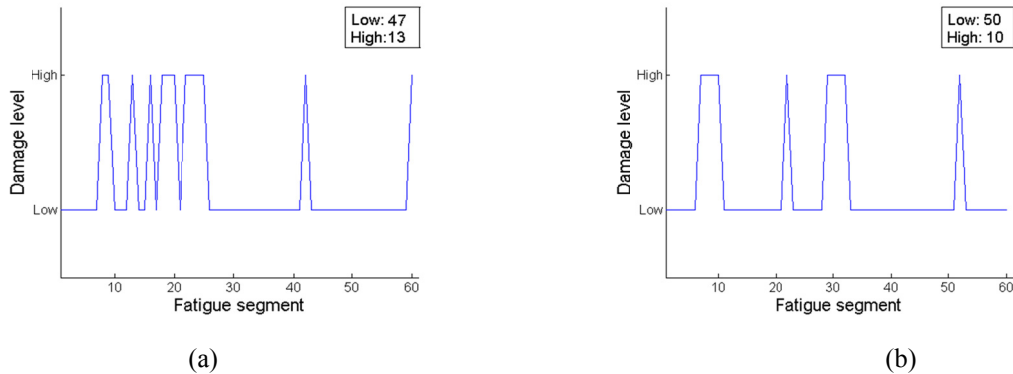


Figure 2. Outputs of labelling procedure for obtained segments corresponds to a) the S1 history and b) the S2 history

Figure 3 shows the scatter diagram of the initial edited history and GA results about the original values of rms and kurtosis. From the diagram, it can be observed that kurtosis is the parameter that is most affected when only high damage segments are considered when preparing the edited history. It is somewhat expected due to the kurtosis which involves mean and variance that have made it very sensitive to data point changes. For S1, the optimal subset found by GA consists of 11 segments that have reduced the reduction rate by $(78\% - 60\% = 18\%)$. However, the new rate is still far better compared to the performance recorded by the RDE technique (Nopiah, Baharin, et al., 2010) i.e. 60% versus 7%. The GA results of the S2 are similar to that of S1 but the decrement relative to the reduction rate is slightly higher i.e. $(83\% - 62\% = 21\%)$. Apart from the optimal results, the diagram also locates the closest solution to the original with respect to the maximum of 50% loss in the reduction rate i.e. size of S cannot exceed half of the size of L (remarks with '+').

Figure 4a and Figure 5a show time history of S1 and S2 in full length compared to two outputs of the editing process where the first ones are without the participation of the GA-FSCo (see Figure 4b-c and Figure 5b-c). The three dimensional cycle histogram with regards to the mean strain and amplitude strain of S1 history before and after editing process (refer to the Figure 4c) are compared in Figure 6a-b. Of the 3980 stress-strain cycles, just 35% participated in the edited version. Moreover it is interesting to note that the scale used in both versions for both cycle properties remains unchanged. The most important thing to highlight is that large number of small amplitude cycles present in the original version was removed (refer bar chart in Figure 7). This evidence points out that the damage value in both versions is highly possibly similar (or approximately similar) since those cycles are responsible for zero or minimal amount of damage. It is valid at least for linear Palmgren-Miner damage rule. The above comparison task was repeated for the S2 in which similar patterns as the S1 have appeared (not presented in this study). The empirical results show the applicability of the proposed GA-based fatigue segment combination for preparing applicable edited loading history; retains the damage value and keeps statistical values within acceptance interval.

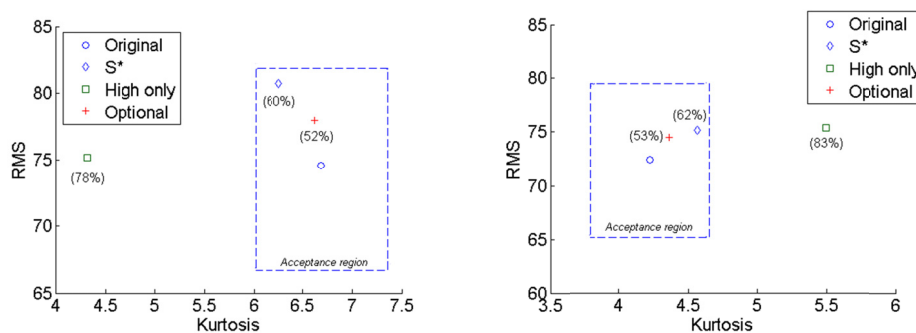


Figure 3. Scatter plots for the original, edited history without GA and two solutions when the GA is operated on a) the S1 and b) the S2

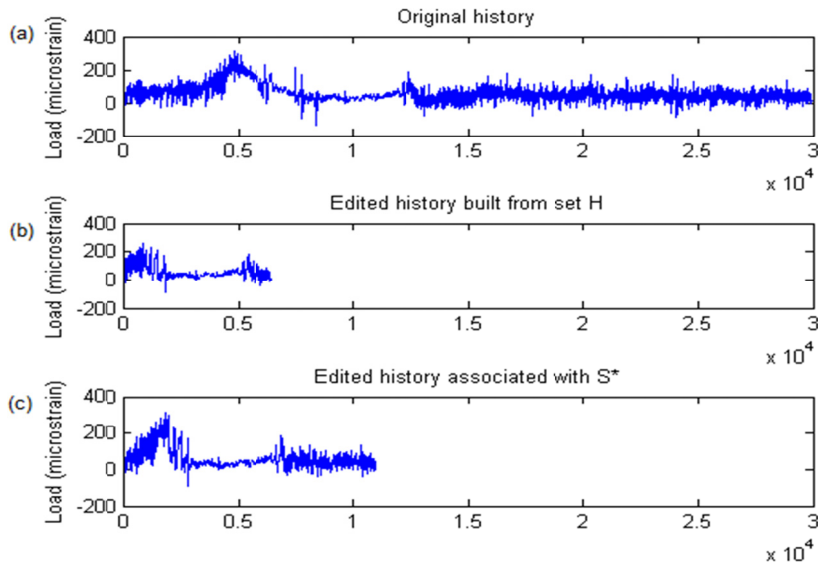


Figure 4. Variable amplitude loading history of S1. a) Full length version, b) Shorten version before application of the GA, and c) Shorten version associates with the S^*

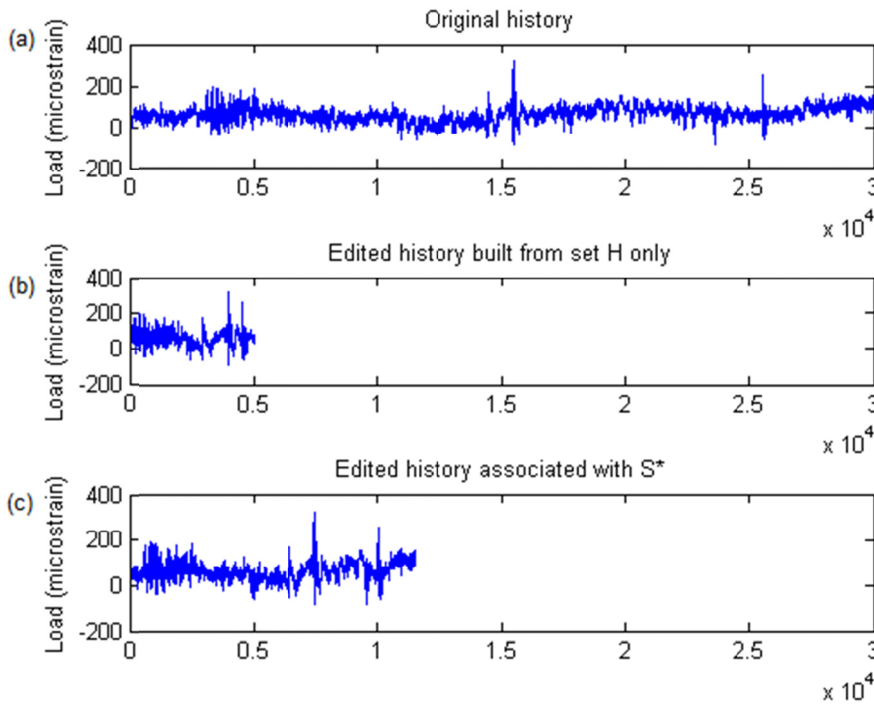


Figure 5. Variable amplitude loading history of S2. a) Full length version, b) Shorten version before application of the GA, and c) Shorten version associates with the S^*

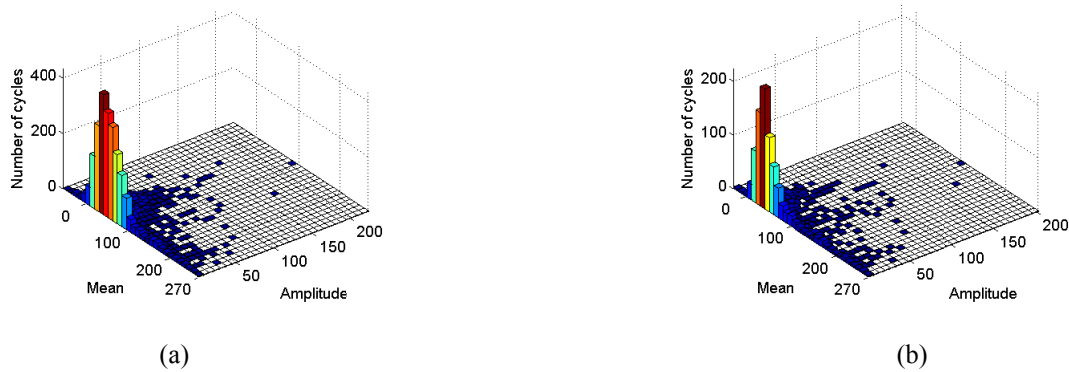


Figure 6. Cycle histogram for a) full length of S1, and b) the S1's shorten version

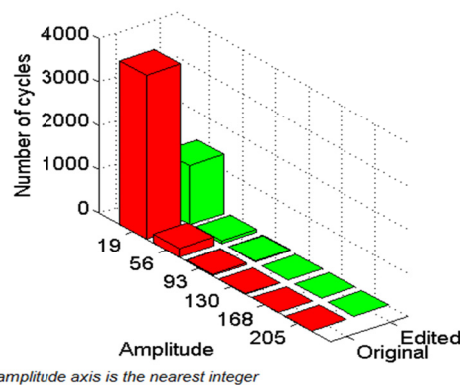


Figure 7. Amplitude cycle histogram of uncut and edited history of S1

4. Conclusion

A genetic search strategy is presented for fatigue segment combination problem of the strain variable amplitude loading history subject to SAE1045 steel. The genetic algorithms aim to find the smallest subset of low fatigue segment that can position the root mean square and kurtosis values of the edited history into the acceptance intervals. The introduction of special penalty function into the fitness function has successfully smoothed the operation search. Besides that, the penalty function clustered the irrelevant combinations found along the search. Through a series of comparisons on two loading histories, it can be stated that through the GA-based for FSCo the unsatisfied edited history can be altered. With small scarification on length reduction rate by importing several selected low fatigue segment, the edited history becomes applicable. For S1 edited history, there was a small decrease in the length reduction rate i.e. 78% to 60%. Meanwhile for S2, loss in the rate is slightly high compares to S1 in which 62% instead of 83%. Taken together, the proposed method did not put in any harm on the cumulative fatigue damage value the edited history had because it has not removed any high damage segments. In fact there was an increment in damage value even in small amount. Comparison on the cycle histogram between the original history and the edited version (after the participation of GA) supports the above claims.

What is now needed is a review study on how to accelerate the GA search process. It is significant because the larger the size of low damage class, the more complex the search space. Another perspective is to further investigate the capability of the proposed method in dealing with outputs gained from other segmentation algorithms. It is also suggested that the joint application of the non-overlapped segmentation method and the GA for the fatigue segment combination performs on longer loading histories.

Acknowledgement

The authors would like to acknowledge financial support from the Ministry of Higher Education of Malaysia MOHE under Grant No. UKM-KK-06-FRGS-0090-2010 and UKM-KK-07-FRGS0216-2010.

References

- Abdullah, S., Choi, J. C., Giacomini, J. A., & Yates, J. R. (2006). Bump extraction algorithm for variable amplitude fatigue loading. *International Journal of Fatigue*, 28(7), 675-691. <http://dx.doi.org/10.1016/j.ijfatigue.2005.09.003>
- Abdullah, S., Nidzwan, C. K. E., & Nuawi, M. Z. (2009). A study of fatigue data editing using the Short-Time Fourier Transform (STFT). *American Journal of Applied Sciences*, 6(4), 565-575. <http://dx.doi.org/10.3844/ajassp.2009.565.575>
- Abdullah, S., Putra, T. E., Nuawi, M. Z., Nopiah, Z. M., Arifin, A., & Abdullah, L. (2010). The Morlet wavelet analysis for fatigue feature llustering. *Wseas Transactions On Mathematics*, 9(5), 345-354.
- Ahmed, Z. H. (2010). Genetic algorithm for the traveling salesman problem using sequential constructive crossover operator. *International Journal of Biometric and Bioinformatics*, 3(6), 96-105.
- Baek, S., Cho, S., & Joo, W. (2008). Fatigue life prediction based on the rainflow cycle counting method for the end beam of a freight car bogie. *International Journal of Automotive Technology*, 9(1), 95-101. <http://dx.doi.org/10.1007/s12239-008-0012-y>
- Downing, N. E. (2004). *Mean stress effects in stress-life and strain-life fatigue*. Paper presented at the Fatigue 2004: Second SAE Brasil international conference on fatigue, Sao Paulo, Brazil.
- Fung, K. Y., Kwong, C. K., Siu, K. W. M., & Yu, K. M. (2012). A multi-objective genetic algorithm approach to rule mining for affective product design. *Expert Systems with Applications*, 39(8), 7411-7419. <http://dx.doi.org/10.1016/j.eswa.2012.01.065>
- How fast should a signal be sampled? (2012). Retrieved from <http://www.mathworks.com/help/toolbox/daq/f5-30343.html#f5-28876>
- Konak, A., Colt, D. W., & E.Smith, A. (2006). Multi-objective optimization genetic algorithms: A tutorial. *Reliability Engineering and System Safety* 91, 992-1007.
- Nopiah, Z. M., Abdullah, S., Baharin, M. N., Putra, T. E., Sahadan, S. N., & Willis., K. O. (2010). Comparative study on data editing techniques for fatigue time series signal. *Advanced Materials Research*, 146-147, 1681-1684. <http://dx.doi.org/10.4028/www.scientific.net/AMR.146-147.1681>
- Nopiah, Z. M., Baharin, M. N., Abdullah, S., Khairir, M. I., & Ariffin, A. (2010). Running damage extraction technique for identifying fatigue damaging events. *WSEAS Transactions On Mathematics*, 9(5), 324-333.
- Petracconi, C. L., Ferreira, S. E., & Palma, E. S. (2010). Fatigue life simulation of a rear tow hook assembly of a passenger car. *Engineering Failure Analysis*, 17(2), 455-463. <http://dx.doi.org/10.1016/j.engfailanal.2009.09.002>
- Putra, T. E., Abdullah, S., Nuawi, M. Z., & Nopiah, Z. M. (2010). Wavelet coefficient extraction algorithm for extracting fatigue features in variable amplitude fatigue loading. *Journal of Applied Sciences*, 10(4), 277-283. <http://dx.doi.org/10.3923/jas.2010.277.283>
- Shiqiong, Z., Longyun, K., Guifang, G., Yanning, Z., Jianbo, C., & Binggang, C. (2008). *The application of combinatorial optimization by Genetic Algorithm and Neural Network*. Paper presented at the Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on.
- Stephens, R. I., Dindinger, P. M., & Gungler, J. E. (1997). Fatigue damage editing for accelerated durability testing using strain range and SWT parameter criteria. *International Journal of Fatigue*, 19(8-9), 599-606. [http://dx.doi.org/10.1016/S0142-1123\(97\)00079-0](http://dx.doi.org/10.1016/S0142-1123(97)00079-0)
- Wu, W. F., Liou, H. Y., & Tse, H. C. (1997). Estimation of fatigue damage and fatigue life of components under random loading. *International Journal of Pressure Vessels and Piping*, 72(3), 243-249. [http://dx.doi.org/10.1016/S0308-0161\(97\)00053-7](http://dx.doi.org/10.1016/S0308-0161(97)00053-7)