Structural Engineering and Mechanics, Vol. 33, No. 5 (2009) 589-604 589
DOI: http://dx.doi.org/10.12989/sem.2009.33.5.589

Load spectra growth modelling and extrapolation with REBMIX

Matej Volk[†], Matija Fajdiga[‡] and Marko Nagode[‡†]

University of Ljubljana, Faculty of Mechanical Engineering, Aškerčeva 6, SI-1000 Ljubljana, Slovenia

(Received June 15, 2009, Accepted September 23, 2009)

Abstract. In the field of predicting structural safety and reliability the operating conditions play an essential role. Since the time and cost limitations are a significant factors in engineering it is important to predict the future operating conditions as close to the actual state as possible from small amount of available data. Because of the randomness of the environment the shape of measured load spectra can vary considerably and therefore simple distribution functions are frequently not sufficient for their modelling. Thus mixed distribution functions have to be used. In general their major weakness is the complicated calculation of unknown parameters. The scope of the paper is to investigate the load spectra growth for actual operating conditions and to investigate the modelling and extrapolation of load spectra with algorithm for mixed distribution estimation, REBMIX. The data obtained from the measurements of wheel forces and the braking moment on proving ground is used to generate load spectra.

Keywords: load spectra growth; extrapolation; reliability; mixed distribution function; alternative algorithm.

1. Introduction

Growing requirements for the reliability of components and cost reduction demand the usage of modern methods for structural evaluation where actual operating and environmental conditions are used in the early phases of design (Ebeling 1997). The service life of a structural component under random loading depends primarily on loading conditions, material properties, geometry and the technology of manufacturing.

In the field of reliability and fatigue analysis, the prediction of the load ranges applied to the structural component during operation in actual operating conditions is still one of the most difficult tasks. Load time histories can be obtained by tests on the proving ground or by simulations. Tests are usually very expensive and time consuming, but they cannot be avoided entirely. They can be reduced only by improving the existing methods and by developing new methods of load spectra modelling. During the service life of the component loads change and repeat constantly, which means that the load spectrum grows in time. It has been shown by many research works, such as

† B.Sc., Young Researcher, Corresponding author, E-mail: matej.volk@fs.uni-lj.si
‡ Professor, Ph.D., E-mail: matija.fajdiga@fs.uni-lj.si
‡† Associate Professor, Ph.D., E-mail: marko.nagode@fs.uni-lj.si

(Zhao and Baker 1992, Grubisic and Fischer 1997, Nagode and Fajdiga 1998b, 1999, Tovo 2000) that if the operating conditions are not constant during service life, the shape of representative load spectra is multi-modal and can not be expressed with unimodal distribution functions. It has been shown by Buxbaum and Zaschel (1979) that all possible basic shapes of load spectra can be approximated well by the two-parameter Weibull distribution. If a load spectrum is composed of more than one basic shape, the mixed Weibull distribution has been suggested by Nagode and Fajdiga (1998a).

The main topic of this paper is to investigate the load spectra growth for the actually measured data and to investigate the quantity of data needed for reliable prediction of future load cycles with the algorithm for prediction of mixed distribution functions, REBMIX which is the acronym for the Rough and Enhanced component parameter estimation that is followed by the Bayesian classification of the remaining observations for the finite MIXture estimation. During the evaluation of the algorithm with statistical testing some improvements were made especially in the field of estimation the lognormal distributed data with newly introduced lnnormal procedure which in certain cases much better approximates the actual distribution. Because the operating conditions are not constant, the shape of load spectra is multi-modal and therefore mixed distribution functions have to be used. Mixed distribution functions in this case are combined of c basic functions belonging to a certain parametric family (e.g., Weibull, normal or lognormal). In the paper it is shown how the most suitable parametric family is obtained and how the optimal number of components (basic functions) in the mixed distribution function is obtained for a certain load time history.

Measured load time histories are divided into four segments from which the load spectra is modelled and extrapolated to the time period of the whole load time history. The extrapolated load spectra are compared with the actually measured load spectrum belonging to the whole load time history. It can be concluded that the extrapolated load spectra modelled with REBMIX in general coincide well with the actually measured load spectrum in the field of the cycles with high probability of occurrence and underestimates the actual load ranges in the field of the cycles with high load amplitudes and low probability of occurrence. It is also shown that in the case of short load time histories algorithm well estimates the load spectra for rather stationary loading conditions whereas larger deviations occur at non-stationary sample data. Additionally the effect of filtering small load cycles on the ability of the algorithm to estimate the mixed distribution function is studied. It turns out that for rather stationary load histories differences between estimated mixed distributions are small whereas for non-stationary load histories filtering plays essential role at mixed distribution estimation.

2. Theoretical background

2.1 Algorithm for parameter estimation of mixed distribution functions

REBMIX is an algorithm developed by Nagode and Fajdiga (1998a, 2000, 2006). In the literature it has been known as an alternative algorithm so far.

The algorithm is used to estimate the number of components, their weights and parameters in mixed distribution functions. A predictive mixed probability density function (p.d.f.) is defined as

$$f(s) = \sum_{l=1}^{c} w_l f_l(s) \tag{1}$$

where constant w_l stands for component weight with characteristics $w_l > 0$ ($l = 1, \ldots, c$), $\sum_{l=1}^{c} w_l = 1$ and $f_l(s)$ represents arbitrary predictive component densities. Generally speaking, $f(s)$ can be composed of c component distributions $f_l(s)$, each of a different type. Because of the difficult detection of the mixed p.d.f. composed of different component distributions, significant simplification is used by selecting all $f_l(s)$ of the same type.

The function of the algorithm is described well in (Nagode and Fajdiga 1998a, 1999, 2000, 2006, Nagode *et al.* 2001). Only the basic functionality of the algorithm will be presented here. First, the number of components, the mixture component parameters and the component weights are modelled jointly. Second, the set of independent observations is counted into a finite number of histogram bins and bin frequencies are calculated. Third, the mixture component parameters and the component weights are estimated. This process consists of the parameter estimation and the separation of observations. Finally, the residual histogram (residue) is distributed between the components by the Bayes decision rule (Nagode and Fajdiga 2006).

The characteristics of the algorithm are:

- The initial component parameter estimation is not required.
- The number of components does not have to be known.
- The procedure is numerically stable and is insensitive to singularities due to the rough parameter estimation.
- The number of components has no influence upon convergence.
- The speed of convergence is very high.
- The mixture components and component weights are estimated successively and not simultaneously.

It turns out that the selection of the proper component distribution is of highest importance, especially in the extrapolated region of the load spectra. Total number of component distributions c is also dependent of the choice of $f_l(s)$. The most convenient choice among all possible selections for $f_l(s)$ seems to be the Weibull distribution, because of its generality and ability to aproximate other distributions with the proper selection of its parameters

$$f_l(s) = \frac{\beta_l}{\theta_l}\left(\frac{s}{\theta_l}\right)^{\beta_l - 1} \exp\left(-\left(\frac{s}{\theta_l}\right)^{\beta_l}\right) \tag{2}$$

where constants β_l and θ_l represent the Weibull shape and scale parameters respectively. The Weibull distribution can approximate the normal or lognormal one and is exactly equal to the exponential and Rayleigh distribution for $\beta_l = 1$ and 2 respectively. The mixed p.d.f. composed of c Weibull distributions is therefore

$$f(s) = \sum_{l=1}^{c} w_l \frac{\beta_l}{\theta_l}\left(\frac{s}{\theta_l}\right)^{\beta_l - 1} \exp\left(-\left(\frac{s}{\theta_l}\right)^{\beta_l}\right) \tag{3}$$

and its cumulative distribution function (c.d.f.)

$$F(s) = 1 - \sum_{l=1}^{c} w_l \exp\left(-\left(\frac{s}{\theta_l}\right)^{\beta_l}\right) \tag{4}$$

Apart from the Weibull distribution function, a normal or lognormal distribution function can be selected for component distributions. The mixed p.d.f. composed of c normal distributions is therefore

$$f(s) = \sum_{l=1}^{c} w_l \frac{1}{\sigma_l\sqrt{2\pi}} \exp\left(-\frac{(s-\mu_l)^2}{2\sigma_l^2}\right) \tag{5}$$

and the lognormal

$$f(s) = \sum_{l=1}^{c} w_l \frac{1}{s\sigma_l\sqrt{2\pi}} \exp\left(-\frac{(\ln(s)-\mu_l)^2}{2\sigma_l^2}\right) \tag{6}$$

where constants μ_l and σ_l represent the means and standard deviations respectively.

The corresponding cumulative distribution functions are therefore for the mixed normal distribution

$$F(s) = \sum_{l=1}^{c} w_l\left(\frac{1}{2}\left(1 + erf\left(\frac{s-\mu_l}{\sigma_l\sqrt{2}}\right)\right)\right) \tag{7}$$

and for the lognormal one

$$F(s) = \sum_{l=1}^{c} w_l\left(\frac{1}{2}\left(1 + erf\left(\frac{\ln(s)-\mu_l}{\sigma_l\sqrt{2}}\right)\right)\right) \tag{8}$$

2.2 Lnnormal procedure

In the statistical testing, where the ability of the algorithm to recognize two or more component distributions (i.e., homogeneity test for the case of Weibull in (Stehlík 2006) and in the limiting case for the lognormal distribution in (Stehlík 2008)) and sensitivity of the algorithm to highly deviating data points (i.e., outlier test (Balakrishnan and David 2001)) was tested, it turned out that for the lognormal test data the algorithm returns poor approximation in first few histogram bins where the majority of the data is present. The poor approximation in the first few histogram bins is a consequence of the way of counting the observations into the histogram. The algorithm counts the observations in a finite number of equally sized histogram bins, which leads to the deviations between the measured and approximated data. This problem is most evident in the case of a small number of bins. To minimize the problem, the so called lnnormal procedure was proposed. The procedure consists of calculating the logarithm of sample data and estimating the parameters of the normal distribution function. By calculating the antilogarithm of histogram bin borders, the bins are not of the same width any more, which makes it possible to condense the number of narrower bins in the region with the majority of the data and widen the bin widths in the field with fewer observations. In this way greater accuracy is achieved in the first few histogram bins. The difference between the lognormal distribution function and lnnormal procedure is shown in Fig. 1. The sample data, consisting of 1000 sample points belonging to one lognormal distribution function (μ = 1 kN and σ = 0,5 kN) are generated with a random number generator. Probability density functions are estimated by selecting the lognormal for generated points and normal parametric family for the logarithms of generated points respectively. As presented in the figure, lnnormal procedure approximates the actual distribution (thin dash dot line) much better in the case of a small number of histogram bins (the number of bins used in this example is 10). Lognormal parametric family approximates these sample data with the mixture of two lognormal distributions (dashed lines in Fig. 1(a)), while lnnormal procedure returns just one. In the case of a larger number of bins, the

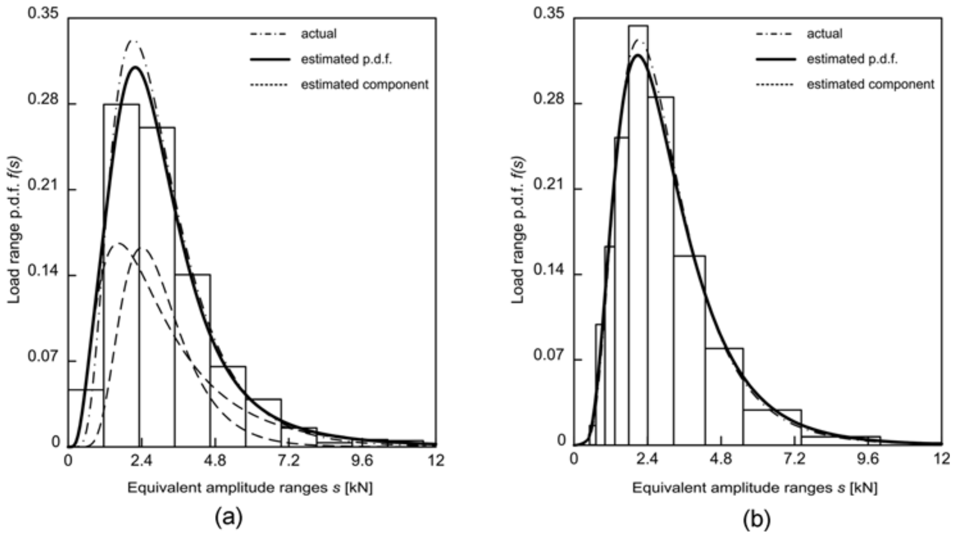

Fig. 1 (a) Actual and estimated lognormal distribution function. (b) Actual lognormal and estimated distribution function with lnnormal procedure

difference between them is not as evident any more, so lognormal distribution seems more appropriate in this case as it is much quicker.

2.3 Basic definition of load spectrum

Load spectrum $H(s)$ is defined on the basis of c.d.f. $F(s)$ by the expression (Nagode and Fajdiga 1998a, Bučar 2006)

$$H(s) = N_c[1 - F(s)] \tag{9}$$

where N_c is the number of all load cycles and represents the range of load spectrum. If the variable s represents the scalar value, e.g., the amplitude of cycle S_a, then the load spectrum is shown in the diagram with the abscissa representing number of cycles N_c and the ordinate representing variable S_a. If the product is subject to random dynamic loading, then the range of load spectrum grows in time. This phenomenon is called the load spectrum growth.

In the early phase of construction, the random loads are estimated by relatively short load time histories because of time and cost limitations. To predict the endurance limit of a randomly loaded structural component, it is important that the sample load time history represents the expected loading of the component as closely as possible to the actual operating conditions. To predict load spectra at any time in the future, it is necessary to extrapolate the measured or modelled load spectra. For extrapolation of load spectra different procedures are described in (Nagode and Fajdiga 1998b, Nagode *et al.* 2001, Socie 2001, Johannesson 2006). Here the load spectra is shifted to the right and extrapolated in the field of load cycles with low probability of occurrence by fitting the statistical distribution function to the data.

In the operating phase of a structural component under real working conditions it is possible, by constant measurements of actual loads, to constantly build and predict the load spectra by the extrapolation in the field of load cycles with low probability of occurrence. In this way it is possible to predict the damage accumulation and fatigue of observed components more accurate than in the case of predicting the service life from a short load time history in the early phase of the construction. This kind of continuous monitoring of components or structures by constant measurements is especially useful at determination of the reliability of superstructures (e.g., very long bridges or very tall skyscrapers) or at determination of the reliability of some vital components of construction from the point of people safety or functionality.

3. Modelling and extrapolation of load spectrum with rebmix

To model and extrapolate load spectra, the measurements of wheel forces and the braking moment on proving ground are used. The load time histories of all three wheel forces and the braking moment are shown in Fig. 2. It can be seen that the operating conditions change over time and therefore it can be expected that the corresponding load spectra are composed of more than one basic shape. It can also be expected that the extrapolated load spectra will not completely coincide with the actually measured load spectrum from all measured data especially at the first and the last load time history representing the wheel force in the *x* direction and the braking moment, because of very big differences in load cycles among the segments of the load time history. Better agreement between the extrapolated load spectra and the actually measured load spectrum from all measured data is expected at the second and third load time history representing the wheel forces in the *y* and *z* direction, because of the rather stationary load time history in all segments.

To study the load spectra growth, each load time history is divided into four segments. First segment represents the first 20 seconds of the load time history, second segment the first 40 seconds, third segment the first 60 seconds and forth segment the whole load time history (see Fig. 2). This enables modelling of the load spectra growth and comparison of the extrapolated load spectra with the actual measured load spectrum at a certain point in time. For all time segments of each load time history the cycles are counted by the Rainflow counting method according to Amzallag *et al.* (1994) and the small cycles are filtered to eliminate the noise in the measured signal. Most of the available fatigue data are obtained with testing by applying the fully reversed loading case (R = −1) where cycle mean S_M is zero. In the used load time histories, the cycle mean

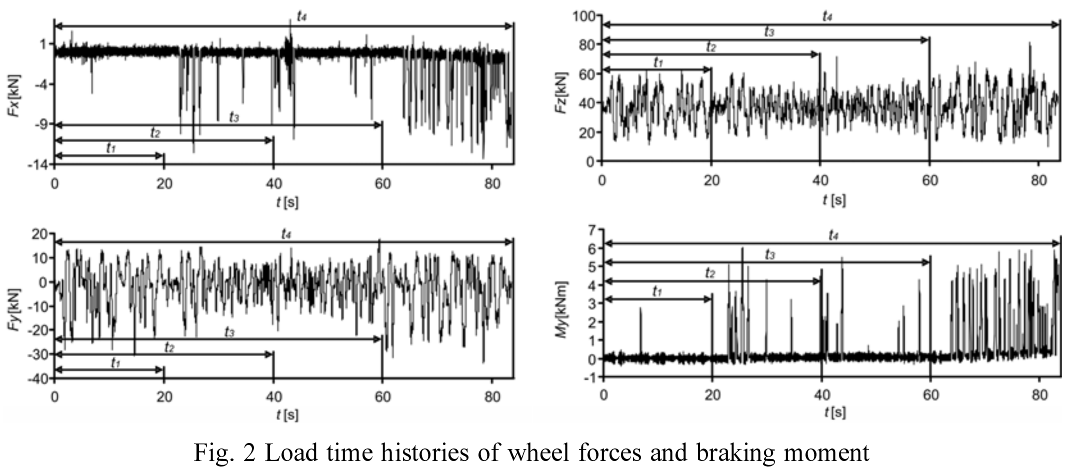

Fig. 2 Load time histories of wheel forces and braking moment

is not constant and is different from zero. Therefore equivalent force amplitudes with the cycle mean set to zero are calculated by the Goodman correction, which is the most commonly used procedure for computing equivalent load amplitudes (Farahmand *et al.* 1997). By the calculation of the equivalent force amplitudes, the effects of load cycle means on the component are considered. For each time segment of all load time histories, the equivalent force amplitude cycles are counted into histogram bins with the REBMIX algorithm. Parameters of mixed distribution functions are calculated for data counted in different number of histogram bins, ranging from 4 to 100. According to the Akaike Information Criterion (AIC) by Akaike (1974), the optimal number of histogram bins and consequently the optimal number of components in the mixed distribution function for each parametric family is selected at the lowest AIC value that is defined by the following expression:

$$AIC = 2k - 2\ln(L) \tag{10}$$

where k represents the number of independent free parameters and L the maximized value of the likelihood function for the estimated model. The AIC value can be negative if the maximized value of the likelihood function is larger than one and much larger than the number of parameters in the statistical model. Akaike Information Criterion is chosen among other available criterions described in (Figueiredo and Jain 2002, Fonseca 2008) because of its simplicity and compatibility of the results with the results from previous studies of the algorithm. At small numbers of histogram bins it turns out that the AIC value decreases rapidly by the increase of the number of bins, whereas at large numbers of histogram bins it only slightly alternates around some constant value. Therefore increasing the number of histogram bins over a certain value has no influence in terms of gaining accuracy. It only increases the number of components and the time needed to estimate the mixed distribution function. The number of histogram bins where the AIC value turns to a nearly constant value depends on the parametric family used and the load time history. Generally, for all the

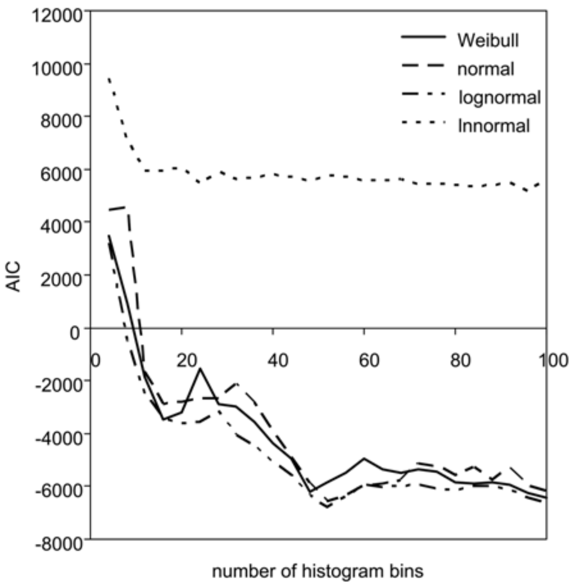

Fig. 3 Changing of AIC value with increasing the number of histogram bins for braking moment load time history

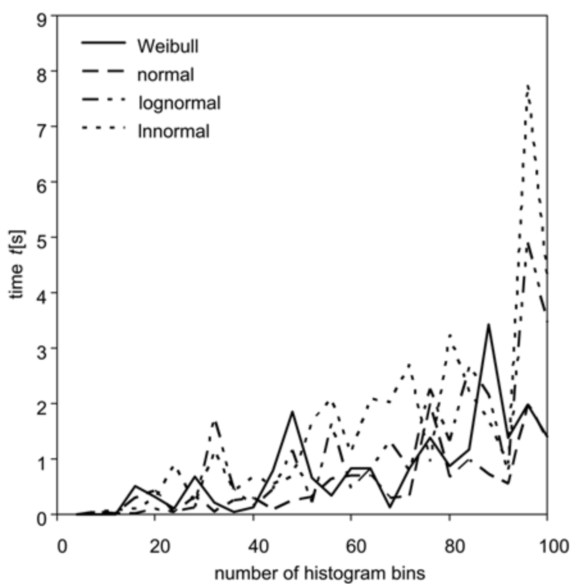

Fig. 4 Changing of estimation time with increasing the number of histogram bins for wheel forces in y direction

presented cases AIC values turn to a nearly constant value when the observations are arranged in between 10 and 50 histogram bins. Fig. 3 shows how the AIC value changes with number of histogram bins for all parametric families in the case of the whole load time history of the braking moment. It can be seen that for this kind of load time history all three parametric families have very similar AIC values stabilized after 50 bins, except at lnnormal procedure stabilization occurs much earlier, namely at around 20 bins. It turns out that for rather stationary load time histories (wheel forces in the *y* and *z* direction) AIC value for all parametric families stabilizes at around 20 or less histogram bins and the stabilization is not dependant of the length of the load time history. Increasing trend of mixed distribution parameters estimation time, for the second segment of the load time history of wheel forces in *y* direction is shown in Fig. 4. As can be seen from the figure, estimation times alternate with different number of histogram bins. This is because the observations are arranged in different number of histogram bins and consequently different number of components and their parameters are calculated with different amount of time. In general it can be seen that in the presented relatively short load history the time needed for estimation of unknown parameters increases for approximately 300% by increasing the number of histogram bins from 20 to 100. At longer load histories and larger number of bins the increment of estimation time might not be so negligible. All calculations of mixed distribution functions are made on Intel®Core™2 Duo 2, 2GHz computer with 3GB of RAM. By the lowest AIC value the optimal number of histogram bins is defined and consequently the optimal number of components too. The optimal number of components of mixed distribution functions changes only slightly with the load spectra growth for all four load time histories. In Fig. 5(a) the optimal number of components for all parametric families and for all four segments of the load time history of the wheel forces in *y* direction are presented. As shown in the figure, the number of components in mixed distribution functions for all segments is nearly constant for the Weibull and lognormal parametric family, and increases slightly for the normal parametric family and at lnnormal procedure. Similarly to the selection of the optimal number of histogram bins, it is possible to select the most suitable parametric family for the probability density estimation. The parametric family having the lowest AIC value is the most suitable one for the given load time history. Fig. 5(b) shows how the minimal

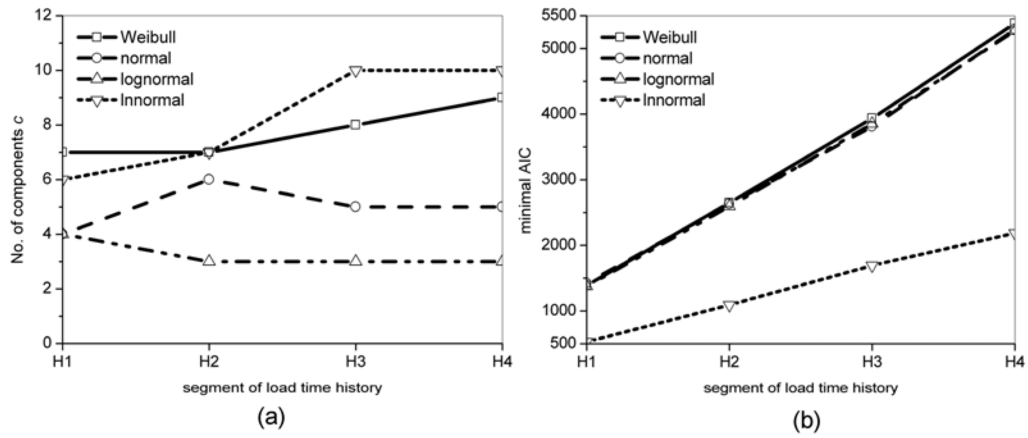

Fig. 5 Influence of load spectra growth on number of components in mixed distribution (a) and changes of minimal AIC value of mixed distribution function as the influence of load spectra growth (b) for wheel forces in *y* direction

Table 1 Changes of the number of free parameters k and maximized values of likelihood function L, for all parametric families as the influence of load spectra growth for wheel forces in y direction

	H1		H2		H3		H4	
	k	L	k	L	k	L	k	L
Weibull	20	1.1E-293	20	2.3E-567	23	4E-845	26	2.3E-1158
Normal	11	1.8E-305	17	5.8E-564	14	1.1E-821	14	5.2E-1144
Lognormal	11	3.3E-295	8	2.6E-560	8	5.9E-834	8	1.8E-1141
Lnnormal	17	2.6E-108	20	4.9E-229	29	1.2E-355	29	4.8E-463

AIC value changes with the load spectra growth for wheel forces in y direction. It turns out that for a rather stationary load time histories (wheel forces in y and z direction), the minimal AIC value increases nearly linear with the load spectra growth. For the non stationary load time histories (wheel forces in x direction and braking moment), the AIC value decreases nearly linear with the load spectra growth. As can be seen from the figure minimal AIC values are very similar for each segment of load time history for Weibull, normal and lognormal parametric family and are quite different at lnnormal procedure (see Fig. 5(b)). In presented example the most complex, lnnormal procedure has the lowest AIC values in all segments of this load time history which means that it is the most suitable for mixed distribution estimation. In this case the lnnormal procedure has much higher value of likelihood function L than other distributions while the number of free parameters k is approximately the same as at Weibull and slightly bigger than at normal and lognormal distribution which can be seen from Table 1.

Load spectrum is calculated based on Eq. (9) for each segment of each load time history. The load spectra growth for load time histories of wheel forces in x and y direction and the optimal parametric family according to the AIC value is shown in Fig. 6(a) and Fig. 7(a) by bold lines and actual load spectra of the segments by thin lines. Modelled load spectra are extrapolated from first

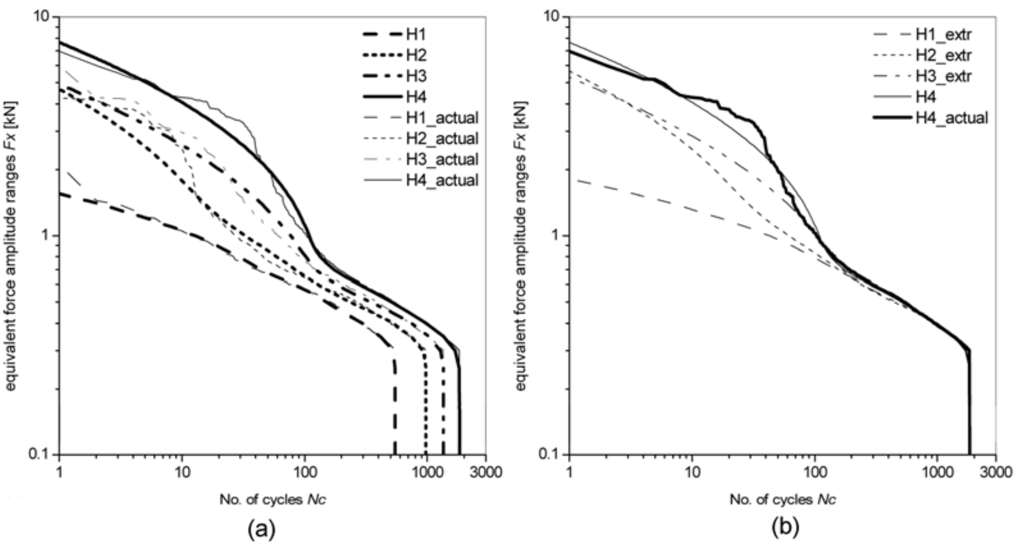

Fig. 6 Load spectra growth (a) and extrapolated load spectra (b) of wheel forces in x direction generated with mixed lognormal distributions

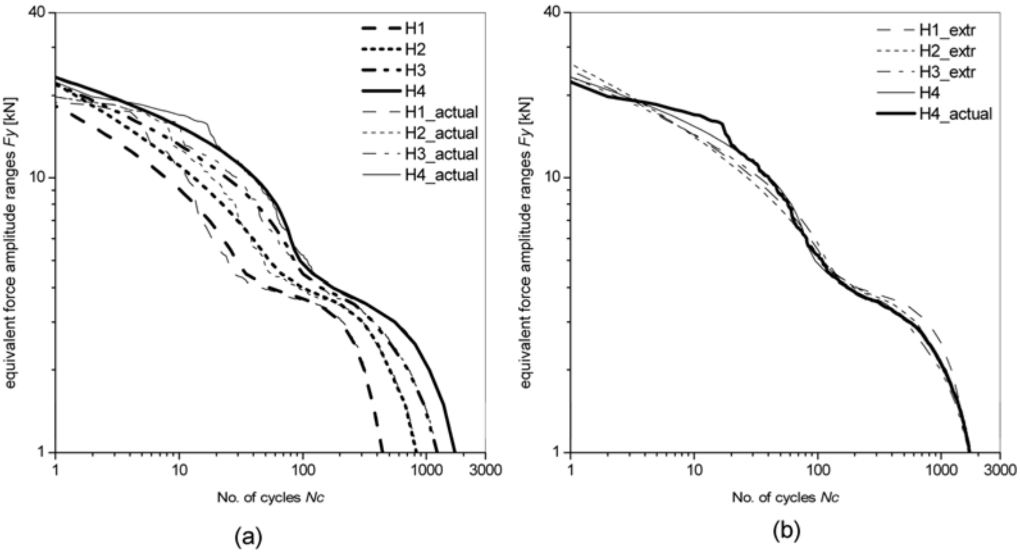

Fig. 7 Load spectra growth (a) and extrapolated load spectra (b) of wheel forces in y direction generated with mixed distributions from lnnormal procedure

three segments of load time history, trying to approach the actual load spectrum of the whole load time history as closely as possible. The extrapolated load spectra are shown in Fig. 6(b) and Fig. 7(b) by the thinner lines and the actual load spectra composed of the whole load time history by the bold line.

As expected, the extrapolated load spectra of all three segments of wheel forces in y and z direction approach much better to the actually measured load spectra from the whole load time history than the extrapolated load spectra of wheel forces in x direction and for braking moment. Small deviations between the extrapolated and actually measured load spectra for the rather stationary load time histories are present in the field of cycles with low probability of occurrence. The extrapolated load spectra of wheel forces in y direction overestimate the number of cycles in the field of the cycles with high probability of occurrence for maximal 30% for the equivalent force amplitude ranges from 2 kN to 4 kN (Fig. 7). Only at this load time history the extrapolated load spectra overestimate the number of cycles for a certain range of equivalent force amplitudes. In all other load time histories the extrapolated load spectra underestimate the number of cycles to expect at a certain point in time. Greater deviations between the extrapolated load spectra and the actually measured load spectrum of wheel forces in the x direction and the braking moment occur in the field of high equivalent force amplitudes and cycles with low probability of occurrence. This is mainly the consequence of non-equally distributed load ranges between the segments of load time history. In this case, the majority of the maximal force amplitudes appear in the last quarter of the load time history and therefore it is practically impossible to predict actual load spectrum satisfactorily from the first segment of load time history where there are practically just small force amplitudes. The extrapolated load spectra from the second and third segment of the load time history are much closer to the actual state, although the discrepancy between load spectra is still large.

3.1 Measure of assessing the agreement between the modelled and measured load spectra

To measure how well the modelled or extrapolated load spectra describe the actual load spectrum, objective measure called relative integrated absolute error (RIAE) is calculated. The RIAE measure is defined as a quotient between the integral of the absolute deviations of the extrapolated (H_{extr}) and the actual (H_{actual}) load spectrum and the integral of the actual load spectrum (H_{actual}) (Klemenc and Fajdiga 2006)

$$\mathrm{RIAE} = \frac{\displaystyle\int_{s_{\min}}^{\max\{s_{\max;extr},\, s_{\max;actual}\}} \left|\log_{10}(H_{extr}(s)) - \log_{10}(H_{actual}(s))\right| ds}{\displaystyle\int_{s_{\min}}^{s_{\max;actual}} \log_{10}(H_{actual}(s))\, ds} \tag{11}$$

The advantage of the RIAE measure over the other integral measures of agreement described in (Schimek 2000) is that it is defined as the agreement between the load spectra logarithms. The logarithms give more weight to the larger load cycles, which generally contribute more to the fatigue crack growth.

The agreement between modelled and actual or extrapolated and actual load spectra is ranked as good, average, poor and very poor. The agreement between modelled and actual load spectrum is ranked as good if RIAE $\leq$ 0.05 (e.g., H4 on Fig. 6(b)). The agreement is ranked as average if 0.05 < RIAE $\leq$ 0.15 (e.g., $H3_{extr}$ on Fig. 6(b)) or poor if 0.15 < RIAE $\leq$ 0.5 (e.g., $H1_{extr}$ and $H2_{extr}$ on Fig. 6(b)). If RIAE > 0.5 the agreement is ranked as very poor.

4. Results

The results of the agreement between modelled and actually measured load spectra from all four segments of each load time history and between extrapolated and actual load spectra from whole load time histories gained from filtered data are shown in Table 2(a). As seen from the table, the RIEA measure between modelled and measured load spectra for all segments and all load time histories is less than 0.05 which means that the agreement for all cases is ranked as good. It can also be seen that the agreement between the extrapolated load spectra and actual load spectrum from the whole load time history increases with the increasing number of observations. The smallest deviations of the extrapolated load spectra occurred at load spectra modelled from the load time history of the wheel forces in z and y direction. This is due to very good representative samples of observations. In all segments of the load time history, all load cycles that occur during the testing of the component are represented equally, which means that a small amount of observations would be enough to predict life time of the component satisfactorily, if the load state does not change. In the case of the load time history of the wheel forces in z direction, the extrapolated load spectrum modelled from the observations in the first 20 seconds deviates from the measured load spectrum of the whole load time history (83 seconds) for less than 1% which is a very good result. At extrapolated load spectra modelled from the load time history of the wheel forces in y direction (Fig. 7) the agreement between first extrapolated ($H1_{extr}$) and measured load spectrum from whole

Table 2 The RIAE measure of agreement between modelled and actually measured load spectra for all segments of load time history and between extrapolated load spectra and actually measured load spectrum of the whole load time history (a) filtered data, (b) shifted data

a	RIAE measure of filtered data				b	RIAE measure of shifted data			
	H1-model	H2-model	H3-model	H4-model		H1-model	H2-model	H3-model	H4-model
Fx	0.0035	0.0155	0.0044	0.0218	Fx	0.0810	0.0427	0.0556	0.1011
	H1-extr.	H2-extr.	H3-extr.			H1-extr.	H2-extr.	H3-extr.	
	0.1885	0.1557	0.0980			0.1196	0.1825	0.1454	
	H1-model	H2-model	H3-model	H4-model		H1-model	H2-model	H3-model	H4-model
Fy	0.0157	0.0094	0.0056	0.0003	Fy	0.0261	0.0150	0.0049	0.0090
	H1-extr.	H2-extr.	H3-extr.			H1-extr.	H2-extr.	H3-extr.	
	0.1236	0.0038	0.0353			0.1145	0.0026	0.0255	
	H1-model	H2-model	H3-model	H4-model		H1-model	H2-model	H3-model	H4-model
Fz	0.0026	0.0026	0.0040	0.0067	Fz	0.0025	0.0042	0.0116	0.0170
	H1-extr.	H2-extr.	H3-extr.			H1-extr.	H2-extr.	H3-extr.	
	0.0091	0.0148	0.0150			0.0090	0.0160	0.0225	
	H1-model	H2-model	H3-model	H4-model		H1-model	H2-model	H3-model	H4-model
My	0.0340	0.0108	0.0409	0.0062	My	0.0037	0.2013	0.1686	0.0309
	H1-extr.	H2-extr.	H3-extr.			H1-extr.	H2-extr.	H3-extr.	
	0.6388	0.4063	0.4280			0.6551	0.5235	0.5053	

load time history is ranked as average, all other agreements between extrapolated load spectra and measured load spectrum are ranked as good. The largest deviation between the extrapolated and actually measured load spectrum occurred at the load spectra composed of load time history of the braking moment. The deviations in this case are in the range from 43% to 63% for all three

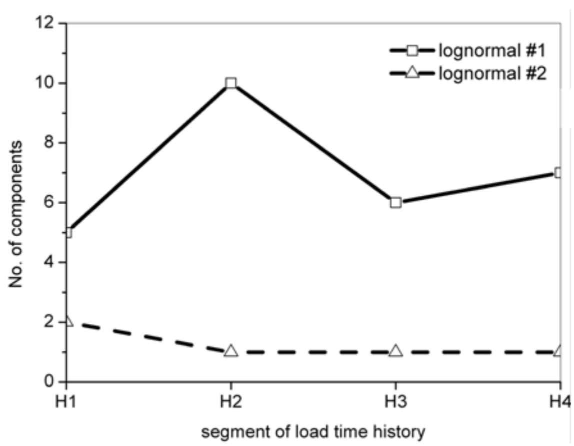

Fig. 8 Reduction of the optimal number of components in mixed distribution with shifting the sample data for filter size in the case of wheel forces in *x* direction

extrapolated load spectra. Therefore the agreement between extrapolated load spectra and measured load spectrum is ranked as poor and very poor. This poor and very poor agreement is the consequence of highly unequally represented load cycles in the segments of the load history. In this case it is impossible to reliably predict life time of the component from a small sample of observations. In the presented example the deviation between the extrapolated load spectrum modelled from the observations in the first 60 seconds and the measured load spectrum of the whole load time history is still over 40%.

Additionally, the influence of filtering on the modelling and extrapolation of load spectra is studied. Instead of just removing small load cycles from the load cycle history, all counted cycles are decreased for the filter size for each load time history by substituting the variable s in Eq. (3) to Eq. (8) with variable $(s - \eta)$, where η is a known parameter and represents the filter size. The procedure of modelling and extrapolation of load spectra is the same as shown earlier. At the end each load spectrum is shifted up for the filter size and compared with the corresponding measured load spectrum. By this kind of filtering, the load cycle data is moved towards the abscissa axis, thus making it much simpler for the algorithm to recognize the mixed distribution function. It turns out that in this case the algorithm estimates the unknown parameters of mixed distribution function on average from 10 to 80 times faster for the same parametric family and the same load time history. Optimal parametric families according to the AIC values for all load time histories stayed the same as in the previous examples, although the AIC values are a little lower for all three parametric families and higher in all cases at the lnnormal procedure. The AIC value decrease to a nearly constant value a little sooner, which means that the observations are arranged in a smaller optimal number of histogram bins (in the presented cases between 10 and 40 histogram bins) and consequently the mixed distribution functions are composed of fewer components. The reduction of the components in the mixed distribution functions is from 1 to 9 components for all presented examples. The maximal reduction occurs at the lognormal parametric family and the minimal reduction at the lnnormal procedure. In Fig. 8 it is shown how the number of components in the mixed distribution function is reduced for all segments of the load time history of the wheel forces in x direction and for lognormal parametric family which is the optimal parametric family for this load time history.

The results of the agreement between modelled and actually measured load spectra from all four segments of each load time history and between extrapolated and actual load spectra from whole load time histories gained from shifted data are shown in Table 2(b).

It turns out that for rather stationary load time histories RIAE measure between modelled and actually measured load spectra for all segments of load histories and between extrapolated load spectra and actual load spectrum from the whole load time history differs for maximal 1% as in the previous examples. Therefore the agreement is still ranked as good in all the cases except for the extrapolated load spectrum from the first segment ($H1_{extr}$) of wheel forces in y direction, where the agreement is slightly better than before but still ranked as average. For non stationary load time histories the agreement between modelled and actually measured load spectra for all segments of load histories is between 2.5% and 19% worse than before in all cases except in the first segment of the braking moment where the agreement is a little better. The agreement between extrapolated load spectra and actual load spectrum from whole load history is from 1.6% to 11.7% worse in all cases except for the extrapolated load spectrum from first segment of load history of wheel forces in x direction (see Table 2(b)). The reason for these bad results lies in the predicted optimal number of components of mixed distribution functions which is too small to gain a good approximation of mixed distribution function. With this kind of filtering of the sample data, the number of components in mixed distribution functions is smaller than in the previous examples (see Fig. 8). This means that for short non stationary load histories REBMIX algorithm gives better agreement of the mixed distribution functions if the sample data is not shifted. Like in the previous examples, the largest deviations between the extrapolated load spectra and the actually measured load spectrum occur at the load spectra of the braking moment, and are between 50% and 65% according to the RIAE measure which is ranked as very poor.

5. Conclusions

The basic ideas in this article are the investigation of load spectra growth and the evaluation of modelling and extrapolation of load spectra with REBMIX from actually measured data using different mixture models. It turns out that for all mixture models it has no effect on increasing the number of histogram bins, in which the observations are arranged in, over a certain limit since the accuracy of the approximation remains practically the same but the calculation time rises by increasing the number of bins. In the presented examples the limit depends on the shape of load time history and the parametric family used and is between 10 and 50 bins. It is shown that the number of components in the mixed distribution do not depend so much on the number of observations as it slightly changes with increasing the number of observations in the load time history. On the other hand, the number of components is more dependent of the parametric family used in the mixture distribution estimation. It can be seen from Fig. 5 that parametric families with almost the same AIC value can have a very different number of components in mixed distributions. The difference between the numbers of components in the presented cases is up to 200%. It can be concluded that the agreement part ($2*ln(L)$) of Eq. (10) influences more to the final value of AIC than the complexity part ($2*k$). This means that one should also take into consideration not only the AIC value but also the parametric family used and consequently the number of components, especially at large number of observations as it can reduce the estimation time considerably.

When a small number of histogram bins is used to approximate lognormal data, it is better to select the lnnormal procedure instead of the lognormal parametric family because the histogram bins can be condensed in the area with the largest number of observations and consequently better approximation of the distribution is gained. At the large number of histogram bins, the lognormal parametric family seems a better choice since the approximation with the lnnormal procedure becomes very slow by increasing the number of bins. As seen in Fig. 4, the increase of the estimation time is the fastest at the lnnormal procedure.

According to the randomness of actually measured data used for the modelling of load spectra it was expected that the extrapolated load spectra will not coincide well with the actually measured load spectrum from the whole load data. The expectations came true for the load data of the wheel forces in x direction and braking moment since the segments that the load spectra is modelled from, have very unequally distributed load ranges. For good agreement between the extrapolated load spectrum and the actually measured load spectrum it is necessary, that all possible load ranges are represented in the sample that the load spectrum is modelled from and extrapolated to any point in time. Good extrapolation of load spectrum also depends on the choice of the parametric family. However, it turns out that good representative sample of loads is much more important than the parametric family. As can be seen from the presented examples, the deviation between the extrapolated and the actually measured load spectrum can be very small even if the extrapolated load spectrum is modelled from a small amount of observations, if the sample is a good representation of the actual loads occurring during the operation of the component. On the other hand, if the representative sample is not defined well, the deviation between the extrapolated and the actually measured load spectrum can be large even if the extrapolated load spectrum is modelled from a large amount of observations. In the cases of highly unpredictable load cycles, it is necessary to constantly supervise the load spectra growth and extrapolate the modelled load spectrum in the next point in time. In general, the treated load time histories show that the extrapolated load spectra coincide very well with the actually measured load spectrum in the field of the cycles with high probability of occurrence. Larger discrepancies are detected in the field of larger equivalent force amplitude ranges. In all cases the algorithm underestimates the loads with high amplitudes and low probability of occurrence, which generally contribute more to fatigue crack growth. On the basis of these observations one should pay special attention to the extrapolated load spectrum in the field of the cycles of higher amplitudes at predicting the reliability of the component. In the fatigue damage calculation the material SN curve should also be taken into consideration since the fatigue damage caused by numerous small amplitude cycles or very few large amplitude cycles might be small compared to that produced by medium amplitude cycles (Socie 2001). All presented extrapolated load spectra underestimates load cycles with medium equivalent amplitudes. Underestimation is small for rather stationary load time histories and not so negligible for non stationary load time histories.

By different methods of treating the small load cycles it turns out that the deviation between the extrapolated and actually measured load spectrum is up to 12% different. Deviation between modelled and measured load spectrum from each segment of load time history is up to 19% different. If all the data are shifted for the filter size instead of just removing small cycles, the number of components in the mixture distribution decreases and the time needed to estimate the unknown parameters decreases rapidly. On the other hand, if the data that the load spectrum is modelled from is not shifted, the agreement between modelled and actually measured load spectrum from each segment of load history is better in almost all cases. Agreement between extrapolated and actually measured load spectra for shifted data is slightly better for all extrapolated load spectra of wheel forces in y direction and for first extrapolated load spectrum of wheel forces in x direction. In all other cases the agreement between extrapolated and actually measured load spectra is worse than in the case when observations are not shifted.

To all interested readers, the software programme REBMIX is available on the http://www.fs.uni-lj.si/lavek/.

References

Akaike, H. (1974), "A new look at statistical model identification", *IEEE* T. Automat. Contr., **19**, 716-723.

Amzallag, C., Gerey, J.P., Robert, J.L. and Bahuaud, J. (1994), "Standardization of the rainflow counting method for fatigue analysis", *Int. J. Fatigue*, **16**(4), 287-293.

Balakrishnan, N. and David, H.A. (2001), "A note on the variance of a lightly trimmed mean when multiple outliers are present in the sample", *Stat. Probabil. Lett.*, **55**(4), 339-343.

Bučar, T. (2006), *Structural Reliability Modelling Depending on Operation Conditions.* Thesis (PhD). University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana.

Buxbaum, O. and Zaschel, J.M. (1979), *Beschreibung Stochasticher Beanspruchungs-Zeit-Funktionen. Verhalten* *von Stahl bei schwingender Beanspruchung. Kontaktstudium Werkstoffkunde Eisen und Stahl III. Verein Deutscher Eisenhüttenleute.* Düsseldorf, 208-222.

Ebeling, C.E. (1997), *An Introduction to Reliability and Maintainability Engineering,* McGraw-Hill, New York.

Farahmand, B., Bockrath, G. and Glassco, J. (1997), *Fatigue and Fracture Mechanics of High Risk Parts,* Chapman & Hall, New York.

Figueiredo, M.A.T. and Jain, A.K. (2002), "Unsupervised learning of finite mixture models", *IEEE T. Pattern Anal.,* **24**(3), 381-396.

Fonseca, J.R.S. (2008), "The application of mixture modeling and information criteria for discovering patterns of coronary heart disease", *J. Appl. Quantitative Meth.*, **3**(4), 292-303.

Grubisic, V. and Fischer, G. (1997), "Methodology for effective design evaluation and durability approval of car suspension components" *SAE technical paper series 970094.*

Johannesson, P. (2006), "Extrapolation of load histories and spectra", Fatigue Fract. Eng. M., **29**(3), 201-207.

Klemenc, J. and Fajdiga, M. (2006), "Predicting smoothed loading spectra using a combined multilayer perceptron neural network", *Int. J. Fatigue*, **28**(7), 777-791.

Nagode, M. and Fajdiga, M. (1998a), "A general multi-modal probability density function suitable for the rainflow ranges of stationary random processes", *Int. J. Fatigue*, **20**(3), 211-223.

Nagode, M. and Fajdiga, M. (1998b), "On a new method for prediction of the scatter of loading spectra", *Int. J. Fatigue*, **20**(4), 271-277.

Nagode, M. and Fajdiga, M. (1999), "The influence of variable operating conditions upon the general multi-modal Weibull distribution", *Reliab. Eng. Syst. Safe*, **64**(3), 383-389.

Nagode, M. and Fajdiga, M. (2000), "An improved algorithm for parameter estimation suitable for mixed Weibull distributions", *Int. J. Fatigue*, **22**(1), 75-80.

Nagode, M., Klemenc, J. and Fajdiga, M. (2001), "Parametric modelling and scatter prediction of rainflow matrices", *Int. J. Fatigue*, **23**(6), 525-532.

Nagode, M. and Fajdiga, M. (2006), "An alternative perspective on the mixture estimation problem", *Reliab. Eng. Syst. Safe*, **91**(4), 388-397.

Schimek, M.G. (2000), *Smoothing and Regression: Approaches, Computation and Application,* Wiley, New York.

Socie, D. (2001), "Modelling expected service usage from short-term loading measurements", *Int. J. Mater. Prod. Tech.*, **16**(4), 295-303.

Stehlík, M. (2006), "Exact likelihood ratio scale and homogeneity testing of some loss processes", *Stat. Probabil. Lett.* , **76**(1), 19-26.

Stehlík, M. (2008), "Homogeneity and scale testing of generalized gamma distribution", *Reliab. Eng. Syst. Safe*, **93**(12), 1809-1813.

Tovo, R. (2000), "A damage-based evaluation of probability density distribution for rain-flow ranges from random processes", *Int. J. Fatigue*, **22**(5), 425-429.

Zhao, W. and Baker, M.J. (1992), "On the probability density function of rainflow stress range for stationary Gaussian processes", *Int. J. Fatigue*, **14**(2), 121-135.