Extrapolation of Rainflow Matrices

Pär Johannesson (parj@math.chalmers.se)
Mathematical Statistics, Chalmers University of Technology, 412 96 Göteborg, Sweden

Jean-Jacques Thomas (thomas13@mpsa.com)
PSA Peugeot Citröen, Centre SAMM – Chemin de la Malmaison, 91570 Bièvre, France

Abstract. Fatigue loads are often measured for a limited period of time, but are used for making fatigue life predictions for the full lifetime. The measured load is reduced by means of rainflow cycle counting, and stored as a 3-dimensional histogram, called the rainflow matrix (RFM). Hence, there is a need to extrapolate the measured RFM to one representing the full lifetime. How to make the extrapolation is addressed in this paper.

The simplest method is to repeat the measurement until the full lifetime is reached, say N times. This is equivalent to multiplying the measured RFM by N. The drawback is that only the measured rainflow cycles appear, even though other cycles are also possible. Especially, measured infrequent load events are multiplied. The solution is to use the limiting RFM, i.e. the shape of the RFM that is obtained when having an infinitely long measurement. Hence, the extrapolation becomes N times the limiting RFM.

This paper treats the problem of estimating the limiting RFM from a measured RFM. The main result is an asymptotic expression for the limiting RFM, that only involves the level crossing intensity. The main assumption is Poisson convergence of level crossings. It gives a good approximation that is valid for low minima and for high maxima, i.e. for cycles with large amplitudes. The level crossing intensity is extrapolated for high and for low levels, by using the generalized Pareto distribution. Since the formula is only valid for large amplitudes, the remaining part has to be estimated by other means, e.g. by using a kernel smoother. The combination becomes the final estimate of the limiting RFM. Full details about the estimation procedure is presented. The method is verified through examples for both Gaussian and Markov loads, and test track measurements are also analysed.

Keywords: fatigue, random loads, rainflow cycles, interval crossings, generalized Pareto distribution

1. Introduction

When analyzing complex fatigue loads in the automotive industry one often uses simple damage and accumulation rules, together with the rainflow cycle counting method. The SN-curve is a model for the fatigue life, originally for constant amplitude loads, where at a stress level S_i the specimen can resist N_i cycles. The damage accumulation hypothesis due to Palmgren-Miner [17, 16] is that each cycle with amplitude S_i uses a fraction $1/N_i$ of the total lifetime. Thus the total fatigue damage is given by

$$D = \sum \frac{n_i}{N_i} \tag{1}$$

where n_i is the number of cycles with the amplitude S_i. A fatigue failure occurs when the whole lifetime is consumed, i.e. when the damage D exceeds one.

For extracting cycles in variable amplitude loads, the rainflow cycle method is widely used. The method was first presented by Endo in 1967, see [15], and counts closed hysteresis loops in the load. An equivalent local definition of rainflow cycles was given by Rychlik [19], see Figure 1. From this definition it is possible to establish the equivalence between counting rainflow cycles and counting crossings of intervals. This fact is very useful when making statistical analyses of loads.

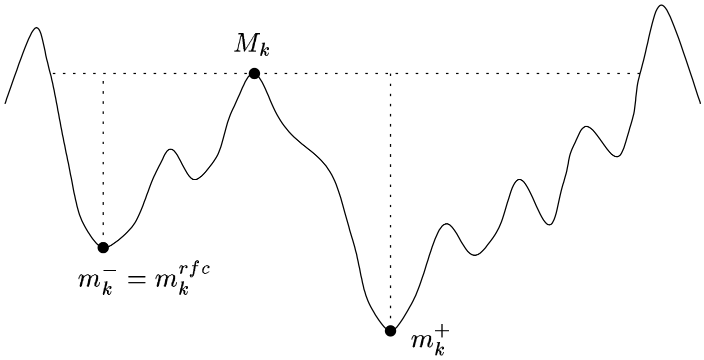

Figure 1. Definition of rainflow cycles by Rychlik [19]. From each local maximum M_k one shall try to reach above the same level, in the backward(left) and forward(right) directions, with as small a downward excursion as possible. The minimum, of m_k^- and m_k^+, which represents the smallest deviation from the maximum M_k is defined as the corresponding rainflow minimum m_k^{rfc}. The rainflow cycle is defined as $(m_k^{\mathrm{rfc}}, M_k)$. For the precise definition see Rychlik [19].

When registering load histories the signal is often discretized to a fixed number of load levels, which enables efficient storage of the rainflow cycles in the form of a matrix. Each element of the rainflow matrix (RFM) then contains the number of cycles for that class of cycles.

Making field measurements of vehicle fatigue loads are a time-consuming and costly procedure, and is therefore only performed for a limited period of time, say for a distance of x km. These measurements are then used for making predictions on the total lifetime of the vehicle, say y km. Hence, there is a need for methods to extrapolate the measured rainflow matrix $\boldsymbol{F}_x$ representing x km to a rainflow matrix $\boldsymbol{F}_y$ representing y km. The simplest way of doing the extrapolation is to rescale the measured rainflow matrix to y km, and hence multiply it by y/x, giving $\boldsymbol{F}_y = \frac{y}{x}\boldsymbol{F}_x$. The drawback of this approach is that only the measured cycles will appear in $\boldsymbol{F}_y$, even though also other cycles will appear in a longer measurement. Especially, the rare load cycles that occurred in the measurement will be multiplied by y/x, while other rare cycles will not be possible.

Another approach is to look for the limiting shape of the RFM, and use this matrix for extrapolation to y km. In our application the limiting RFM, $\boldsymbol{G}$, say, should be understood as the RFM, normalized by the number of km, one would get when measuring for an infinitely long distance, i.e. $\boldsymbol{G} = \lim_{x\to\infty} \frac{1}{x}\boldsymbol{F}_x$. There are several possible techniques for estimating $\boldsymbol{G}$ from the measured RFM $\boldsymbol{F}_x$:

1. Estimate a model for the load, and calculate the limiting shape of the RFM for this model.
2. Estimate the limiting shape of the RFM by applying some smoothing technique on $\boldsymbol{F}_x$.
3. Model only the RFM for large cycles, since they cause most of the damage. If necessary, model the remaining part of the RFM by some other technique from 1 or 2.

To estimate a load model from a measured RFM is a hard problem. Further, to calculate the limiting RFM is only possible for some models, e.g. for loads with Markov structure (see e.g. Johannesson [11, 12] and Example 2). For Gaussian loads there exist accurate approximations (see Rychlik [20, 21] and Example 1). Point 1 will not be pursued in this paper.

Concerning point 2, a feasible smoothing technique is the so called kernel smoothing, which is described in Section 5. This is a well established statistical method for non-parametric estimation of an unknown function. A study of smoothing RFMs is found in Dressler et al. [6].

The estimate given by the kernel smoother is accurate where there are many observations, but can perform poorly otherwise. Often, a large amount of the fatigue damage is due to the rare large cycles, hence it is motivated to find methods in category 3. The main result in this paper is a method for finding the limiting shape of the most extreme part of rainflow matrix, i.e. for cycles with low minima and large maxima. The approach is based on asymptotic theory for crossings of extreme (high and low) levels, and uses the connection between rainflow cycles and crossings of intervals. The method gives an explicit asymptotic expression for the intensity of interval crossings, which is the same as the cumulative rainflow matrix. The resulting formula for the extreme RFM only depends on the intensity of level crossings. When analysing a measured RFM, we propose to estimate the intensity of level crossings from the observed one, by extrapolation for high and low levels. The extrapolation is based on methods from the extreme value theory, which are adapted to our problem. The generalized Pareto distribution is used when extrapolating the level crossing intensity.

The results will be presented by numerical examples. In Example 1 the extreme RFM is calculated for a Gaussian load. A Markov load is used to demonstrate and verify the proposed methods, see Examples 2–6. Measured loads from a test track is analysed in Example 7. For the calculation Matlab has been used together with the toolbox *Wave Analysis for Fatigue and Oceanography* (WAFO), with manual [27]. An overview of WAFO is found in Brodtkorb et al. [2]. The toolbox is free and can be downloaded from the Internet[1].

2. Limiting Rainflow Matrix for Large Cycles

The goal is to find an expression for the limiting RFM that is valid for cycles with high maxima and low minima. To obtain the result, the equivalence between counting rainflow cycles and counting crossings of intervals are used. The main assumption is that the times of upcrossings of high and of low levels converge to two independent Poisson processes. The method of approximating extreme level crossings by Poisson processes when calculating the rainflow distribution has also been suggested in Rychlik [22, 21]. The Poisson convergence of level crossings is valid for stationary processes with continuous sample paths, that are subjected to suitable conditions of 'mixing type', and especially, it is valid for stationary Gaussian processes with finite second spectral moment, see Leadbetter et al. [13] for details. The resulting formula for the limiting cumulative rainflow matrix is

$$\mu^{\text{rfc}}(u,v) \approx \frac{\mu(u)\mu(v)}{\mu(u)+\mu(v)} \tag{2}$$

[1] `http://www.maths.lth.se/matstat/wafo/`

where the approximation is supposed to be good when u is a low level and v is a high level. Note that Eq. (2) only depends on the intensity of level upcrossings $\mu(u)$, which is a function of u.

Next we will give some definitions of cycle counts and of intensity of crossings. Let $N_T(u)$ be the number of upcrossings of the level u by the process $\{X(t)\}$ in time $0 < t < T$, and define the intensity of upcrossings as

$$\mu(u) \;=\; \lim_{T\to\infty} \frac{\mu_T(u)}{T} \qquad \text{with} \qquad \mu_T(u) = \mathbf{E}[N_T(u)] \tag{3}$$

assuming the limit to exist. (For ergodic processes the limit exists.) Let $N_T(u, v)$ be the number of upcrossings of the closed interval $[u, v]$ by $\{X(t)\}$ in time $0 < t < T$, and define the corresponding intensity as

$$\mu(u,v) \;=\; \lim_{T\to\infty} \frac{\mu_T(u,v)}{T} \qquad \text{with} \qquad \mu_T(u,v) = \mathbf{E}[N_T(u,v)] \tag{4}$$

assuming that the limit exists. Further, define the cumulative rainflow count

$$N_T^{\text{rfc}}(u,v) = \#\{(m_k^{\text{rfc}}, M_k) : m_k^{\text{rfc}} < u,\ M_k > v\}, \tag{5}$$

where $\#\{\cdot\}$ denotes the number of elements in the set $\{\cdot\}$, and define the cumulative rainflow (count) intensity

$$\mu^{\text{rfc}}(u,v) = \lim_{T\to\infty} \frac{\mu_T^{\text{rfc}}(u,v)}{T} \qquad \text{with} \qquad \mu_T^{\text{rfc}}(u,v) = \mathbf{E}[N_T^{\text{rfc}}(u,v)] \tag{6}$$

assuming that the limit exists.

There is an interesting and important connection between rainflow cycles and crossings of intervals. Assume that the process is at a local maximum at level v. The definition of rainflow cycles (Figure 1) states that in order to have a rainflow cycle with minimum less than u and maximum v, the process has to reach below level u, before it upcrosses v. This has to hold in both the forward and the backward time directions. Hence, to have rainflow cycles with minimum less than u and maximum bigger than v, is equivalent to the event of upcrossing the closed interval $[u, v]$, and thus we have the relation

$$N_T^{\text{rfc}}(u,v) \;=\; N_T(u,v) = \#\left\{ \begin{array}{l} \text{upcrossings of the closed} \\ \text{interval } [u,v] \text{ for } X(t),\ t \in [0,T] \end{array} \right\}. \tag{7}$$

A mathematically precise proof of this statement is found in Rychlik [23, Lemma 7]. Consequently, for an ergodic process $\{X(t)\}_{t\geq 0}$ the relation between the cumulative rainflow intensity and the intensity of upcrossings of $[u, v]$ is

$$\mu^{\text{rfc}}(u,v) \;=\; \lim_{T\to\infty} \frac{N_T^{\text{rfc}}(u,v)}{T} = \lim_{T\to\infty} \frac{N_T(u,v)}{T} = \mu(u,v) \tag{8}$$

The result in this section is based on Poisson convergence of level crossings. Define the time-normalized point processes of upcrossings

$$\begin{aligned} U_{\hat{T}}^*(B) &\;=\; \#\{u\text{-upcrossings by } X(t);\ t/\hat{T} \in B\} \\ V_{\hat{T}}^*(B) &\;=\; \#\{v\text{-upcrossings by } X(t);\ t/\hat{T} \in B\} \end{aligned}$$

for any Borel set B. Consider $u = u_{\hat{T}} \to -\infty$ and $v = v_{\hat{T}} \to \infty$, when $\hat{T} \to \infty$, such that

$$\begin{array}{ll} \hat{T}\mu(u_{\hat{T}}) \to \lambda_u, & u_{\hat{T}} \to -\infty \\ \hat{T}\mu(v_{\hat{T}}) \to \lambda_v, & v_{\hat{T}} \to \infty \end{array} \tag{9}$$

where $\lambda_u > 0$ and $\lambda_v > 0$ are fixed numbers. Further, let $\hat{U}(B)$ and $\hat{V}(B)$ be two independent Poisson processes with intensities λ_u and λ_v, respectively.

Now we are ready to state the result.

THEOREM 1. *Suppose that the process $\{X(t)\}$ is stationary and ergodic, and has continuously differentiable sample paths. Further suppose that $(U^*_{\hat{T}}, V^*_{\hat{T}})$ converges in distribution to the independent Poisson processes $(\hat{U}, \hat{V})$, if Eq. (9) holds when $\hat{T} \to \infty$. Then the cumulative rainflow count intensity satisfies*

$$\frac{\mu^{\text{rfc}}(u,v)}{\dfrac{\mu(u)\mu(v)}{\mu(u)+\mu(v)}} \quad \to \quad 1 \tag{10}$$

when $u \to -\infty$, $v \to +\infty$.

Proof: Since $\{X(t)\}$ is stationary and ergodic, the cumulative rainflow count intensity can, according to Eq. (8), be written as

$$\begin{aligned} \mu^{\text{rfc}}(u,v) &= \lim_{T\to\infty} \frac{1}{T} N_T^{\text{rfc}}(u,v) = \lim_{T\to\infty} \frac{1}{T} N_T(u,v) \\ &= \lim_{T\to\infty} \frac{1}{T} N_T(u) \frac{N_T(u,v)}{N_T(u)} \\ &= \mu(u)\mathbf{P}\left(\left\{ \begin{array}{l} t=0 \text{ is a } u\text{-upcrossing of } X(t), \\ \text{and } \exists r>0 \text{ such that } X(r)>v \\ \text{and } u \le X(s) \le v,\ 0<s<r \end{array} \right\} \middle| \left\{ \begin{array}{l} t=0 \text{ is a} \\ u\text{-upcrossing} \\ \text{of } X(t) \end{array} \right\} \right) \end{aligned}$$

where the conditional probability should be understood as a Palm measure. It means that after an upcrossing of u, the process shall upcross v before it upcrosses u again. This can also easily be described in terms of the point processes $U^*_{\hat{T}}$ and $V^*_{\hat{T}}$. It means that, after $t = 0$, the next point in the processes shall appear in $V^*_{\hat{T}}$, which can be expressed as

$$\mu^{\text{rfc}}(u,v) \;=\; \mu(u)\mathbf{P}(\tau_v < \tau_u \,|\, U_0)$$

where τ_u and τ_v, respectively, denote the waiting times to the first point after $t = 0$ in $U^*_{\hat{T}}$ and $V^*_{\hat{T}}$, respectively, and U_0 denotes the event "$U^*_{\hat{T}}$ has a point at $t = 0$".

Next, we shall use the Poisson convergence. Since $U^*_{\hat{T}}$ and $V^*_{\hat{T}}$ converges to two independent Poisson processes, the waiting times τ_v and τ_u converges in distribution to two independent exponential variables, $\hat{\tau}_v$ and $\hat{\tau}_u$, say, with the means $1/\lambda_v$ and $1/\lambda_u$, respectively. Further, note that by assumption Eq. (9), $\mu(u_{\hat{T}})/\mu(v_{\hat{T}}) \to \lambda_u/\lambda_v$ when $\hat{T} \to \infty$. Now we get

$$\frac{\mu^{\text{rfc}}(u,v)}{\dfrac{\mu(u)\mu(v)}{\mu(u)+\mu(v)}} = \frac{\mu(u)\mathbf{P}(\tau_v < \tau_u \,|\, U_0)}{\dfrac{\mu(u)\mu(v)}{\mu(u)+\mu(v)}}$$

$$
\begin{aligned}
&= \frac{\mu(u)+\mu(v)}{\mu(v)} \mathbf{P}(\tau_v < \tau_u \,|\, U_0) \\
&= \left(\frac{\mu(u)}{\mu(v)} + 1\right) \mathbf{P}(\tau_v < \tau_u \,|\, U_0) \\
&\to \left(\frac{\lambda_u}{\lambda_v} + 1\right) \mathbf{P}(\hat{\tau}_v < \hat{\tau}_u \,|\, \hat{U}_0) \\
&= \frac{\lambda_u + \lambda_v}{\lambda_v} \cdot \mathbf{P}(\hat{\tau}_v < \hat{\tau}_u)
\end{aligned}
$$

where we used the lack of memory property of the Poisson process, when dropping the conditioning on $\hat{U}_0$, denoting "$\hat{U}$ has a point at $t=0$". Since $\hat{\tau}_v$ and $\hat{\tau}_u$ are independent, the probability $\mathbf{P}(\hat{\tau}_v < \hat{\tau}_u)$ is easy to calculate, giving

$$
\frac{\mu^{\text{rfc}}(u,v)}{\frac{\mu(u)\mu(v)}{\mu(u)+\mu(v)}} \to \frac{\lambda_u+\lambda_v}{\lambda_v} \cdot \frac{\lambda_v}{\lambda_u+\lambda_v} = 1
$$

concluding the proof. □

The practical importance of Theorem 1 is that for cycles with high maxima and low minima, i.e. cycles with large amplitudes, Eq. (2) gives an approximation of the rainflow intensity. In the following examples we shall examine how well the approximation works in practice.

EXAMPLE 1 (Extreme rainflow matrix for a Gaussian load). In many applications a frequency spectrum is given and a Gaussian process is used, e.g. for modelling stationary sea states, and loads. Unfortunately, no explicit expression for its rainflow intensity exists. However, many approximations have been suggested, and here we will compare Theorem 1 with one such approximation.

Consider a zero-mean Gaussian process $X(t)$ with $\mathbf{V}[X(t)] = 1$ and $\mathbf{V}[X'(t)] = 1$. The level crossing intensity is given by Rice's formula

$$
\mu(u) = \frac{1}{2\pi} \exp(-u^2/2). \tag{11}
$$

Hence, by Theorem 1 we get an approximation of the cumulative rainflow intensity

$$
\mu^{\text{rfc}}_{\text{app}}(u,v) = \frac{\mu(u)\mu(v)}{\mu(u)+\mu(v)} = \frac{1}{2\pi(\exp(u^2/2)+\exp(v^2/2))} \tag{12}
$$

and the rainflow intensity is obtained by differentiating with respect to u and v

$$
\lambda^{\text{rfc}}_{\text{app}}(u,v) = -\frac{\partial^2}{\partial u \partial v} \mu^{\text{rfc}}_{\text{app}}(u,v) = -\frac{uv\exp((u^2+v^2)/2)}{\pi(\exp(u^2/2)+\exp(v^2/2))^3} \tag{13}
$$

and is valid for $u \ll 0$ and $v \gg 0$.

An accurate approximation of the rainflow intensity is obtained by approximating the sequence of local extremes by a Markov chain (see Example 2). The transition probabilities are obtained from the min-max probability density function, which is computed from the spectral density by using a Slepian model process and regression approximation, see Lindgren & Rychlik [14] or Rychlik [20, 21]. A broad band spectrum with irregularity factor 0.56 was used.

The two approximations are compared in Figure 2, where we can see that they are almost identical in the upper left corner of the graph, i.e. for cycles with large amplitudes. Hence, the simple explicit expression Eq. (13) works very well. □

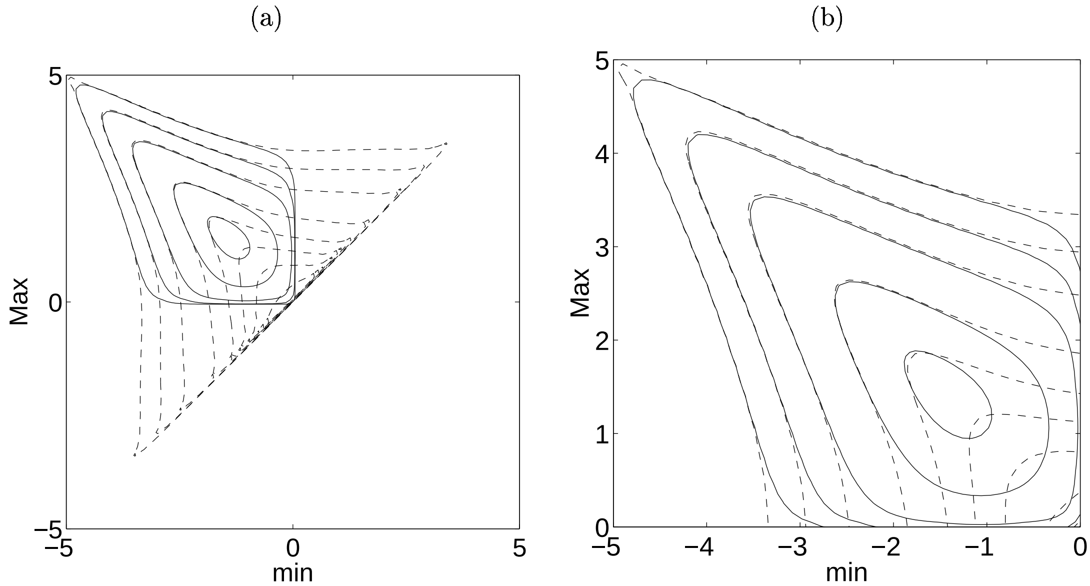

Figure 2. Example 1, Gaussian load. Iso-lines of rainflow matrices: Extreme RFM (−) given by Eq. (13), compared with Markov approximation (− −); (b) shows a magnification of the upper left part of the RFM.

For discrete loads the cumulative rainflow matrix is the same as the cumulative rainflow count. The relation between the cumulative RFM, $\boldsymbol{N}^{\mathrm{rfc}}$, and the RFM, $\boldsymbol{F}^{\mathrm{rfc}}$, is

$$\boldsymbol{F}^{\mathrm{rfc}} = \left(f_{ij}^{\mathrm{rfc}}\right)_{i,j=1}^{n} \qquad \text{with} \qquad f_{ij}^{\mathrm{rfc}} = n_{i+1,j-1}^{\mathrm{rfc}} - n_{i,j-1}^{\mathrm{rfc}} - n_{i+1,j}^{\mathrm{rfc}} + n_{i,j}^{\mathrm{rfc}} \tag{14}$$

and

$$\boldsymbol{N}^{\mathrm{rfc}} = \left(n_{ij}^{\mathrm{rfc}}\right)_{i,j=1}^{n} \qquad \text{with} \qquad n_{ij}^{\mathrm{rfc}} = \sum_{i'=1}^{i-1} \sum_{j'=j+1}^{n} f_{i'j'}^{\mathrm{rfc}} \tag{15}$$

and consequently the RFM and the cumulative RFM contains the same information. The same relations, Eqs. (14,15), are also valid for the cumulative rainflow count intensity, $\boldsymbol{\mu}^{\mathrm{rfc}}$, and the limiting rainflow matrix, $\boldsymbol{G}^{\mathrm{rfc}}$.

EXAMPLE 2 (Extreme rainflow matrix for a Markov load). In fatigue applications it is often practical to consider only the turning points of the load process, and model them by a Markov chain. One useful property of the Markov model is that its limiting rainflow matrix can be calculated exactly, see e.g. Rychlik [20], Frendahl & Rychlik [8], or Johannesson [11, 12]. Hence, the level crossing intensity can also be calculated. For the Markov model we can, therefore, compare the Poisson approximation with the exact result. For a model with 100 discrete load levels in the range $[-1, 1]$, the comparison is presented in Figure 3. Observe that the approximation and the exact RFM are identical in the upper left part of the RFM, i.e. for cycles with large amplitudes. To model the RFM for those cycles was the goal of the approximation, and here we can see that the result is satisfactory. □

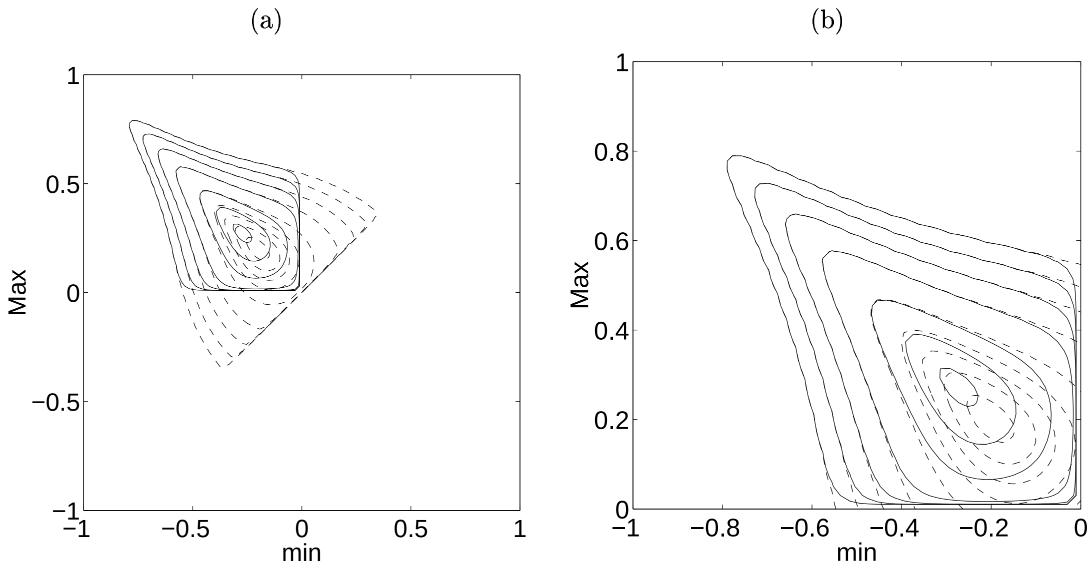

Figure 3. Example 2, Markov load. Iso-lines of rainflow matrices: Extreme RFM (−), Exact RFM (− −); (b) shows a magnification of the upper left part of the RFMs.

3. Extrapolation of Level Crossing Intensity

When extrapolating the rainflow matrix for a measured signal, its level crossing intensity is not known, and hence it has to be estimated from data. For high (and for low) levels we propose to extrapolate the observed level crossing spectrum by a parametric shape, the Generalized Pareto Distribution (GPD), suggested by the extreme value theory.

The number of crossings of a high level, v, say, is approximately equal to the number of peaks exceeding v. In extreme value analysis one uses a technique called Peak Over Threshold (POT), where the exceedances over a high threshold are considered. (Davison & Smith [3] gives a nice presentation of the technique.) Hence, for high levels, we propose to approximate the level crossing intensity by

$$\mu(v) = \mu(v_0)H(v - v_0), \qquad \text{for} \quad v > v_0 \tag{16}$$

and the same for low levels, $u < u_0$. The form of $H(x)$ is given by the GPD with the survival function

$$H(x) = \begin{cases} (1 - \frac{kx}{\sigma})^{1/k}, & k \neq 0 \\ \exp(-\frac{x}{\sigma}), & k = 0 \end{cases} \tag{17}$$

with the shape parameter $-\infty < k < \infty$, and the scale parameter $\sigma > 0$. When $k > 0$ the GPD has an upper end point, $0 < x < \sigma/k$, while for $k \leq 0$ the support is $0 < x < \infty$. The theory of extreme values is closely related to the GPD. The Generalized Extreme Value distribution (GEV) is $G(x) = \exp(-H(x-b))$. There are several facts that motivate the use of GPD for exceedances over high levels.

1. Attraction to a GEV is equivalent to POT convergence to GPD. (Pickands [18])
2. The maximum of a Poisson distributed number of independent GPD variables has a GEV distribution.
3. The GPD is threshold stable, i.e. the exceedances for GPD variables gives a GPD, but with different parameters.

Both the last two properties characterize the GPD in the sense that no other family of distributions has either property.

For the GPD, we have two parameters k and σ to estimate. In the Matlab toolbox WAFO, four methods for estimating the parameters in a GPD is implemented, namely Pickands' method (Pickands [18]), the method of moments (see e.g. Hosking & Wallis [10]), and the method of probability weighted moments (Hosking & Wallis [10]), and the maximum likelihood method (see e.g. Davison & Smith [3], or Grimshaw [9]). Standard errors for the estimates are also delivered. The different estimators were tried. The experience is that the maximum likelihood method is the best estimator, and the method is presented below.

Consider the exceedances $x_1, x_2, \ldots, x_n$. By writing $\sigma = k/\tau$ the maximum likelihood estimation is reduced to solving the single equation for τ

$$\frac{n}{\tau} = \frac{1}{k(\tau)} \sum_{i=1}^{n} \frac{x_i}{1-\tau x_i} \tag{18}$$

with

$$k = k(\tau) = -\frac{1}{n} \sum_{i=1}^{n} \log(1-\tau x_i). \tag{19}$$

Smith [26] showed that the maximum likelihood estimator exists if $k < 1$ and is asymptotically normal if $k < 1/2$. The limiting covariance matrix for (σ, k) is

$$\frac{1-k}{n} \begin{pmatrix} 2\sigma^2 & \sigma \\ \sigma & 1-k \end{pmatrix} \tag{20}$$

provided that $k < 1/2$. Eq. (16) also contains $\mu(v_0)$, which is estimated as the observed upcrossing intensity of level v_0.

If a distribution is in the domain of attraction to the Gumbel distribution (GEV with $k = 0$), then the exceedances of high levels converge to the exponential distribution (GPD with $k = 0$). (This is the case for Gaussian processes.) If one knows or believes that one has convergence to the Gumbel distribution, then one can use the exponential distribution. Hence, one only has to estimate the parameter σ, with maximum likelihood estimator

$$\sigma^* = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{21}$$

Another problem is to choose a proper threshold, which is high enough in order to get a good fit to the GPD, but still contains enough exceedances to get good estimates.

To choose an optimal threshold is a notoriously hard problem which is addressed in e.g. Dress & Kaufmann [5], Dupuis [7], Draisma et al. [4], and Beirlant et al. [1]. Here, we suggest that one tries different thresholds and judge which gives a decent fit to the GPD, e.g. by using quantile-quantile plots. For the GPD, the mean exceedance over a level v, when $k > -1$, is a linear function of v. Hence, it can be used as a diagnostic plot when choosing the threshold. For an automatic choice of the threshold one could systematically try different thresholds and choose the one that gives the best fit according to some criteria, e.g. minimizing the Mean Square Error (MSE) of a given diagnostic plot. This was tested for the exponential case, but the method was not very robust, and could be hard to use in practice. For a more crude automatic choice of the threshold, one could use some rule of thumb, e.g. choose the threshold as the mean plus two standard deviations, or choose the threshold so that it represents the upper 20 % of the range of the data, or choose the threshold so that it represents a certain fraction of the data.

EXAMPLE 3 (Extrapolation of level crossing intensity). The Markov model in Example 2 is used for generating a simulated signal, containing 10 000 cycles, from which the level upcrossings are extracted. The upper and lower tails of the level upcrossing spectrum are extrapolated by using the GPD as described in Section 3. The maximum likelihood estimates gives the extrapolation in Figure 4a. By setting $k = 0$ in the GPD we obtain the exponential distribution, whose maximum likelihood estimates lead to the graph in Figure 4b. The exponential tails, which give a straight line in the log-scale, often tend to overestimate the intensity for extreme crossings, as in this example. Here the GPD (with $k \neq 0$) gives a better extrapolation. □

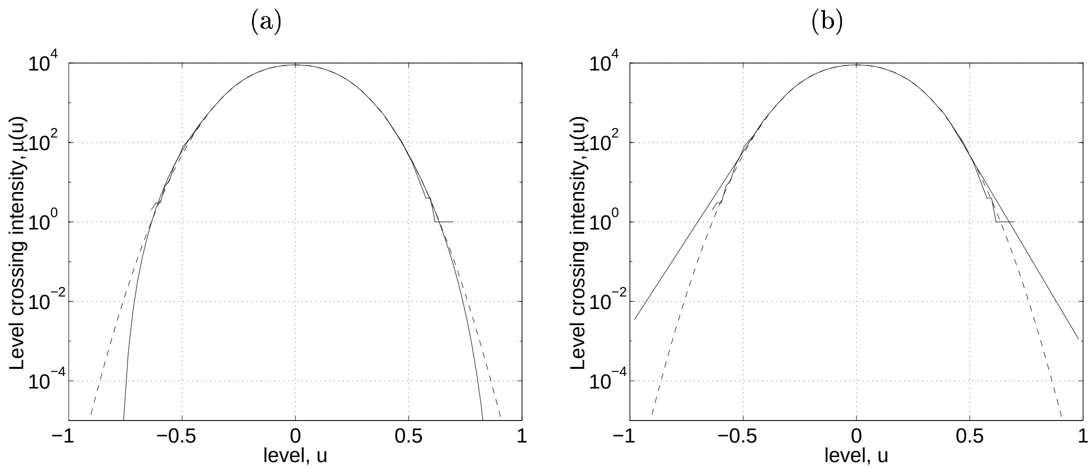

Figure 4. Example 3, the extrapolation of level crossing intensity for a simulated Markov load. The level crossing intensities are extrapolated by using (a) GPD, maximum likelihood, (b) Exp, maximum likelihood, and are compared with the exact one (− −).

4. Extreme Rainflow Matrix

Here we will present an algorithm for calculating an estimate of the extreme RFM from a measured RFM. Thus we will combine Theorem 1 with the extrapolation of the level upcrossings (Section 3). Automatic choices of the necessary parameters for the extrapolation is proposed.

The steps are as follows.

1. *Observed RFM.* The input is a measured RFM, $\boldsymbol{F} = (f_{ij})_{i,j=1}^n$, where f_{ij} is the number of cycles with the minimum i and the maximum j.

2. *Extrapolation of the level upcrossings.* The level upcrossings are calculated from the RFM. The threshold for extrapolation is chosen and the curve is estimated. We present the extrapolation procedure for high levels. For the low levels the procedure is the same.

   a) The level upcrossing spectrum $\boldsymbol{N} = (n_k)_{k=1}^n$ is calculated as

   $$n_k = \sum_{i<k<j} f_{ij} \tag{22}$$

   i.e. the summation is for all cycles with minimum $> i$ and maximum $< j$.

   b) We suggest two methods for an automatic choice of the threshold, denoted t.

   - The threshold represents 20% of the range of the data. Let $\Delta_i = n_i - n_{i-1}$. Choose the threshold so that 20% of the non-zero Δ_i is above the threshold.
   - The threshold represents 5% of the sample. Choose the threshold as the level that is crossed as often as 5% of the maximum value of the level upcrossing spectrum.

   These choices seem to work well in practice. In order to avoid clustering of exceedances, which comes from cycles with a minimum larger than the threshold, we set $f_{ij} = 0$ for all $i > t$. If $\boldsymbol{F}$ was changed, repeat step 2a and 2b.

   c) Now we shall estimate the shape of the level upcrossing intensity above the threshold t. Let $\Delta n_i = n_{t+i} - n_{t+i-1}$, which is the number of peaks a distance i above the threshold. Construct the vector (x_j) of observed exceedances

   $$\boldsymbol{x} = (\underbrace{0.5\ \dots\ 0.5}_{\Delta n_1}\quad \underbrace{1.5\ \dots\ 1.5}_{\Delta n_2}\quad \underbrace{2.5\ \dots\ 2.5}_{\Delta n_3}\quad \dots) \tag{23}$$

   where we made a correction of 0.5, because the estimation procedure is for non-discrete data. On the basis of $\boldsymbol{x}$ we calculate estimates of the parameters k and σ in the GPD, which gives the extrapolation

   $$n_i^+ = n_t \left(1 - \frac{k(i-t)}{\sigma}\right)^{1/k}, \qquad i \geq t. \tag{24}$$

   The recommendation is to use the maximum likelihood method for the GPD. An alternative is to use the exponential distribution (GPD with $k = 0$) giving

   $$n_i^+ = n_t \exp\left(-\frac{i-t}{\sigma}\right), \qquad i \geq t. \tag{25}$$

Based on $\boldsymbol{x}$ the maximum likelihood estimate is $\sigma = \frac{1}{N}\sum x_i$. However, in this case one can obtain a maximum likelihood estimate for the discretized data

$$\sigma = \frac{1}{log(1 + N/S_N)} \tag{26}$$

with

$$N = \sum_i \Delta n_i, \qquad S_N = \sum_i (i-1)\Delta n_i = \sum_j (x_j - 0.5). \tag{27}$$

The last estimator, Eq. (26), is recommended for the exponential case.

d) Let n_i^+ and n_i^- be the extrapolation for high and for low levels, respectively, with thresholds t^+ and t^-, respectively. The final estimate of the level upcrossing intensity is

$$n_i^* = \begin{cases} n_i^- & \text{if } r \le t^- \\ n_i^+ & \text{if } r \ge t^+ \\ n_i & \text{otherwise} \end{cases} \tag{28}$$

3. Now we can use Theorem 1 to calculate the cumulative RFM (see also Eq. (2))

$$\mu_{ij} = \frac{n_i^* n_j^*}{n_i^* + n_j^*} \tag{29}$$

and calculate the corresponding RFM $\boldsymbol{G}$ by using Eq. (14). The result is only valid for large cycles, and for regions where the approximation is not valid, the resulting RFM may contain negative values, which is not possible. Hence, to get an admissible RFM we set those elements to zero, i.e. $g_{ij} = \max(0, g_{ij})$.

EXAMPLE 4 (Extreme rainflow matrix for a Markov load). This is a continuation of Examples 2 and 3. From the extrapolated level crossings, see Figure 4, the extreme RFM is calculated by using Eq. (29). The two RFMs in Figure 5 represent two methods for extrapolating the level crossings, namely GPD and Exp, respectively. In this case the GPD gives the best result. □

5. Smoothing of a Rainflow Matrix

The kernel density estimation is a general statistical technique for non-parametric estimation of an unknown density. Since there are no parametric forms for the rainflow matrix, a Kernel Density Estimator (KDE) is suitable. The KDE is obtained by convoluting the observations by a so-called Kernel function, which means that a "bump" (the Kernel function) is placed at each observation, and then all the "bumps" are summed up to get the estimated density. A parameter of the smoothing is the so-called bandwidth, which is the width of the kernel and hence controls the degree of smoothing.

For general references to KDE the reader is referred to Silverman [25], Scott [24], or Ruppert & Wand [28]. A study of smoothing a RFM is found in Dressler et al. [6], where they also describe the method of kernel smoothing. They give an example of test track measurements, where they extrapolate a one-lap RFM to a 6 lap RFM, and compare it with a measured 6 lap RFM.

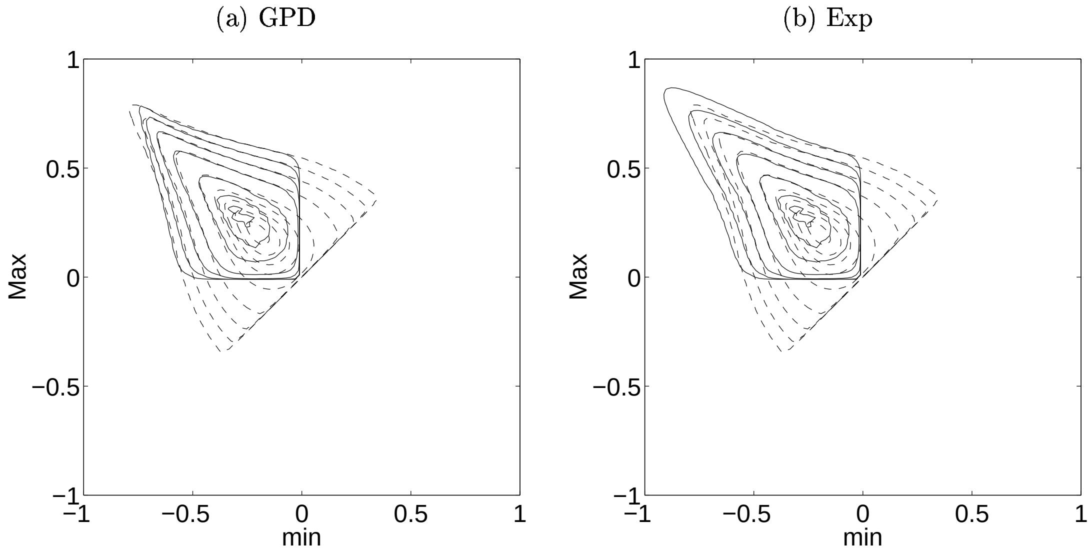

Figure 5. Example 4, the extreme RFMs calculated from estimated level crossing intensities. The level crossing intensities are extrapolated by using (a) GPD, maximum likelihood, (b) Exp, maximum likelihood. Iso-lines of RFMs (−) are compared with the exact RFM (− −).

EXAMPLE 5 (Smoothing of a rainflow matrix). The same simulated signal, form a Markov load, as in Example 4 was used here. The observed RFM was the input to the smoothing. The result is presented in Figure 6. The kernel function is a 2-dimensional normal distribution. □

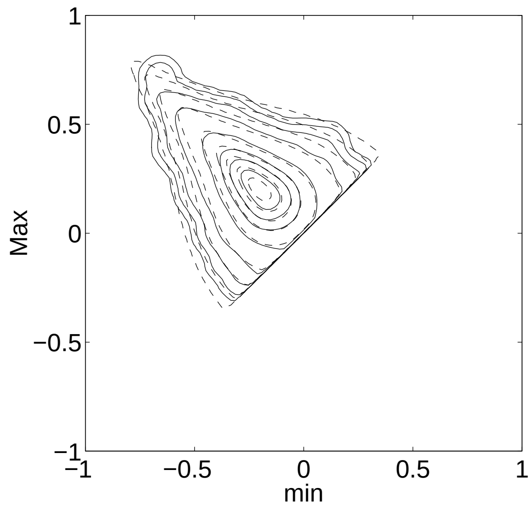

Figure 6. Example 5, the estimate of the limiting shape of the RFM obtained by using a kernel smoother. Iso-lines of the RFM (−) are compared with the exact RFM (− −).

6. Limiting Rainflow Matrix

The goal is to find a good estimate of the limiting RFM. The idea is to use the extreme RFM where it is valid (i.e. for large amplitudes) and to use the smooth RFM elsewhere. Hence, we use the extreme RFM for amplitudes above a threshold amplitude. To determine where the extreme RFM starts to be valid is assessed by comparing the damage (per amplitude) between the measured RFM and the extreme RFM. The threshold amplitude is chosen as the first amplitude that gives e.g. at least 95% of the damage of the observed RFM. Since the extreme RFM is not valid for high minima and low maxima (see Figure 5), these limitations are also applied.

EXAMPLE 6 (Limiting rainflow matrix for a Markov load). Here we combine the results from Examples 4 and 5. The final estimates of the limiting RFMs are presented in Figure 7. □

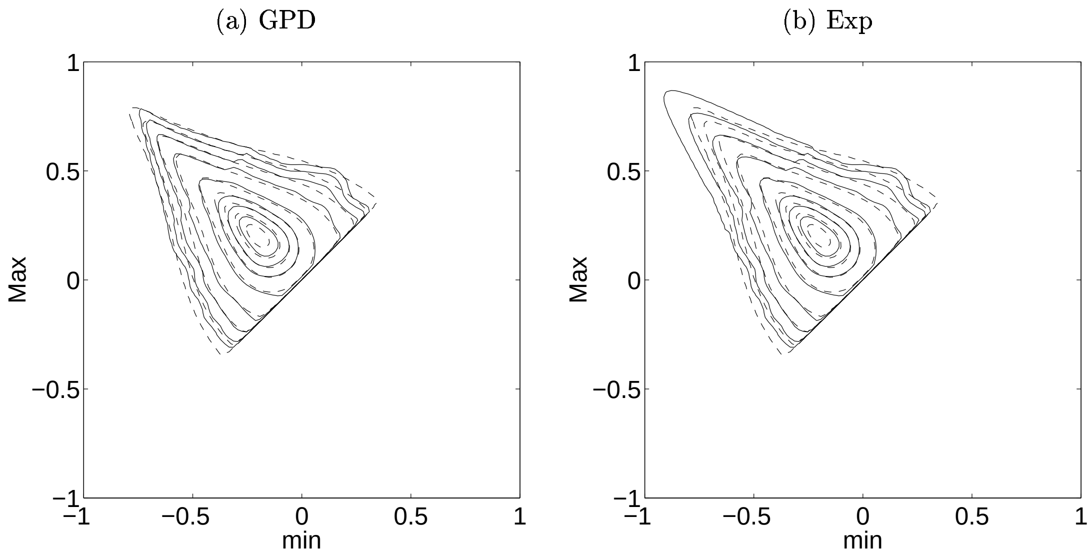

Figure 7. Example 6, the limiting RFM for a Markov load. It is a combination of two RFMs, the extreme RFM for large amplitudes, and the smoothed RFM elsewhere. The level crossing intensities are extrapolated by using (a) GPD, maximum likelihood, (b) Exp, maximum likelihood. Iso-lines of RFMs (−) are compared with the exact RFM (− −).

EXAMPLE 7 (Extrapolation of a rainflow matrix for a measured load). Here we will analyse two load signals that are measured on test tracks. The time signals are forces (units Newton) on the left front wheel of a Peugeot automobile. The two signals are one-lap measurements from two different parts of the test track (here called A and B). The time signal is rainflow filtered, so that small oscillations are removed, and the measured rainflow matrix is calculated by using 64 discrete load levels, ranging from -2000 to 12000. The measured RFM is the input to the calculations, where the first step is to extrapolate the level crossings to unobserved load levels. The thresholds for extrapolation are chosen automatically as the levels that are crossed as often as 5% of the maximum value of the level crossing spectrum.

The extreme RFM is calculated according to Eq. (2). The idea is to use the extreme RFM where it is valid (i.e. for large amplitudes) and to use the smooth RFM elsewhere. Hence, we use the extreme RFM for amplitudes above a threshold amplitude, chosen according to the damage criterion. The limiting RFMs is shown in Figure 7, where the observed cycles are marked with dots.

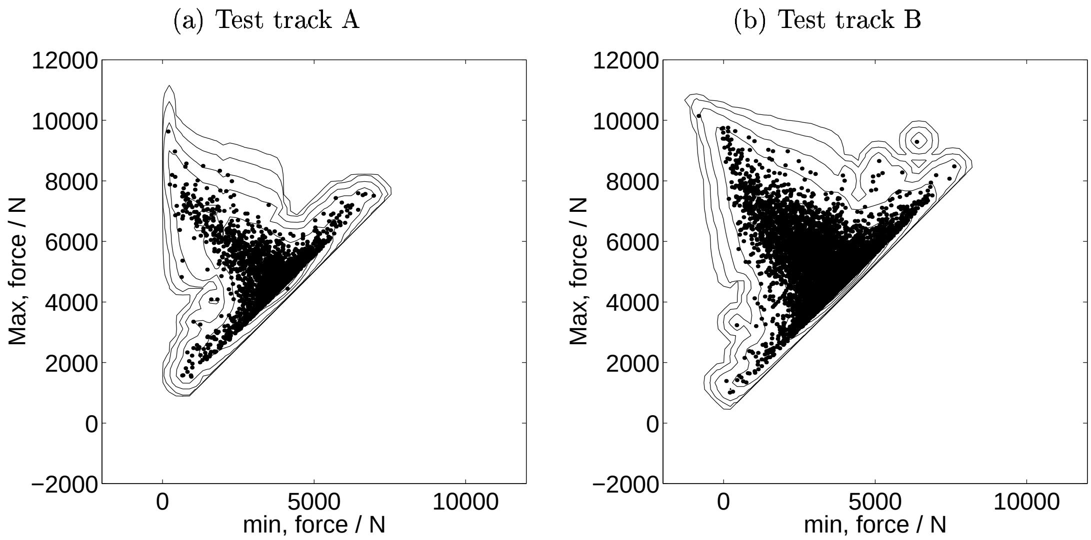

Figure 8. Example 7, the extrapolation of a measured load, vertical forces on the front wheel of a Peugeot. The estimated limiting RFM is a combining the extreme RFM, for large amplitudes, and the smooth RFM elsewhere. The level crossing intensities are extrapolated by using the GPD, maximum likelihood estimation. (a) Test track A, (b) Test track B. Iso-lines of RFMs (−) are compared with the rainflow cycles in the signals (•).

The limiting RFMs (in Figure 7) now represent the "long run" RFMs for one lap on the test tracks A and B, respectively. The last step is to extrapolate to the desired distance (or time). This is done by multiplying the limiting RFM by a suitable scaling factor. Assume that a full lifetime is equivalent to 5000 laps on the test track A, or 500 laps for test track B. The extrapolated RFMs then becomes $G_A^{\text{life}} = 5000 \cdot G_A$, and $G_B^{\text{life}} = 500 \cdot G_B$, respectively. The extrapolation based on the measured RFMs becomes $F_A^{\text{life}} = 5000 \cdot F_A$, and $F_B^{\text{life}} = 500 \cdot F_B$, respectively.

The level crossings are easily extracted from the RFM. In Figure 9 the extrapolation of the level crossings based on the measured RFM, and on the limiting RFM are compared. Clearly, when one wants to extrapolate to high levels one should use the limiting RFM. Now, we can predict how often a certain high level will be crossed. We can also specify the frequency of crossings and get an estimate of the corresponding load level, see Table I.

The frequency distribution of the rainflow amplitudes is also easily extracted from the RFM. Figure 10 shows the extrapolation of the cumulative frequency of the rainflow amplitudes, based on the observed RFM, and on the limiting RFM. When extrapolating to high amplitudes the limiting RFM should be used. By specifying the cumulative frequency we get a prediction of the corresponding amplitude, see Table II. □

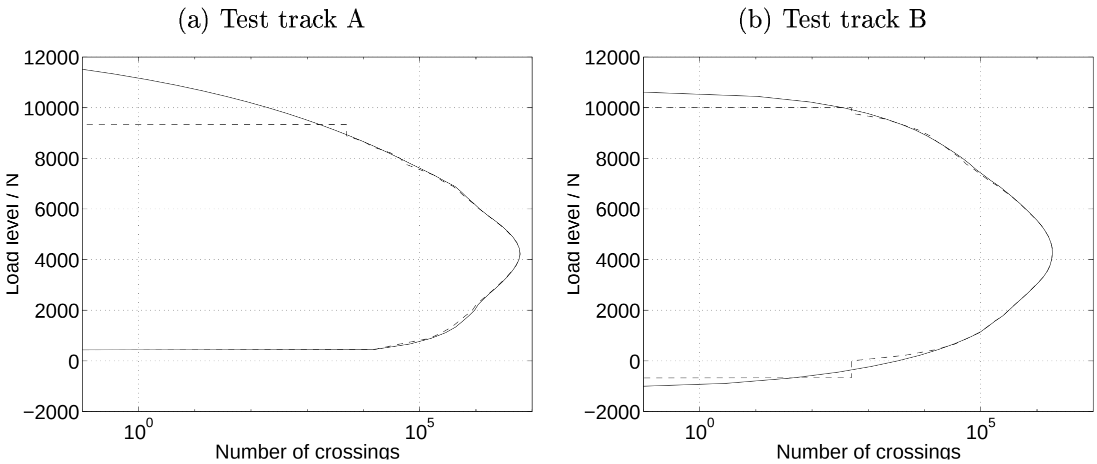

Figure 9. Example 7, extrapolated level crossings for a measured load. (a) Test track A, (b) Test track B. The extrapolated level crossings (−) are compared with the observed ones (− −).

Table I. Example 7, extrapolated level crossings.

Frequency of the number of crossings	Estimated load level / N G_A^{life}	F_A^{life}	G_B^{life}	F_B^{life}
1	11200	9330	10650	10000
10^3	9520	9330	9770	9700
10^4	8660	8670	8950	9010

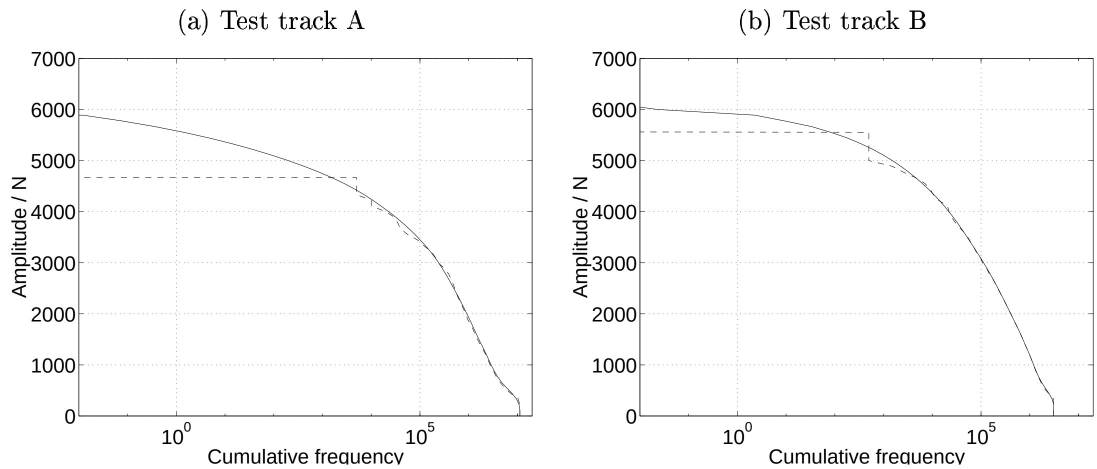

Figure 10. Example 7, extrapolated rainflow amplitudes for a measured load. (a) Test track A, (b) Test track B. The extrapolated cumulative frequencies of rainflow amplitudes crossings (−) are compared with the observed ones (− −).

Table II. Example 7, extrapolated rainflow amplitudes.

Cumulative	Estimated amplitude / N			
frequency	G_A^{life}	F_A^{life}	G_B^{life}	F_B^{life}
1	5600	4670	5950	5560
10^3	4750	4670	5110	4940
10^4	4250	4110	4370	4370

7. Conclusions

When extrapolating a measured RFM to longer time periods a simple rescaling of the measured RFM does not permit other cycles than the observed ones, and not all possible cycles. Therefore, we propose to estimate the limiting shape of the RFM, and use this matrix for extrapolation to longer periods.

We have derived an asymptotic expression for the cumulative rainflow intensity, which is valid only for large cycles. The assumption is that the exceedance of high and low levels converges to two independent Poisson processes. (This is true for stationary random processes under suitable conditions.) The expression for the extreme rainflow matrix only depends on the level crossing intensity, which is easier to estimate, than the complete limiting RFM. The extreme RFM is a very good approximation of the limiting RFM for large amplitudes, see Examples 1 and 2. To estimate the level crossing intensity for high (and for low) levels we suggest to use results from extreme value analysis, POT (Peak Over Threshold) analysis.

We have proposed a method for estimating the limiting shape of the rainflow matrix.

– *For large cycles: Extreme rainflow matrix.* We propose to use the extreme RFM where it is valid. The level crossing intensity is estimated from the measured RFM. A detailed algorithm for the calculations is presented in Section 4.

– *For small and moderate cycles: Smoothed rainflow matrix.* Here we propose to use a kernel smoother to estimate the RFM for cycles with small and moderate amplitudes.

The proposed estimation procedure of the limiting RFM is illustrated in Examples 4-6, where a Markov model for the load is used, which enables us to compare the results with the true limiting RFM. In Example 7 measured vehicle loads from test tracks are extrapolated by using both the measured RFM and the estimated limiting RFM. From a RFM it is easy to extract level crossings and rainflow amplitudes, and the extrapolations of these are also examined.

Acknowledgements

We are grateful to PSA (Peugeot Citroën) for their financial support and for supplying us with test track measurements. We would also like to thank professor Jacques de Maré and Dr. Thomas Svensson for their encouragement, stimulating discussions and helpful comments.

References

1. J. Beirlant, J. L. Teugels, and P. Vynckier. *Practical Analysis of Extreme Values.* Leuven University Press, Leuven, 1996.
2. P. A. Brodtkorb, P. Johannesson, G. Lindgren, I. Rychlik, J. Rydén, and E. Sjö. WAFO – a Matlab toolbox for analysis of random waves and loads. In *Proceedings of the 10th International Offshore and Polar Engineering conference, Seattle*, volume III, pages 343–350, 2000.
3. A. C. Davison and R. L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Ser. B*, 52:393–442, 1990.
4. G. Draisma, L. de Hann, L. Peng, and T. T. Pereira. A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes*, 2:367–404, 1999.
5. H. Dress and E. Kaufmann. Selecting the optimal sample fraction in univariate extreme value statistics. *Stochastic Processes and their Applications*, 75:149–172, 1998.
6. K. Dressler, B. Gründer, M. Hack, and V. B. Köttgen. Extrapolation of rainflow matrices. SAE Technical Paper 960569, 1996.
7. D. J. Dupuis. Exceedances over high thresholds: A guide to threshold selection. *Extremes*, 1:251–261, 1998.
8. M. Frendahl and I. Rychlik. Rainflow analysis: Markov method. *International Journal of Fatigue*, 15:265–272, 1993.
9. S. D. Grimshaw. Computing maximum likelihood estimates of the generalized Pareto distribution. *Technometrics*, 35:185–191, 1993.
10. J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29:339–349, 1989.
11. P. Johannesson. Rainflow cycles for switching processes with Markov structure. *Probability in the Engineering and Informational Sciences*, 12:143–175, 1998.
12. P. Johannesson. *Rainflow Analysis of Switching Markov Loads.* PhD thesis, Mathematical Statistics, Centre for Mathematical Sciences, Lund Institute of Technology, 1999.
13. M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes.* Springer, 1983.
14. G. Lindgren and I. Rychlik. Slepian models and regression approximations in crossing and extreme value theory. *International Statistical Review*, 59:195–225, 1991.
15. M. Matsuishi and T. Endo. Fatigue of metals subjected to varying stress – fatigue lives under random loading. In *Preliminary Proceedings of The Kyushu District Meeting*, pages 37–40. The Japan Society of Mechanical Engineers, March 1968. *In Japanese.*
16. M. A. Miner. Cumulative damage in fatigue. *Journal of Applied Mechanics.*, 12:A159–A164, 1945.
17. A. Palmgren. Die Lebensdauer von Kugellagern. *Zeitschrift des Vereins Deutscher Ingenieure*, 68:339–341, 1924.
18. J. Pickands III. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975.
19. I. Rychlik. A new definition of the rainflow cycle counting method. *International Journal of Fatigue*, 9:119–121, 1987.
20. I. Rychlik. Rain-Flow-Cycle distribution for ergodic load processes. *SIAM Journal on Applied Mathematics*, 48:662–679, 1988.
21. I. Rychlik. Rainflow cycles in Gaussian loads. *Fatigue & Fracture of Engineering Materials & Structures*, 15:57–72, 1992.
22. I. Rychlik. The two-barrier problem for continuously differentiable processes. *Advances in Applied Probability.*, 24:71–94, 1992.
23. I. Rychlik. Note on cycle counts in irregular loads. *Fatigue & Fracture of Engineering Materials & Structures*, 16:377–390, 1993.
24. D. W. Scott. *Multivariate Density Estimation.* Wiley, New York, 1992.
25. B. W. Silverman. *Density estimation for statistics and data analysis.* Chapman and Hall, 1986.
26. R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72:67–92, 1985.
27. WAFO Group. WAFO – a Matlab toolbox for analysis of random waves and loads, tutorial. Mathematical Statistics, Lund University, 2000.
28. M. P. Wand and M. C. Jones. *Kernel Smoothing.* Chapman & Hall, 1994.