DISTRIBUTIONS OF RESPONSES EXPERIENCED BY A VESSEL
-
A COMKISS REPORT

ANASTASSIA BAXEVANI, KRZYSZTOF PODGÓRSKI, AND IGOR RYCHLIK

September 2000

ABSTRACT. The purpose of these notes is to gather some ideas and possible approaches on how to analyze distributions of responses experienced by a vessels during a voyage. In this presentation we attempt to give a complete account of definitions and results. The main purpose is to clarify the methodological basis for future reference so one can avoid possible errors in interpretation and oversimplifications that could cause inaccuracies in statistical analysis.

1. Introduction

Let us consider a ship which is undertaking voyages over a certain relatively long period of time. The load she experiences and consequently her response to this load are random and thus they would be best analyzed by some reliable statistical methods. Clearly, from the engineering point of view, the most important is to find an accurate approach to study the extremal events occuring during this period of exploration. The principal methodological challenge in the so-posed problem comes from several different sources of the involved random variation. They are directly related to temporal-spatial scale within which one consider the sea surface. Although, conditions occuring on the sea surface are changing continuously in time and in space, it is quite common in the literature that the randomness is different in different scales and different methods should be chosen depending on the scale. We consider the following three different threshold levels. First, there is randomness related to short-time variability of the sea surface in time intervals measured in minutes or in few hours at most and in a restricted region in which the weather conditions appear to be the same. Another level of variability is due to the change of the sea states which is the consequence of changes of weather conditions occuring within several hours or even several days and also due to different conditions in distant regions of the sea. Then there is a stochastic variability of different journeys which are undertaken at different times and possible along different routes. Of course, these are only few of many possible factors and the complete analysis would be extremely complex if possible at all. In these notes we focus on these factors which are most important for the statistical properties of extremal waves (or responses). However, the proposed methods can be also applied to studies of arbitrary, not only extremal, waves.

In various practical problems, for example in safety considerations, it is important to determine the probability $P(X > r)$ that the response X exceeds some critical level r. The problem is not a simple one, since it is not even obvious how such a probability should be defined. Some more precise formulation is needed which has to take into account the intended use of the probability. In order to define it correctly, often the ideal situation of unlimited available data is considered and then frequencies of the events that are denoted by $P(X > r)$ has to be extracted and correctly interpreted in the terms of the used model.

To be more specific let us consider the process of load or response (for most of our consideration there is no need to discriminate one from another as the response is a linear or non-linear

Date: September 2000.

filter of the load) denoted here by $X(t)$. The generic quantity we are interested in is

$$P(\max_{t\in[0,T]} X(t) > r,).$$

Of course, it has to be specified what is the meaning of this quantity by identifying the model for $X(t)$ as well as what is the time interval which is considered in the above expression (one voyage, duration of a storm, lifetime of the vessel). Also various techniques can be used depending how high is the level r as asymptotic results are often accessible for high levels. High levels occurs, in principle, only in the extreme conditions of storms on the sea. This brings the importance of statistical models of storms and loads resulting from them.

Most of the material included in our considerations appeared in some form in the literature but this is, as far as we know, the first attempt to put together the consistent methodological framework which, as we hope, will be helpful l for future advances in research on these topics.

In the remaining part of this introduction we describe the concepts and notions which allows for the precise formulation of the discussed research problems.

1.1. **Sea state.** One of the crucial concepts in our studies is the notion of the sea state. It is natural to assume that over a certain short time, say 20 minutes, the conditions which are determining behavior of the sea surface are steady enough to consider it as a stationary random process. In Gaussian modeling of such processes the directional spectrum $S(\omega, \alpha)$ defines completely random properties of the surface. Thus it would be natural to describe the sea states through the spectrum governing the sea-surface process. In practice though, this is not possible, since one has available only some statistics (i.e. estimates of parameters) related to the spectrum. Most often these are significant wave height H^s, mean period T^z and the main wave direction φ, along the path. In the following those three parameters, denoted by $\vartheta = (H^s, T^z, \varphi)$, will be called the sea-state parameters or shortly sea-state. Once they are established the corresponding sea spectra are taken from some parametric family.

An important issue is to determine what is the actual duration of a given sea-state. Of course, it is a random variable distribution of which is depending on the state itself. In Doucent et al. (1987) analysis of Frigg field data set has been presented. It was shown that the distribution of the sea-state duration T_{stat} can be well approximated by an exponential distribution $P(T_{stat} > t) = \exp(-(t-0.4)/3.2)$, $t > 0.4$ (in hours), and a mean of 3 hours. This is unconditional distribution disregarding a sea state in hand. Even more interesting, although not surprising, is that more violent storms (higher significant wave height) last shorter. The conditional expectation of T_{stat} given the significant wave height, decreases exponentially, see Figure 4 in Doucent et al. (1987). As pointed out in the paper, the duration of the sea-state can be shorter than 20 minutes: "This is especially true for large values of $H_{1/3}$ " ($H_{1/3}$ stands for the significant wave height). The duration of a sea-state is important when sampling frequency of the data is considered. We will talk more about this in the next subsection.

1.2. **Route.** Assume that a vessel is sailing between two locations from A to B. Here we picture a route as a curve drawn on a map of ocean. We intend to give methods to combine the distributions of the loads or ship's responses from several voyages.

We denote time by t and a position on the ocean by $pos = (lo, la)$, where lo, la, stands for longitude, latitude, respectively. The route will be denoted by AB and parameterized by the distance from the starting point A, i.e. $AB = \{pos(s); s \in (0, S)\}$, where S is the distance from A to B. If we know the time T it takes from A to B, the starting time of the voyage t_0 and the speed $v(t)$ at any time during the voyage, then the route can be also parameterized by time $t \in [t_0, t_0 + T]$ by means of the relation $s = \int_{t_0}^{t} v(x)\, dx$. In practical computations the route is sampled and represented by a sequence of points $\widetilde{AB} = \{pos(s_i); i = 1, \ldots, N\}$, where $0 = s_1 < s_2 < \ldots < s_N = S$.

There are several ways of discretizing the route. The most natural one is to choose points equidistantly. We want the grid to be so dense that we do not need to consider variability of the sea-states between the grid points. This rises the question of a choice of sampling frequency. One could use the satellite data to investigate this issue. Since for fixed satellite observations we have fixed sampling frequency, we cannot come up with denser points without using oversampling techniques. But usually the data are sampled frequently enough so the question is whether one can take as a grid point every second or third satellite observation, or rather their averages to reduce the noise in the data. To verify approaches and a choice of sampling methods the statistics of responses based on sampled satellite observations could be compared with statistics obtained from the complete sequence of observations.

Another possibility would be to have a dynamic grid (not equidistant) based on the speed of the vessel, which often depends on the sea-state. One can sample time equidistantly and then compute positions s_i. The sampling frequency should be chosen high enough so, for example, we are not missing big storms. This is related to the problem of the duration of a sea-state since we want to avoid the possibility that a sea-state can be underestimated by taking too long sample intervals.

In this presentation, we shall often assume that the route is known in advance. However, it is also of interest to consider cases when only parts of a route are known (in extreme cases only points A and B). The general formulas should address this possibility, since in reality the route cannot be known exactly in advance because of uncertainty of such factors as sea conditions, captains actions, etc.. The variability of the route could be modeled as a random path in (lo,la) coordinates between A and B. In fact there exist programs that simulate the route of a ship and compute the time it takes to advance along such a route, see Aalbers et al. (1996).

1.3. **Response process – conditional Gaussian model.** For a voyage, i.e. a vessel sailing along a route, there are several responses that are of interest for the safety of the transport. Complete ship's response would be describe by a multivariate vector including such variables as sway, pitch, roll, heave and so on. For simplicity, we will limit ourselves to a one-dimensional signal denoted by $X(t)$, where as before t is the time variable.

In order to be able to simplify the analysis we assume that the sea surface can be accurately described by a Gaussian surface, with zero mean and a directional spectrum $S(\omega, \alpha)$. Furthermore we shall approximate the ship's movements (and other responses) by means of linear filters of the sea elevation and the fact that the speed of the vessel is constant for a relatively long, compared with wave periods, time. Under those assumptions the response X is a Gaussian process with some mean (assumed to be zero only for simplicity) and some spectral density $S(\omega)$. Both the mean and spectral density depend on many geometrical factors, loadings, ship's velocity (i.e. speed and heading) and the sea spectrum. (See Appendix I, for introduction to Gaussian modeling of sea surface and responses.)

The directional spectrum $S(\omega, \alpha)$ is needed for computation of the response spectrum and hence it should be known at any time as the vessel advances along the path. In our approach, it will be enough to know the sea-state $\vartheta = (H^s, T^z, \varphi)$, where as before H^s is the significant wave height, T^z is the mean wave period, and φ is the main wave direction.

ASSUMPTION I: A vessel encounters different sea states $\vartheta(s)$, $s \in (0, S)$ on a route from A to B. We assume that the sea-state parameters are constant in time intervals, which we call *periods of stationarity*. These periods are include at least hundred of waves of the average duration, i.e. they last at least a hundred of T^z. This corresponds to approximately 20 minutes of measurements. Often however, we assume them to last for two hours which corresponds to ca. 600 waves in time.

In addition we assume that for a given voyage the *probabilistic properties of the response depend only on the value of sea-state parameters ϑ and not on the position s of the vessel along the route.*

If the second part of the assumption is violated, then the route should be divided into shorter parts for which this assumption is satisfied. The interesting properties of responses can be studied for each part and then combined.

For a given voyage (vessel, loading), during stationarity period ΔT (under the assumptions: sea-state parameters are known, constant ship's velocity, etc.) the response spectrum can be computed by means of some dedicated computer software, e.g. VAC, see Aalberts et al. (1996). The derived response power spectrum is denoted by $S_\vartheta(\omega)$, the response variance by $\sigma^2(\vartheta)$ and its mean frequency, defined as the intensity of up-crossings by the response of its mean, by $f(\vartheta)$. As a consequence of Assumption I, the speed $v(t)$ of a vessel at time t, is a function of a voyage and of the present value of sea-state parameter $\vartheta(t)$. Consequently, we can view the VAC software as a function V, say,

$$V(\vartheta(t)) = (\sigma(t), f(t), v(t)), \qquad \text{where } \vartheta(t) = (H^s(t), T^z(t), \varphi(t)). \tag{1}$$

These parameters of the response as the function of the sea-state are called **short-term responses**.

In the case when it is not possible to compute the short-term response, since, for example, suitable software like VAC is not available, we can analyze a simplified response

$$(\sigma, f) = \left(\frac{H^s}{4}, \frac{1}{T^z}\right),$$

and $v(t) = v_0$, where v_0 is a constant speed while we also assume that $\varphi = 0$. This assumption is not uncommon in the literature, see Lindemann (1986): "the loading effect may be assumed proportional to the wave-height". Then we represent the response X as an encountered sea elevation level. In this simplified approach we do not consider the sea level measured from the moving vessel, which would be possible since we know the constant speed v_0. This allows to simplify the analysis by avoiding consideration of the Doppler effect.

1.4. **Distribution of response.** Our purpose is now to define the distribution of response, i.e. the generic quantity we denote by $P(X > r)$. Since this probability can be interpreted in a variety of non-equivalent ways, the used notation should be treated in a rather non-formal way. The precise definitions will follow later. Here we will consider two important different interpretation for the response X. First, we are interested in probabilities of **exceedances of absolute tolerance limits** by the absolute maximum of the response process $X(t)$ over the period of an entire voyage. This distribution allows, for example, to determine chance of the complete damage of a ship during long-term of exploration or, in the design stage, to construct a reliable ship serving a given sea.

The second meaning to the response X is the value of individual crest and we are interested in the probability of **exceedances of a threshold by such an individual crest**. Thus X now has the interpretation of the maximum of the response over the period of an individual wave. The distribution this sort is used in fatigue analysis, where one short excursion above a critical threshold not need be fatal but it only makes partial damage, e.g. a plastic deformation.

Intuitively, the difference between these two cases is similar to the problem of finding distributions of global maximum and local maxima for the response during a voyage. In the first case we have one observation per voyage in the second hundred thousands of local maxima (the highest, which is the global maximum, included).

We begin with the definition of the probability of exceedance of a tolerance limit r by the maximum response X during a voyage. *Consider N voyages and define $P(X > r)$ as the limiting fraction of voyages for which the maximum response exceeds level r, as N tends to infinity, if such* a *limit exists*. Then the probability $P(X > r)$, can be written as follows. Let $X_j(t), j = 1, \dots, N$, denote the response measured during the jth voyage, then

$$(2) \qquad P(X > r) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \{ \max_{t\in[0,T_j]} X_j(t) > r\},$$

where T_j is the time of the jth voyage, {"condition"} denotes both the set and the indicator of the set, i.e. it is defined to be one if the "condition" is true and zero otherwise. Observe that we have not assumed that the sea-state or even the route connecting A with B is fixed – these could vary between voyages. However, first we consider a special case called *deterministic route*, where we assume that during all the voyages the vessel encounters exactly the same and known sea-state process $\vartheta(t)$.

This special case can be visualized through the model in which N identical vessels are experiencing exactly the same sea states although in an independent way. For each vessel we record the maximal wave height in the response process. The proportion of vessels for which the response X exceeds at least once the level r is an estimate of the conditional probability of exceedance given the sea-state process $\vartheta(t)$. By the Law of Large Numbers, the limit (2) exists. In this situation the limit is denoted by $P(X > r|\vartheta(\cdot))$ we call it the conditional probability of excursion given that the sea-state process is known.

In general the sea-state process $\vartheta(t)$ is not fixed and varies between voyages. Simply it can happen that some voyages are performed in good weather and some in stormy weather. For each voyage the process $\vartheta(\cdot)$ is unknown, and surely differs from voyage to voyage, but we assume that the random mechanism (weather, captain, etc.) that generates $\vartheta(\cdot)$ is ergodic (the Law of Averages or the Law of Large Numbers are valid). In such case the limit $P(X > r)$ does still exist and equals

$$P(X > r) = E[P(X > r|\vartheta(\cdot))].$$

Note that $P(X > r|\vartheta(\cdot))$ is a random variable taking values between zero and one, where randomness comes through the now random sea state process $\vartheta(t)$. Our task is to find computable expression for $P(X > r)$ in the framework of some reasonable theoretical model. We will come back to this question in the next section. In our opinion, when computing probabilities it is best to use measurements and compute frequencies of occurrences of events of interest. However, if the data are limited or not available (not each response in a vessel can be measured), then we need to rely on mathematical modeling. Let us first illustrate such a situation by considering the so-called k-voyage value.

Example 1. **The k-voyage value** is denoted by r_k and often is used as a design value of a typical "big" wave that can be encountered by a vessel during k voyages. It is defined as the value that, in average, is exceeded once during k voyages. More precisely, if $Y_j(r) = \{\max_{0\le t\le T_j} X_j(t)\}$, where the underlying voyage is random (the sea-state process ϑ is random not fixed), then the expected value $EY_j(r) = P(X > r)$ depends on r and r_k is defined as the solution to the equation $P(X > r_k) = 1/k$ so that $E\sum_{j=1}^{k} Y_j(r_k) = 1$ giving a reason for the above interpretation.

Our task is to estimate r_k for, say, $k = 100$. Suppose that we have recorded the absolute maxima in $N = 100$ voyages. Since we have 100 observations of the maximal response $x_1, \dots, x_{100}$, we can estimate r_{100} by $\max(x_1, \dots, x_{100})$, expecting that by the Law of Large Numbers r_{100} should be approximately equal to the highest recorded value of of the extremal responses. However, if $N = 10$ which is more realistic sample size and we wish to know the 100-voyages value r_{100}, then we are looking for a value that probably has not been observed. We need to resort to some more subtle methods. For example, we could use the equation $P(X > r_{100}) = 0.01$, under the assumption that $P(X > r)$ is a smooth function of r, and hence the observed part of $P(X > r)$ (the empirical cumulative distribution function of the observed maximal values in 10 voyages) can be extrapolated to higher, unobserved values by some smoothing estimation techniques. Here the extreme value theory is particularly useful providing the appropriate tail behavior of the distributions, and we refer to two books Aage et al. (1999), containing oceanographic application, and Leadbetter et al. (1983) which is a monograph on these topics.

The second problem in which we deal with the distribution of the response in a different context is the distribution of crest heights. To avoid the confusion with the previous problem we use now a generic A^c for the individual crest height (amplitude).The whole problem can be defined in a similar way as before. First we will introduce the concept of "oscillations of a response encountered during a voyage". Oscillations are defined as apparent waves observed in the response: *for a fixed* **reference level** *m, define an oscillation as the part of the signal between two consecutive up-crossings of m.*

Now assume that one is recording the individual heights of oscillations that a vessel has encountered during each of N voyages. The values are presented in the form of an empirical distribution. That is for any level h one finds the fraction of oscillations with crests below h. Again by invoking the Law of Averages we expect that the fraction converges to a probability distribution function, as N tends to infinity. Denote by A^c a random variable having this distribution. Then the variable A^c represents the variability of crest heights during a voyage.

Using mathematical notation, the exceedance probability $P(A^c > h)$ is defined as follows. As before, consider N voyages. For each voyage estimate the empirical exceedance probability of $P(A^c_j > h)$. Then $P(A^c > h)$ is the limiting value of the average of $P(A^c_j > h)$, if such limit exists, i.e.

$$P(A^c > h) = \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} P(A^c_j > h).$$

As before we can also define the distribution of A^c conditionally on the value of sea-state process during a voyage, i.e. $P(A^c_j > h|\vartheta(t))$.

2. CONDITIONAL PROBABILITIES $P(\cdot|\vartheta(\cdot))$ – DETERMINISTIC ROUTE

In this section we assume that the random variability in experienced maximal wave height is only caused by the random variability of sea surface (waves) under deterministic sea conditions. Thus all probabilities considered in this section are conditional probabilities given the sea-state process $\vartheta(t)$.

2.1. Some definitions and assumptions.

Because we assume that the sea-state process is slowly varying, we can consider the response $X(t)$ to be stationary in relatively short time intervals. Thus we call stationary portions of $X(t)$ **short term responses**. More formally, if the sea-state is a constant function of time $\vartheta(t) = \vartheta$, then the resulting response will be a stationary response which corresponds to the short time response over the time interval in which any other sea-state (not necessarily constant) $\vartheta(\cdot)$ is approximately equal to ϑ. For short term responses our problems are reduced to the extreme value theory of stationary processes on which there is an extensive literature [see, for example, the monograph Leadbetter, Lindgren, and Rootzén (1983)].

In the following we introduce some measures for properties along the route, called also **long-term responses** which are mixtures of the short term responses. The mixing factors are sometimes called **wave climate along the route**.

We begin with two general assumptions.

ASSUMPTION II: During a stationarity period the reference level will be chosen as the level which is most frequently crossed and we assume that this level does not change between the different stationary sea conditions encountered by a vessel. This common reference level is denoted by m, and for simplicity only, we assume that $m = 0$.

In order to derive approximations for $P(X > r|\vartheta(\cdot))$ and $P(A^c > h\vartheta(\cdot))$, we will use the properties of crossings of levels r, and h, respectively. Denote by $N(r)$ the number of times the response $X(t)$ crosses upwards the level r during a voyage. Obviously $N(r)$ depends also on the encountered sea-states $\vartheta(\cdot)$, but for notational convenience we do not write explicitly this dependence.

Let us also introduce an additional concept of the *crossing intensity* under a sea-state ϑ, denoted by $\mu(r;\vartheta)$ and which is the number of crossings per time unit. More precisely, if during the whole voyage one had the same sea conditions ϑ, then $E[N(r)|\vartheta] = T\mu(r;\vartheta)$ and $\mu(0;\vartheta) = f(\vartheta)$.

Since the reference level is $m = 0$, the *number of oscillations* is equal to $N(0)$. If the sea-state process $\vartheta(t) = (H^s(t), T^z(t), \varphi(t))$ is known, then for a given function V, see (1), one can compute the mean frequency from

$$\hat{f} = \frac{1}{T}\int_{t_0}^{t_0+T} f(\vartheta(t))\, dt. \tag{3}$$

Using the crossing intensity, the expected number of crossings during a voyage can be obtained from the following mixture

$$E[N(r)] = \int_{t_0}^{t_0+T} \mu(r;\vartheta(t))\, dt. \tag{4}$$

For a long voyage, integration over time may not be practical and hence one can rewrite the formula by introducing the distribution of sea-state along the route. Let Θ be a random variable having the following distribution

$$P(\Theta \in A) = \frac{1}{T}\int_{t_0}^{t_0+T} \{t : \vartheta(t) \in A\}\, dt, \tag{5}$$

i.e. the fraction of time sea-state spends in the set A. Here $\{t : \vartheta(t) \in A\}$ is an indicator function which is equal to one for these t for which the condition in the brackets is satisfied and zero otherwise. Using the introduced variable Θ, the expected number of up-crossings can be written in a compact form

$$E[N(r)] = TE[\mu(r;\Theta)] = T\int \mu(r;\vartheta)p(\vartheta)\, d\vartheta, \tag{6}$$

where $p_{\vartheta(\cdot)}(\vartheta)$ is the probability density function (pdf) of the distribution given by (5) that measures the time a vessel is exposed to the sea-state parameters ϑ. In Section 4 we give methods to estimate $p_{\vartheta(\cdot)}(\vartheta)$, the so-called *exposure time density*.

Consequently, the mean frequency $\hat{f}$ is given also by

$$\hat{f} = E[f(\Theta)|\vartheta(\cdot)] = \int f(\vartheta)p_{\vartheta(\cdot)}(\vartheta)\, d\vartheta.$$

Next, we will introduce another probability density

$$q_{\vartheta(\cdot)}(\vartheta) = \frac{f(\vartheta)}{\hat{f}} p_{\vartheta(\cdot)}(\vartheta) \tag{7}$$

that measures the intensity of oscillations a vessel encounters while being in a sea-state ϑ. The function $q_{\vartheta(\cdot)}(\vartheta)$ will be called *oscillations density function*. Let us denote by $\tilde{\Theta}$ a generic random variable that has $q_{\vartheta(\cdot)}(\vartheta)$ as its pdf. This variable will be applied when the distribution of oscillation crest heights is approximated. More precisely, in the following subsection we shall use the following normalized crossing intensity

$$\frac{E[N(r)]}{E[N(0)]} = \int \frac{\mu(r;\vartheta)}{\mu(0;\vartheta)} \frac{f(\vartheta)}{\hat{f}} p(\vartheta)\,d\vartheta = \int \frac{\mu(r;\vartheta)}{\mu(0;\vartheta)} q(\vartheta)\,d\vartheta = E\left[\frac{\mu(r;\tilde{\Theta})}{\mu(0;\tilde{\Theta})}\right]. \tag{8}$$

We finish with the third and last assumption that we impose on the response process. That is, unimodality of crossing intensity as function of r, which implies that $1 - E[\frac{\mu(r;\tilde{\Theta})}{\mu(0;\tilde{\Theta})}]$ is a distribution function, often used as an approximation of $P(A^c < r|\vartheta(\cdot))$ (see the following subsection).

ASSUMPTION III: Assume that under stationary conditions, i.e. for any ϑ, the crossing intensity $\mu(r;\vartheta)$ has only one local maximum (at $r = 0$).

Under this assumption, $E[N(r)]$ has also only one maximum. The last assumption is reasonable in the case of linear responses driven by ocean waves.

Remark 1. For a deterministic route and sea-state parameters $\vartheta(t) = (H^s(t), T^z(t), \varphi(t))$ along a given path, both $p_{\vartheta(\cdot)}(\vartheta)$ and $q_{\vartheta(\cdot)}(\vartheta)$ are probability density functions. The random variables having these densities were denoted by Θ, $\tilde{\Theta}$, respectively. They have a very natural interpretation despite the fact that they seem so artificial. *The variable Θ tells us what is the value of sea-state process at randomly chosen time points during the deterministic voyage, while $\tilde{\Theta}$ tells us the value of sea-state parameter at a randomly chosen oscillation.* Obviously both variables are functions of the fixed sea-state process $\vartheta(\cdot)$, which we could be written more explicitly as $\Theta|\vartheta(\cdot)$ and $\tilde{\Theta}|\vartheta(\cdot)$. In the general case when the sea-state and the route are random, the wave climate along the route will be described by the unconditional variables Θ, $\tilde{\Theta}$.

2.2. **The height of the crest of an oscillation.** If the sea state parameters during a voyage are constant and equal ϑ, we shall denote the random crest height by $A^c(\vartheta)$. Assuming that one can find the distributions for $A^c(\vartheta)$, then, after some derivations, we obtain that

$$P(A^c > h|\vartheta(\cdot)) = P(A^c(\tilde{\Theta}) > h|\vartheta(\cdot)) = \int P(A^c(\vartheta) > h) q_{\vartheta(\cdot)}(\vartheta)\,d\vartheta,$$

where $q_{\vartheta(\cdot)}(\vartheta)$ is the pdf of $\tilde{\Theta}$, and has to be estimated. We will try to accomplish this by using satellite data.

If measurements are available one can fit some parametric distribution. Gran (1990) discusses several families of suitable distributions that could be applied for $P(A^c(\vartheta) > h)$ and $q_{\vartheta(\cdot)}(\vartheta)$.

Theoretical derivation of the density of $A^c(\vartheta)$ using the model of the response $X(t)$, is an extremely difficult problem and hence some approximation methods are of interest.

2.2.1. *Rayleigh model.* For Gaussian responses one is often approximating $A^c(\vartheta)$ by a standard Rayleigh distributed variable R

$$P(A^c(\vartheta) > h) = P(\sigma(\vartheta)R > h) = e^{-\frac{h^2}{2\sigma(\vartheta)^2}},$$

where $\sigma(\vartheta)^2$ is the variance of the response $X(u)$, see formula (1). It is well known that this is an accurate approximation for narrow band spectra. However as it was discussed in Rychlik and Leadbetter (1997), the Rayleigh approximation is accurate for distributions of $A^c(\vartheta)$ for high h levels, even for broad band spectra. More generally it was shown that for any response $X(t)$,

$$P(A^c > h) \leq \frac{E[N(h)]}{E[N(0)]}. \tag{9}$$

The inequality is usually "very" sharp for high values of h. Thus, for stationary periods of constant sea-state and if the response is Gaussian. we have

$$P(A^c(\vartheta) > h) \leq \frac{\mu(h;\vartheta)}{\mu(0;\vartheta)} = e^{-\frac{h^2}{2\sigma(\vartheta)^2}}.$$

2.2.2. *Approximation based on crossing intensities.* It is well known that crests of high but narrow waves are slightly higher than the Gaussian model would predict. This is also seen in the crossing intensity $\mu(h;\vartheta)$, for ϑ corresponding to higher significant wave heights.

Now, if $\mu(h;\vartheta)$ is known, then by using (9) and Assumptions II and III, we can directly check that

$$F(r;\vartheta) = 1 - \frac{\mu(r;\vartheta)}{\mu(0;\vartheta)}, \qquad F(r) = \int F(r;\vartheta)q(\vartheta)\,d\vartheta,$$

are distribution functions. Since $F(r;\vartheta)$ is increasing, $F(0;\vartheta) = 0$ and $F(+\infty;\vartheta) = 1$, we also have,

$$F(r) = 1 - \frac{E[N(r)]}{E[N(0)]}, \qquad r \geq 0.$$

Thus the relation

$$P(A^c(\vartheta) > h) \leq 1 - F(h;\vartheta),$$

provides with a conservative bound for the distribution of $A^c(\vartheta)$, which can be useful since we are interested in safety applications.

Clearly the problem is how to estimate the crossing intensity function $\mu(h;\vartheta)$. One could fit some parametric distribution functions to the intensities, using measurements at fixed locations. This approach is still under investigation. One could argue that if we are going to use observations why not then use some of the parametric distributions proposed in the literature, [see, for example, Gran (1990)]? The answer to this is that analysis based on crossings is much simpler than the study of crests of apparent waves. Particularly, if the non Gaussian model for the response is specified, then one can compute the crossing intensity by means of Rice formula

$$\mu(h;\vartheta) = \int_0^{+\infty} z f_{X'(0),X(0)}(z,h)\,dz.$$

For example, for second order nonlinear waves the crossing intensity has been computed in Machado (2000).

2.2.3. *Transformed Rayleigh model.* For Gaussian seas the Rayleigh approximation for the crest height is commonly used. For slightly non-Gaussian responses one can express the crest height as a transformed Rayleigh variable. We will use this representation while computing fatigue damage from rainflow cycles.

Relying on the assumption that the reference level is constant and the crossing intensities are unimodal, it can be proved that there is a unique increasing function g, $g(0) = 0$, with an inverse function $G = g^{-1}$, such that $P(A^c(\vartheta) > h) \leq P(G(R;\vartheta) > h)$. The function g is defined by

$$g(r;\vartheta) = \begin{cases} \sqrt{-2\ln(\mu(r;\vartheta)/\mu(0;\vartheta))}, & \text{if } r \geq 0, \\ -\sqrt{-2\ln(\mu(r;\vartheta)/\mu(0;\vartheta))}, & \text{if } r < 0. \end{cases} \tag{10}$$

Consequently, we have that

$$P(A^c > h) \leq E[e^{-g(h;\tilde{\Theta})^2/2}] = \int e^{-g(h;\vartheta)^2/2} q(\vartheta)\,d\vartheta = \int e^{-\frac{g(h;\vartheta)^2 - 2\ln(q(\vartheta))}{2}}\,d\vartheta.$$

The last integral has to be computed numerically. To do so, quite often, one uses the so called saddle point method, which is a particularly accurate method for high h values.

The approach is as follows, see Gran (1992) for details: 1) check for a fixed h if the function $f(\vartheta) = g(h;\vartheta)^2/2 - \ln(q(\vartheta))$ has a local minimum at $\vartheta = \vartheta_0$, say. Then, use Taylor series for a quadratic approximation of f around the point $\vartheta = \vartheta_0$. Then the integral looks like the integral of a Gaussian density and can be computed explicitly;

$$\int e^{-f(\vartheta)}\,d\vartheta \approx e^{-f(\vartheta_0)}\int e^{-\frac{1}{2}(\vartheta-\vartheta_0)[f''(\vartheta_0)](\vartheta-\vartheta_0)^T}\,d\vartheta = (2\pi)^{3/2}\det[f''(\vartheta_0)]^{1/2}e^{-f(\vartheta_0)},$$

under the assumption that $[f''(\vartheta_0)]$ is a positive definite matrix.

2.3. **The distribution of rainflow cycles.** We turn now to the fatigue caused by the load oscillations and in particular to the study of rainflow cycle distribution. We shall not go into details on how to compute the rainflow cycles amplitude distribution, but concentrate rather on the properties of accumulated damage that is $E[D_\beta(t)]$.

By Assumptions II, III, the expected damage accumulated in the time interval $(0,t)$, under stationary conditions defined by ϑ, can be bounded as follows

$$E[D_\beta(t)] \le Tf(\vartheta)E[(G(R;\vartheta) - G(-R;\vartheta))^\beta],$$

where R is a standard Gaussian variable and G is the inverse of g, defined by (10), see Rychlik and Leadbetter (1997) for a detailed discussion on how the bound is derived. In the special case of Gaussian responses we have

$$E[D_\beta(t)] \le Tf(\vartheta)E[(2\sigma(\vartheta)R)^\beta] = Tf(\vartheta)\sigma(\vartheta)^\beta\, 2^{\beta/2}\Gamma(\frac{\beta}{2}+1).$$

The parameters $f(\vartheta)$ and $\sigma(\vartheta)$ are important in order to perform a crude analysis of the damage intensity of the response. The mean frequency indicates a rate of bigger cycles while $\sigma(\vartheta)$ is the scaling factor for amplitudes. (Observe that in some special cases, when $\beta = 1$ or as β goes to infinity, the method is exact).

Now, under Assumptions I, II and III, we have that the stationarity periods are relatively long, and since the most frequently crossed level is constant, then the rainflow cycles associated with sea state variation are of the same order as the cycles within the stationarity periods and hence can be neglected. This leads to the following approximation

$$E[D_\beta(t)] \approx T\hat{f}E[(G(R;\tilde{\Theta}) - G(-R;\tilde{\Theta}))^\beta] = T\hat{f}\int\int E[(G(r;\vartheta) - G(-r;\vartheta))^\beta]re^{-\frac{r^2}{2}}q(\vartheta)\,dr\,d\vartheta.$$

For the Gaussian case the formula reduces to

$$E[D_\beta(t)] \approx T\hat{f}\,2^{\beta/2}\Gamma(\frac{\beta}{2}+1)\int \sigma(\vartheta)^\beta q(\vartheta)\,d\vartheta.$$

Observe that for any response satisfying Assumptions II and III we can write that

$$E[D_\beta(t)] \le \hat{f}\int E[(G(R) - G(-R))^\beta]q(\vartheta)\,d\vartheta,$$

where the inverse function g is defined by (10) with $\mu(r;\vartheta)$ replaced by $E[N(r)]$.

Finally, if Assumptions I and II are violated, then the method based on crossings may become too conservative. If more accurate predictions of fatigue life are needed, then more detailed models have to be developed for the sequence of load cycles. Here the Markov chain theory has shown to be particularly useful. There are two reasons for this:

- the Markov models constitute a broad class of processes that can accurately model many real loads,
- for Markov models, the fatigue damage prediction using rainflow method is particularly simple, Rychlik (1988) and Johannesson (1999).

In the simplest case, the necessary information is the intensity of pairs of local maxima and the following minima (the so-called Markov matrix or min-max matrix). The dependence between other extremes is modeled using Markov chains, see Frendahl & Rychlik (1993).

2.4. **Poisson approximation for the number of exceedances.** The following approach is sometimes called the Rice method and is based on the fact that crossings of high levels can often be very accurately modeled by a clustered Poisson process.

Assume that $X(t_0) < r$, i.e. we are not starting with an extreme response, then

$$P(X > r) = P(\max_{t \in T} X(t) > r) = P(N(r) > 0) \leq E[N(r)].$$

The last upper bound is close to the accurate value for very high levels under the assumption that crossings are not clustering, i.e. a crossing of a high level is not directly followed by another crossing. However, if up-crossings are always coming in pairs (cluster of two), then we could improve the inequality by $P(N(r) > 0) \leq 0.5 * E[N(r)]$.

It is a well known result for stationary processes, see Leadbetter et al. (1983), that under mild assumptions the up-crossings of high levels behave asymptotically as a clustered Poisson point process. From this it follows that for high levels of r

$$P(X > r) = 1 - P(N(r) = 0 | \vartheta(\cdot)) \approx 1 - e^{-\lambda E[N(r) | \vartheta(\cdot)]}, \tag{11}$$

where λ, $0 < \lambda \leq 1$, the so-called extremal index is the inverse of the average size of clusters of high maxima.

Let us make several remarks.

- For responses during stationary periods defined by ϑ we will use $\lambda = 1$. This is correct when the response is Gaussian. Even if for narrow band spectra waves are coming in groups, asymptotically the extreme waves are not clustering. However, for the Poisson approximation to be accurate for a narrow band sea the finite level need to be taken relatively high. On the other hand during storms the sea is not narrow band so the clustering is not occuring and the quality of approximation is rather good.
- Another problem can be that some responses are non-Gaussian, for example in the case of slamming. This is an important problem but we will not consider such effects here.
- A voyage is a mixture of stationary periods and hence one can ask if the method will work for the whole route. Obviously, for some sea-states ϑ, the chosen threshold r is high enough so that the use of Poisson approximation is well motivated, while it may not be for other encountered sea-states. A pragmatic answer to this question is that the quality of approximation increases as the level r, or the duration of the worst sea-state encountered, increases. More precisely, if the worst sea-state lasts for ca. 20 minutes the levels r such that $P(X > r) \leq 0.2$, are well approximated. If the time increases from 20 minutes to two hours, then the same levels will have the probabilities of exceedance increased, so that $P(X > r) \leq 2/3$ (ca.), see the following remark for more detailed analysis.
- Here we present some simple analysis of validity of Poisson approximation. For simplicity assume that the response X is just the sea surface elevation and let $\vartheta(t)$ be the history of the sea-states encountered during a voyage. Let H_0^s denote the highest significant wave height, encountered by a vessel. We assume that the sea is Gaussian, duration of a sea-state is ca. 20 minutes and the average wave period 12 seconds. Consider a level u, say, that is 3.5 standard deviations, i.e. $0.875 * H_0^s$. (Such level u is high enough so that the Poisson approximation for the exceedance probability can be used). For the period of 20 minutes the expected number of up-crossings of the level u is $E[N(u)] \approx 0.22$, and the probability that the maximum exceeds u is ca. 0.2. It is important to notice that if we assume that the sea was calm when $H^s(t) < 0.6 * H_0^s$, then we are underestimating the $E[N(u)]$ by at

most 0.006, in a 20 day period. The similar analysis for $H^s(t) < 0.5 * H^s(t_0)$ gives value of underestimation $3.3 \cdot 10^{-6}$.

The conclusion is that if we are interested in the probabilities $P(X > r)$ less than 0.2, then one can neglect all the sea-states having significant wave-height $H^s < 0.67 * H_0^s$. The period of 20 minutes is quite short, but if the spectrum of the sea is not extremely narrow banded then the level u can be decreased to $0.75H_0^s$ with approximative exceedance probability 2/3 and so we expand the region of applicability of Poisson approximation.

In the above analysis we have assumed that the storm was "box"-shaped which corresponds to stationary periods of sea under the extreme weather conditions. This assumption is somehow idealistic. Therefore it is interesting to consider other models of storms. If we perform the same analysis under the assumption that the storm was shaped like a parabola then we get $E[N(u)] \approx 0.1576$, and the probability that the maximum exceeds u is approximately 0.1458. We can see that the corresponding values are smaller, something that was expected since we haven't change the storm duration.(see picture the parabola is inside the box). Again it is important to notice that if we consider the sea to be calm for $H^s(t) < 0.6 * H_0^s$ then we underestimate the $E[N(u)]$ by 0.0025 which decreases to $1.181 \cdot 10^{-6}$ when the above assumption is made for $H^s(t) < 0.5 * H_0^s$.

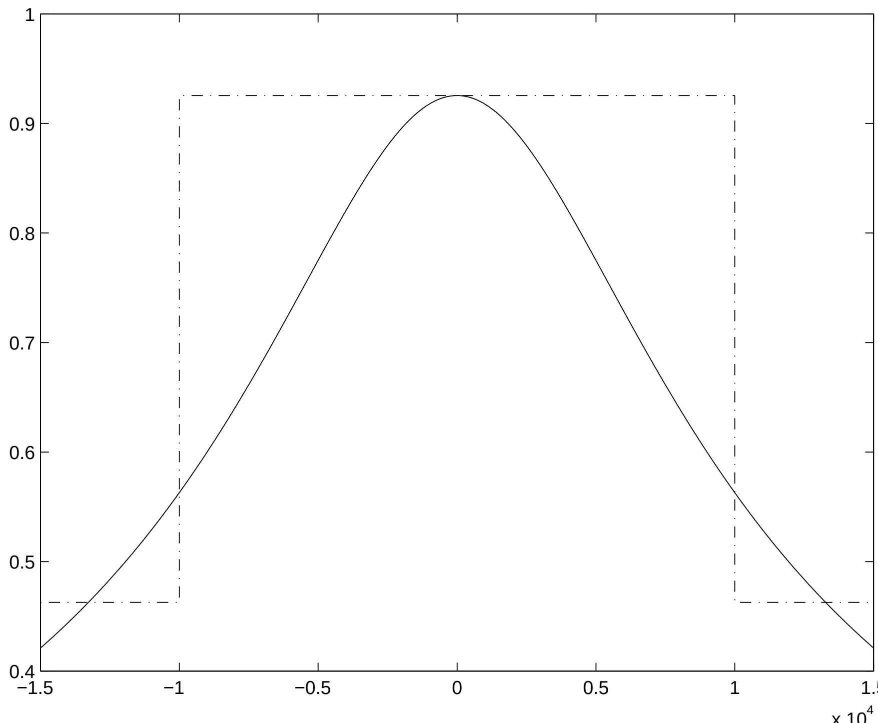

Figure 1: *Comparison of "box"-shape and parabola storms.*

When one is discussing the so called k voyage values, with $k = 10$, say, then the level u should be around 4.2 standard deviations high. In such case we can neglect all the sea-states for which $H^s(t) < 0.7 * H_0^s$. (Higher k-voyage values will further limit the number of sea states that need to be considered.)

The general conclusion is that, if we are interested in exceedance probabilities $P(X > r)$, for Gaussian responses, 0.25 or smaller, then the Rice method is working very accurately, and we need to compute $E[N(u)]$ for sea-states that are at least 0.6 as serious as the worst case. Unfortunately, so exact bounds can not be given for non-Gaussian responses.

2.5. **Evaluation of Poisson approximation for** $P(X > r)$**.** Obviously in order to make further computations we need to model the shape of $E[N(r)]$ as a function of r. We shall mention here some possible approaches and relate them to standard methods used in the oceanografic literature.

- The first approach is to consider the distribution of encountered crest heights during a voyage $P(A^c > r)$. Then

$$P(X > r) \approx 1 - e^{-T\hat{f}(E[N(r)]/E[N(0)])} = 1 - e^{-T\hat{f}(1-F(r))} \geq 1 - e^{-T\hat{f}P(A^c>r)}.$$

This is just one of the approaches taken in oceanographic literature. For example following Lindemann: "The maximum encountered during N wave cycles (in our formula $N = T\hat{f}$) will itself exhibit an extreme value distribution. For the long term distribution of waves, the extreme value distribution will suffice

$$P(H_{max}) = e^{-e^{-(h(H_{max})-h_0(N))}}, \ldots"$$

By taking $P(A^c > r) = \exp(-h(r))$ we have calibrated the formulas. However there is one important difference. In the last formula one needs to analyze the distributions of crest heights which is a nontrivial problem. Several approximations are needed. Our approach is more direct since it is relying on a more basic concept of crossings, see the discussion on computation of $P(A^c > r)$, given in the previous subsection.

- Another possibility is again to consider that the voyage is a mixture of periods of stationarity. Then

$$P(X > r) \approx 1 - e^{-T\hat{f}(E[N(r)]/E[N(0)])} = 1 - e^{-T\hat{f}\int \frac{\mu(r;\vartheta)}{\mu(0;\vartheta)}q(\vartheta)\,d\vartheta} \geq 1 - e^{-T\hat{f}\int P(A^c(\vartheta)>r)q(\vartheta)\,d\vartheta}.$$

Again a saddle point method could be used to evaluate the integral. The method is of interest when we have models for the response during stationary periods. If the mixture density $q(\vartheta)$ is available then the exceedance probability could be computed approximately.

- Clearly, the drawback of the cases discussed above, is that one needs to have information on either $E[N(r)]$ or $\mu(r;\vartheta)$ for values of r higher than those available in the data. This is not a problem if we have mathematical models for the responses, but if not, then one needs to fit some parametric distribution. The fit should be more accurate for high values of r and hence one may directly concentrate to this region, which leads us to the techniques called *peak over threshold* POT.

$$P(X > r) \approx 1 - e^{-T\hat{f}(E[N(u)]/E[N(0)])(E[N(r)]/E[N(u)])} \geq 1 - e^{-T\hat{f}P(A^c>u)P(A^c>r|A^c>u)}, \quad r > u,$$

Since A^c is a mixture of different sea states, it is sometimes easier to find a suitable family of parametric distributions. Consequently one can use

$$P(X > r) \approx 1 - e^{-E[N(u)]\int \frac{\mu(r;\vartheta)}{E[N(u)]}p(\vartheta)\,d\vartheta} = 1 - e^{-E[N(u)]\int \frac{\mu(r;\vartheta)}{\mu(u;\vartheta)}q_u(\vartheta)\,d\vartheta}, \quad r > u,$$

where $q_u(\vartheta) = \frac{\mu(u;\vartheta)}{E[\mu(u,\Theta)]}p(\vartheta)$. Obviously both $\frac{\mu(r;\vartheta)}{\mu(u;\vartheta)}$ and $q_u(\vartheta)$ have to be computed or estimated. One possible simplification is to assume that responses for different stationary conditions have crossing intensities that differ only by some scale parameters, i.e. standard deviation $\sigma(\vartheta)$ and that there is a random variable Y such that $A^c(\vartheta) = \sigma(\vartheta)Y$. Then

$$\frac{\mu(r;\vartheta)}{\mu(u;\vartheta)} \approx \frac{P(A^c(\vartheta) > r)}{P(A^c(\vartheta) > u)} = \frac{P(\sigma(\vartheta)Y > r)}{P(\sigma(\vartheta)Y > u)}.$$

If Y is distributed according to a generalized Pareto distribution, which is a natural assumption for high u values, then

$$\frac{\mu(r;\vartheta)}{\mu(u;\vartheta)} \approx \left(1 - k\frac{r-u}{c\sigma(\vartheta)}\right)^{1/k}.$$

The scale parameter c and the shape parameters k need to be estimated.

We shall now make some numerical experiment in the case of Gaussian responses. Then the Pareto shape parameter $k = 0$, i.e. Y is exponentially distributed and we have

$$P(X > r) \approx 1 - e^{-E[N(u)]e^{-\frac{r^2-u^2}{2\sigma(\vartheta)^2}}} \approx 1 - e^{-E[N(u)]e^{-\frac{r-u}{\sigma(\vartheta)^2/u}}},$$

where the first approximation is the Poisson approximation and the second is the Pareto model. As before the accuracy of the second approximation is improving with increasing u level. However, if the parameters of the Pareto distribution have to be estimated then the level u can not be too high. Consider the same example of 20 minutes worst storm with significant wave height H_0^s and average period 12 seconds. Let us choose the level u to be three standard deviations high, then $E[N(u)] = 1.1$ (we have in average one up-crossing of the level), then for $k = 10$ voyage the Pareto model gives $r_{10} = 0.945H_0^s$, while the Poisson approximation gives slightly lower value $r_{10} = 0.925H_0^s$. Both values seem to be very close, however we are in the tail of the distribution and the value $0.945H_0^s$ exceeds one in average in 13 voyages and not in 10.

- Another possibility is to follow the approach given in Leadbetter et al. (1983), then

$$P(X > r) \approx 1 - e^{-e^{-a_T(\frac{r}{\sigma(\vartheta)}-b_T)}},$$

where $a_T = \sqrt{2\log(T)}$ and $b_T = \sqrt{2\log(T)} + \frac{\log(\sqrt{f(\vartheta)})}{\sqrt{2\log(T)}}$. We consider again the same example as before of 20 minutes worst storm with significant wave height H_0^s and average period 12 seconds. If we assume as above the level u to be three standard deviations high, then for $k = 10$ then the model for *extremes of Gaussian responses* gives $r_{10} = 0.9463H_0^s$.

We can apply the above consideration to find approximations of the k-voyage value. Under the assumption that the high excursions do not cluster, $P(X > r_k) \cong 1 - \exp(-E[N(r_k)])$. Consequently r_k is a solution to the equation $E[N(r_k)] = -\ln(1 - 1/k) \approx 1/k$. Observe that one could define r_k as a level that exceeds only once during k voyages (there is in average one wave that is higher than the level r_k). Then the level r_k is a solution to the equation $kE[N(r_k)] = 1$. Here we are assuming that the level r_k is high, so that at least for Gaussian seas, there is only one up-crossing of the level r_k per wave that exceeds the level. Consequently, both approaches give approximately the same value for r_k, if voyage is long and if one is interested in large number k.

2.6. **Computation of $p(\vartheta)$ and connecting different routes.** In Assumption I, we require that the properties of the response on the route AB depend only on some parameters individual for a ship (loading, geometry etc.) and only on sea state parameters ϑ. Let divide the route AB into K routes denoted by $B_{k-1}B_k$, $k = 1, \dots, K$ and $B_0 \equiv A$. We assume that the exposure time densities $p_k(\vartheta)$ and the total time that the voyage on the route $B_{k-1}B_k$, T_k have been found. Then the exposure time density $p(\vartheta)$ for the route AB is given by

$$p(\vartheta) = \sum_{k=1}^{K} \frac{T_k p_k(\vartheta)}{T},$$

where $T = \sum_{k=1}^{K} T_k$. The last formula can be used to compute $p(\vartheta)$. More precisely, one can choose K large enough to assume that on the routes $B_{k-1}B_k$, $\vartheta(t) \approx \vartheta_k$. Denote by s_k the length of the route $B_{k-1}B_k$. Since we assumed that the speed depends only on the value of the sea-state parameter ϑ, i.e. it is given by a function $v(\vartheta)$, then using for example kernel $h(x)$ with bandwidth d, the function p can be estimated by

$$p(\vartheta) \approx \sum_{k=1}^{K} \frac{T_k\, h((\vartheta - \vartheta_k)/d)/d}{T} = \sum_{k=1}^{K} \frac{\frac{s_k}{v(\vartheta_k)}\, h((\vartheta - \vartheta_k)/d)/d}{\sum_{k=1}^{K} s_k/v(\vartheta_k)}$$

Now let us consider the case that each sub route satisfies Assumption I, but the function V, see equation (1) is different for different sub routes (that is, the route AB does not satisfy Assumption I). Then the properties of the response depend on the position along the route and hence there is no meaning in computing the exposure density for AB. In such cases one needs to approximate each of the k exceedance probabilities $P(X_k > r)$, separately. Now we turn to computation of the exceedance probability $P(X > r)$ for AB composed of the K routes $B_{k-1}B_k$. Here the important assumption is that the route $\vartheta(t)$ is known, i.e. deterministic. Then, by Assumption I, stationarity periods are long and hence we can assume that maximal responses encountered on routes $B_{k-1}B_k$ are independent. (Observe that this not need to not be correct if $\vartheta(t)$ is a random process.) Then

$$P(X > r) = 1 - (1 - P(X_1 > r)) \cdot \ldots \cdot (1 - P(X_K > r)).$$

Obviously we have $\max P(X_k > r) \leq P(\max X_k > r) \leq \sum_{k=1}^{K} P(X_k > r)$.

We turn now to the distribution of oscillations amplitudes, and let A^c_k be the maximum height encountered during path k. Let $\hat{f}_k$ be the mean frequency for the voyage k, then the distribution of oscillations amplitude can be computed as follows

$$P(A^c > h) = \sum_{k=1}^{K} \frac{T_k \hat{f}_k P(A^c_k > h)}{T \cdot \hat{f}},$$

where $\hat{f} = \sum_{k=1}^{K} \frac{T_k \hat{f}_k}{T}$.

3. Random sea-state process $\vartheta(t)$

In Section 1.4, we defined the distribution of response under the general assumptions of an ergodic sea-state process. We also saw that in that case the time averages of functionals converge. Under these assumptions the probabilities $P(X > r)$ and $P(A^c > h)$, are well defined. We shall now discuss some aspects of computations of these probabilities.

In order to make the assumption of ergodicity more believable, we may assume that voyages start during a limited period of time and are not too long, so the weather system can be assumed to be time stationary and homogeneous over big regions of the sea. We also assume that the captain changes route on the basis of weather conditions and experience. For example voyages start in January and take 20 days.

A warning is in place here. We have only the old data available and we wish to predict the future. So everything is based on the assumption that either the future will be just like the past or it can be accurately predicted based on ergodicity. Of course here one could ask about the climate variation since there are some indications that storms get worst in the last decades, see Carter (1999). So, should we assume that voyages are performed in a limited period of time, or should we include trends in the climate evolution?

In Section 2 we discussed the case of conditional probabilities $P(\cdot|\vartheta(\cdot))$, where the sea-state process is known. Obviously, the probabilities $P(X > r)$ and $P(A^c > h)$ can been computed by means of

$$P(X > r) = E[P(X > r|\vartheta(\cdot))], \qquad \text{and} \qquad P(A^c > h) = E[P(A^c > h|\vartheta(\cdot))].$$

3.1. **The oscillation heights distribution** $P(A^c > h)$.

We begin with the second probability, which is somewhat easier to analyze. We have

$$P(A^c > h) = E[\int P(A^c(\vartheta) > h) q(\vartheta)\, d\vartheta] = P(A^c(\tilde{\Theta}) > h), \tag{12}$$

where here the variable $\tilde{\Theta}$ has another pdf than in Section 2.2. Simply it is the average of $q(\vartheta)$ over performed voyages. If one discretizes the possible sea-state parameters into N^{sea} possible combinations, then the method used in VAC is just the computation of formula (12), see the following remark.

The presented method also solves the fatigue problem, if only the average fatigue damage, $E[p(\vartheta)]$, can be computed.

If one wishes to investigate also the variability of the damage, then one needs to know the variance of $P(A^c > h|\vartheta(\cdot))$ for different voyages, which means that we need $E[N(r)^2]$, which is a much harder problem.

Remark: Often in practice, one is discretizing values of sea-state parameters and hence one has only N^{sea}, say, different combinations of values. If i is an index numerating the allowed combinations, then we will say that we have the sea-state no. i and we replace in notation ϑ by i. In Aalbers et al (1996), the formula for $P(A^c > r)$ is given, see formula (2). Using the notation of this paper, and considering only one stationarity region, the formula (2) can be written as follows

$$P(A^c > r) = \int q(\vartheta) e^{-\frac{r^2}{2\sigma^2(\vartheta)}}\, d\vartheta = \sum_{i=1}^{N^{sea}} \frac{p_i}{\bar{N} \cdot T_i^z} e^{-\frac{r^2}{2\sigma_i^2}}$$

where $\bar{N}$ is defined by formula (3) and is equal to the average number of oscillations per second in a stationarity period.

The computation of $E[p(\vartheta)]$ and $E[q(\vartheta)]$ is not so obvious if we allow routes to vary, but for a fixed route it is just a reformulation of the VAC approach.

There are several methods discussed in Section 2.2 to compute the conditional probability $P(A^c > h|\vartheta(\cdot))$ any of those can be used here. We just conclude that mostly those methods are based on different ways of computing the conditional expectation of up-crossings of level h, $E[N(h)|\vartheta(\cdot))]$. Consequently in order to compute $P(A^c > h)$ we need to compute the unconditional expectation $E[N(h)] = E[E[N(h)|\vartheta(\cdot))]]$. We shall next discuss the case of exceedance probability $P(X > r)$, which is a more complicated case. Clearly, we still have $P(X > r) \leq E[N(r)]$, but this bound can be very conservative if there is a tendency of clustering of storms.

3.2. **The exceedance distribution** $P(X > r)$.

Under assumptions of ergodicity and using the Poisson approximation for $P(X > r|\vartheta(\cdot))$, see Section 2.4, we have

$$P(X > r) = E[P(X > r|\vartheta(\cdot))] \approx 1 - E[\exp(-E[N(r)|\vartheta(\cdot)])].$$

In the previous section we have sketched how to compute the conditional expectation $E[N(r)|\vartheta(\cdot)]$, that is the expected number of up-crossings of the level r given that the sea-state process is known. Obviously $E[N(r)|\vartheta(\cdot)]$ is a random variable having a distribution that we do not really know how to compute. In the previous subsection we needed only to compute $E[N(r)]$, here this is not enough, since

$$P(X > r) \cong 1 - E[\exp(-E[N(r)|\vartheta(\cdot)])] = E[E[N(r)]] - \frac{E[E[N(r)]^2]}{2} + \frac{E[E[N(r)]^3]}{3} - \frac{E[E[N(r)]^3]}{6}, \ldots.$$

so we need all moments of $E[N(r)]$. Obviously, for high levels of r we need to compute only first, second and maybe third term of the series, in order to obtain sufficient accuracy. Note that by taking only the first term we have an upper bound to the probability, including second term a lower bound, etc. So the accuracy of the approximation can be controlled. Another possible approach is based on the formula

$$E[N(r) \mid \vartheta(\cdot)] = \int \mu(r;\vartheta) p(\vartheta)\, d\vartheta.$$

Obviously, the density $p(\vartheta)$ is a function of sea-state process $\vartheta(\cdot)$. We have a good chance to compute the expectation $E[p(\vartheta)]$. Then one could model variability of $p(\vartheta)$, i.e. multinomial probability function or Gaussian and Poisson.

The most straight forward approach is to get the distribution of $E[N(r)|\vartheta(\cdot)]$ by means of Monte Carlo simulation, probably using the following formula

$$E[N(r)|\vartheta(\cdot)] = \int_{t_0}^{t_0+T} \mu(r;\vartheta(t))\,dt.$$

(Clearly here T is a function of $\vartheta(\cdot)$.) Since we are interested mostly in high values of r we need to design a suitable algorithm for it. Especially since, as we have discussed in Section 2.4, the value of $E[N(r)|\vartheta(\cdot)]$ depends mostly on the duration of the worst storm (corrected by storms that are above half or more of the strength of the worst one).

We finish with a simple example.

Example: Assume that a voyage takes 20 days and only two sea-states are possible; the calm sea and the stormy sea with significant wave height 4 meters. Assume that the storms along the route can be described by means of a Poisson process with intensity one per two days and that each storm lasts for exactly two hours. We let the mean wave period f during a storm to be 14 seconds. Now let the response X be just the sea level elevation and the critical level $x \geq 4$ meters. If we denote by J the number of storms encountered by a vessel and by T the storm duration, then

$$\begin{aligned} P(X > x) \approx 1 - E[\exp(-E[N(x)|\vartheta(\cdot)])] &= 1 - E[\exp(-TfJ\exp(-x^2/2))] \\ &= 1 - \exp(-E[J](1 - \exp(-Tf\exp(-x^2/2))). \end{aligned}$$

If the number of up-crossings during a single storm is small, then

$$1 - \exp(-Tf\exp(-x^2/2)) \approx Tf\exp(-x^2/2)$$

and we have that

$$P(X > x) \leq 1 - \exp(-E[N(x)]) = 1 - E[J]\exp(-Tf\exp(-x^2/2)).$$

For $x = 4$ this approximation gives the value 0.82, compare to the theoretical 0.79. Increasing level to $x = 5 = 1.25H^s$, the approximation gives value 0.0019 which is exactly equal to the theoretical one.

For $x = 4 = H^s$ the method based on the approximation

$$E[N(x)] - E[N(x)^2]/2 \leq P(X > x) \leq E[N(x)]$$

is not accurate, while for $x = 5$ we have that $E[N(x)] - E[N(x)^2]/2 = 0.019$, that is we get the correct value. Clearly the last method works if $J * 514\exp(-x^2/2)$ is small.

Let $X(t)$ be a stationary Gaussian process. The probability values considered above were exclusively estimated under the assumption that the storm was “box”-shaped. That is, the storm should have constant significant wave height and average wave period over a well-defined time and all wave amplitudes should be Rayleigh- or Rice- distributed with the same parameters. In the following we assume that the significant wave height follows a log-normal distribution, that is $lnH_u(s)$ is normally distributed. A convenient analytical model for the sea state behavior as a function of time is

$$H(t) = \frac{H_0}{\sqrt{1 + (\frac{t}{\tau})^2}},$$

where τ is the half storm duration parameter and H_0 is the storm maximum that occurs at $t = 0$. It can be proved that

$$H_0 = \exp(u + \frac{\sigma^2_{ln(H)}R^2}{2u}).$$

in reason for using such a model is that we get nice closed formulas for the probabilities of exceedance. Moreover by using the more general

$$H(t) = \frac{H_0}{\sqrt{1 + (\frac{t}{\tau})^s}}$$

we can even model the sea-state evolution at a location during and after a storm of limited duration. This, more peaked storm profile, exhibits a discontinuity and a merely linear rise and fall about the maximum sea-state. As shown in the picture (here we can put the picture of the formula for different s values) this storm has the advantage of being fitted to non-stationary storm profiles of different forms. It also covers the statioanery "box"-shape profile for very large values of s. This model can even be extended to the case of non symmetric storm waves by choosing different values for s before and after the maximum.

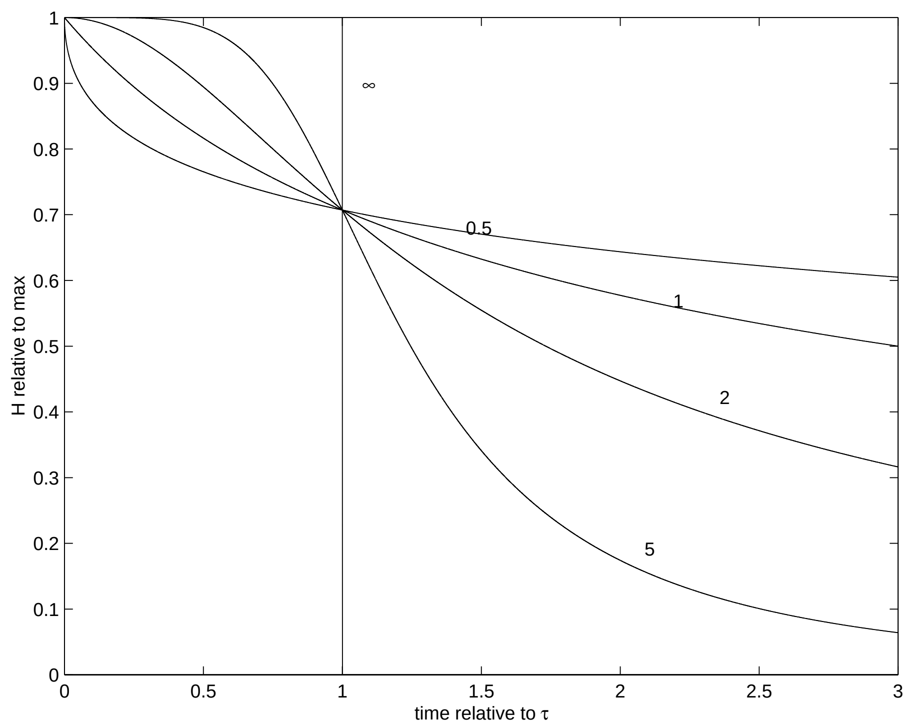

Figure 2: *Different storms for different s values.*

It is also interesting here to note that the storm duration τ, is in general equal to the time needed for the significant wave height to fall from H_0 to $0.707 H_0$. In general, it is a matter of uncertainty which values should be used for H_0 and τ. A simplified theoretical approach would be the use of the sequence $lnH_u(s)$, under the assumption of normality.

It can be proved that the distribution of the process $X(t)$ time s after u up-crossings is the same as the distribution of the variable

$$X_u(s) = \exp(-2u(\pi f_0)^2 s^2 + \frac{\sigma_{lnH}^2}{2u} R^2 + u)$$

where

$$f_0 = \frac{1}{2\pi} \frac{\sigma_{lnH'}}{\sigma_{lnH}}.$$

and R is a Rayleigh distributed variable for large values of u, that is for $u \approx 4\sigma_{lnH}$, or more.

Now by taking a Taylor series expansion we can get for the height of the first local maximum

$$H_u(s) = e^{u + \frac{\sigma_{lnH}^2 R^2}{2u}} - 2u(\pi f_0)^2 s^2 e^{u + \frac{\sigma_{lnH}^2 R^2}{2u}},$$

since for large u the random parabola is dominating the behavior of $X_u(s)$. Therefore,

$$H_u(s) = H_0(1 - u\frac{\sigma^2_{lnH'}}{\sigma^2_{lnH}}\frac{s^2}{2}).$$

After some simple but tedious arithmetic we can find that

$$\tau = \frac{1}{2\pi f_0\sqrt{u}}.$$

Then the above formula simplifies to

$$H_u(s) \cong \frac{H_0}{\sqrt{1+s^2(2\pi f_0)^2}}.$$

To get back to the problem of evaluating the exceedance probability, let us assume that a voyage consists of only two different sea-states; the calm sea and the stormy sea. Here we assume that we can describe the stormy sea by a parabola. For simplicity we also assume that all storms are identical and also that the storms along the route can be described by means of a Poisson process. If we denote by J the number of storms encountered by a vessel, then

$$P(X > x) \approx 1 - E[\exp(-E[N(x)|\vartheta(\cdot)])] = 1 - exp\left(E[J]\tau\frac{f(\vartheta_0)}{f_0}\frac{1}{\sqrt{2\pi}}\frac{1}{4x}E\left(H_0(i)e^{\frac{-8x^2}{H_0(i)^2}}\right)\right)$$

under the additional assumption of indepedance between J and R (Rayleigh distribution at up-crossings) as well as the independence between the $H_0(i)$, for $i = 1, 2, .., E[J]$

In order to evaluate

$$E\left(H_0(i)e^{\frac{-8x^2}{H_0(i)^2}}\right),$$

we are going to use the "saddle point" method.

$$P(X > x) \approx 1 - exp(E[J]2\tau\frac{f(\vartheta_0)}{f_0}\frac{\pi}{\sqrt{2}}\frac{u^2}{x^2}\frac{1}{\sigma^4_{lnH}}\exp(\frac{2u^2 - u(1+2ln(h_0))}{\sigma_{\ln H^2}})),$$

where $h_0 = \frac{2\sqrt{2}x\sigma_{lnH}}{\sqrt(u)}$ is the local min we found in the saddle method.

3.3. **Conditional Probabilities.** We have discussed long-run (over many voyages) distributions of $P(X > r)$ and $P(A^c > h)$, but an interesting special case that should be discussed separately is conditional distribution in the sense, what are distributions of $P(X > r)$ and $P(A^c > h)$ conditionally on some random event, extremely high storm that caused some damages and we asks for frequencies of safe returns to a harbor, nearest not necessarily point B. This would lead to models that remains marked crossing, Slepian model for sea-state processes.

4. Statistics based on Satellite data

This is the main object of studies in COMCISS. Our problem here is that we have data for sea-state along lines (that just crosses the path from A to B). The data are then sampled in time with return periods of c.a. 10 days. The data contain 7 years and hence for some positions storms may be missing (7 years with storm durations of 1 day say and 10 days sampling intervals is somewhat short). Consequently we have to use the data from other locations to improve estimates at the particular place.

Here we will be concerned with the randomness of sea-state parameters. We will omit the main wave direction ϑ at all. But start with only the significant wave height, for simplicity of notation denoted by H. We consider H in space and time, so it has three coordinates $H(lo, la, t)$.

The values of significant wave-height are usually splited in few discrete classes numerated by index $j = 1, \ldots, N^H$. Our objective is to find occupation probabilities

$$q_j = \frac{\text{time spend on the sea with significant wave height no. } j}{\text{duration of voyage}}.$$

REFERENCES

[1] Aage, C., Allan T.D., Carter, D.J.T., Lindgren, G. and Olagnon, M. (1999). *Ocean from Space - a textbook for offshore engineers and naval architects* IFREMER, **ISSN 1240-1153**

[2] Aalbers, A.B., Leenaars, C.E.J. and Quadvlieg, F.H.H.A. (1996) Voyage acceleration climate: a comprehensive statistical method for the evaluation of design loads for offshore transport. *Offshore Technology Conference 1996*

[3] Carter, D.J.T. (1999) Variability and Trends in the Wave Climate of the North Atlantic. *Proceedings of the Ninth ISOPE Conference*, **ISBN 1-880653-42-7**, pp. 12-18.

[4] Davison, A.C. and Smith, R.L. (1990) Models for Exceedances over Heigh Thresholds. *J. R. Statist. Soc. B*, **52**, pp. 393–442.

[5] Doucent, Y., Labeyrie, J. and and Thebault, J. (1987) Validation of Stochastic Environmental Design Criteria in the Frigg Field. *Advances in Underwater Technology, Ocean Science and Offshore Engineering*, Graham & Trotman, **Vol 12**, pp. 45–59.

[6] Frendahl, M. and Rychlik I. (1993) Rainflow Analysis: Markov Method. *Int. J. Fatigue*, **15** pp. 265–272.

[7] Gran, S. (1992) *A course in ocean engineering* Elsevier.

[8] Johannesson, P. (1999), *Rainflow Analysis of Switching Markov Loads*, PhD thesis.

[9] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes* Springer Verlag.

[10] Lindemann, K. (1986) Extreme value statistics and exposure time: a simple method to establish well defined criteria for marine operations. *Offshore Technology Conference 1986*

[11] Machado, U. (2000) Thesis in preparation.

[12] Matsuishi, M. and Endo, T. (1968) *Fatigue of metals subject to varying stress*. Paper presented to Japan Soc. Mech. Engrs, Jukvoka, Japan.

[13] Rice, S. (1944). The mathematical analysis of random noise. *Bell Syst. Techn. J.*, **23**, pp. 282-332.

[14] Rychlik, I. (1988) Rain flow cycle distribution for ergodic load processes. *SIAM J. Appl. Math.* **48**, 662-679.

[15] Rychlik I. (1993a): Note on Cycle Counts in Irregular Loads. *Fatigue Fract. Eng. Mater. and Struc.*, **16**, pp. 377-390.

[16] Rychlik, I. (1993b) On the "narrow-band" approximation for expected fatigue damage. *Probabilistic Engineering Mechanics*, **8**. 1-4

[17] Rychlik I., Johannesson P. and Leadbetter M. R. (1997) Modelling and Statistical Analysis of Ocean-Wave Data using transformed Gaussian processes, *Marine Structures*, **10**, 13-47.

[18] Rychlik I. and Leadbetter M. R. (1997) Analysis of ocean waves by crossing- and oscillation-intensities. *Proceedings of the Seventh ISOPE Conference*, **ISBN 1-880653-31-1**, pp. 206-213.

APPENDIX I: FREQUENCY MODELING OF LOAD HISTORIES

Assume that the response is specified in the frequency domain. This means that the signal is represented by a Fourier series

$$x(t) \approx m + \sum_{i=1}^{N} a_i \cos(\omega_i\, t) + b_i \sin(\omega_i\, t)$$

where $\omega_i = i \cdot 2\pi/T$ are angular frequencies, m is the mean of the signal and a_i, b_i are Fourier coefficients.

The important characteristic of signals in frequency domain is their power spectrum $\hat{s}_i = (a_i^2 + b_i^2)/(2\Delta\omega)$, where $\Delta\omega$ is the sampling interval in frequency domain, i.e. $\omega_i = i \cdot \Delta\omega$. The two-column matrix $\hat{s}(\omega_i) = (\omega_i, \hat{s}_i)$ will be called the power spectrum of $x(t)$.

The sequence $\vartheta_i = \arccos(a_i/\sqrt{2\hat{s}_i \Delta\omega})$ is called a sequence of phases and the Fourier series can be written as follows

$$x(t) \approx m + \sum_{i=1}^{N} \sqrt{2\hat{s}_i \Delta\omega} \cos(\omega_i t + \vartheta_i).$$

If the sampled signal contains exactly $2N + 1$ points then $x(t)$ is equal to its Fourier series at the sampled points. In the special case when $N = 2^k$, the so-called FFT (Fast Fourier Transform) can be used in order to compute the Fourier coefficients (and the spectrum) from the measured signal and in reverse the signal from Fourier coefficients.

As we have written before, the Fourier coefficient to the zero frequency is just the mean of the signal, while the variance is given by $\sigma^2 = \Delta\omega \sum \hat{s}(\omega_i) \approx \int_0^\infty \hat{s}(\omega)\, d\omega$. The last integral is called the zero-order spectral moment λ_0. Similarly higher-order spectral moments are defined by

$$\lambda_i = \int_0^\infty \omega^i \hat{s}(\omega)\, d\omega.$$

Random Functions in Spectral Domain. Assume that we get new measurements of a signal that one is willing to consider as equivalent, but it is seldom identical to the first one. Obviously it will have a different spectrum $\hat{s}(\omega)$ and the phases will be changed. A useful mathematical model for such a situation are the so-called random functions (stochastic processes) which will be denoted by $X(t)$. Here $x(t)$ is seen as a particular randomly chosen function - a sample path. The simplest case that models stationary signals with a fixed spectrum $\hat{s}(\omega)$ is

$$X(t) = m + \sum_{i=1}^{N} \sqrt{\hat{s}_i \Delta\omega} \sqrt{2} \cos(\omega_i t + \Theta_i),$$

where Θ_i are independent uniformly distributed phases. However, it is not a very realistic model, since in practice we often observe variability in spectrum $\hat{s}(\omega)$ between measured functions and hence $\hat{s}_i$ should be modeled as random variables too. In such a case, we assume that there is a deterministic function $S(\omega)$ such that the average value of $\hat{s}(\omega_i)\Delta\omega$ can be approximated by $S(\omega_i)\Delta\omega$ and in many cases one can model $\hat{s}_i = R_i^2 \cdot S(\omega_i)/2$ where R_i are independent random factors, all Rayleigh distributed. (Observe that the average value of R_i^2 is 2.) This gives the following random function

$$X(t) = m + \sum_{i=1}^{N} \sqrt{S(\omega_i)\Delta\omega} R_i \cos(\omega_i t + \Theta_i).$$

The process $X(t)$ has many useful properties that can be used in analysis like: for any fixed t, $X(t)$ is normally distributed, called also Gaussian distributed. A probability of any event defined for $X(t)$ can, in principal, be computed when the mean m and the spectral density S are known.

If $Y(t)$ is an output of a linear filter with $X(t)$ on the input, then $Y(t)$ is also normally distributed and we need to derive the spectrum of $Y(t)$ to be able to analyze its properties. This is a simple task, since if the transfer function of the filter $H(\omega)$ is given, then the spectrum of $Y(t)$, denoted by S_Y, is given by $S_Y(\omega) = |H(\omega)|^2 S(\omega)$. For example, the derivative $X'(t)$ is a Gaussian process with mean zero and spectrum $S_Y(\omega) = \omega^2 S(\omega)$. The variance of the derivative is $\sigma^2_{X'} = \int S_Y(\omega)\, d\omega = \lambda_2$.

The Gaussian process is a sum of cosine terms with amplitudes defined by the spectrum; hence, it is not easy to relate the power spectrum and the fatigue damage. The crossing intensity $\mu(u)$, which yields the average number of up-crossings of the level u, is given by the celebrated Rice formula

$$\mu(u) = f \exp(-(u-m)^2/2\sigma^2).$$

Using spectral moments we have that $\sigma^2 = \lambda_0$ while $f = \frac{1}{2\pi}\sqrt{\frac{\lambda_2}{\lambda_0}}$.

Appendix II: Fatigue Life Prediction – Rainflow Method

For completeness we shall give some short review of fatigue life prediction methodology.

In laboratory experiments, one often subjects a specimen of a material to a constant amplitude load, e.g. $L(t) = s\sin(\omega t)$ where s and ω are constants, and counts the number of cycles (periods) until it breaks. The number of load cycles $N(s)$ as well as the amplitudes s are recorded. Note that for small amplitudes, $s < s_\infty$, $N(s) \approx \infty$, i.e. no damage is observed. The amplitude s_∞ is called *the fatigue limit* or *the endurance limit*. In practice, one often uses a simple model for $N(s)$,

$$N(s) = \begin{cases} K^{-1}s^{-\beta} & s > s_\infty, \\ \infty & s \le s_\infty, \end{cases} \tag{13}$$

where K is a (material dependent) stochastic variable, usually log-normally distributed, i.e. with $K^{-1} = E\gamma^{-1}$ where $\ln(E) \in \mathrm{N}(0, \sigma_E^2)$, and γ, β, σ_E^2 are fixed constants.

For irregular loads, also called variable amplitude loads, one is often combining the S-N curve with a cycle counting method by means of the Palmgren-Miner linear damage accumulation theory, to predict fatigue failure time. The cycle counting forms equivalent load cycles. The now commonly used cycle counting method is rainflow counting and was introduced by Endo (1968). It was designed to catch both slow and rapid variations of the load by forming cycles by pairing high maxima with low minima even if they are separated by intermediate extremes. More precisely, each local maximum is a top of a hysteresis loop with an amplitude that is computed using rainflow algorithm.

Let t_k be the time of the k:th local maximum and s_k the amplitude of the attached hysteresis loop. Define the total damage by

$$D(t) = \sum_{t_k \le t} \frac{1}{N(s_k)} = K \sum_{t_k \le t} s_k^\beta = K D_\beta(t), \tag{14}$$

where the sum contains all cycles up to time t. The fatigue life time T^f, say, is shorter than t if $D(t) > 1$. In other words, T^f is defined as the time when $D(t)$ crosses level 1. A very simple predictor of T^f is obtained by replacing K in Eq. (14) by a constant, for example the median value of K that is γ. For high cycle fatigue, the time to failure is long (more than $10^5/\hat{f}$, where $\hat{f}$ is the mean frequency defined in section on crossings). Then for stationary (and ergodic and some other mild assumptions) loads, the damage $D_\beta(t)$ can be approximated by its mean $E[D_\beta(t)] = d_\beta \cdot t$. Here d_β is the damage intensity, i.e. how much damage is accumulated per time unit. This leads to a very simple predictor of fatigue life time $\hat{T}^f = \frac{1}{\gamma d_\beta}$

One way of computing expected accumulated damage during a voyage is to find the distribution of rainflow cycles. As before we are assuming that N voyages has been performed. During each voyage all rainflow cycles are found and a histogram of its amplitudes is computed. Then the distribution of rainflow cycle is defined as the average histogram when number of voyages N increases to infinity.

Although at the first sight the exceedance probabilities and the distribution of rainflow cycles seem to have little in common, in fact, it is not the case. We demonstrated that analysis of exceedances probabilities is related to properties of crossings of levels (preferable height), while rainflow cycles distribution can be computed from intensities of crossings of intervals, see Rychlik (1993a) for the definitions and proofs. Crossings of intervals can be bounded by crossings of levels, what, for Gaussian responses, leads to a commonly used Rayleigh approximation for cycles amplitudes see Rychlik (1993b) and Rychlik and Leadbetter (1997) for details.

The first and third authors: Centre of Mathematical Sciences, Department of Mathematical Statistics, Lund Institute of Technology, Box 118, SE-221 00 Lund, Sweden; *The second author:* Department of Mathematical Sciences, Indiana University – Purdue University Indianapolis, Indianapolis, IN 46202, USA.