
CONFIDENCE INTERVALS FOR EXCEEDANCE PROBABILITIES WITH APPLICATION TO EXTREME SHIP MOTIONS

Authors: DYLAN GLOTZER, VLADAS PIPIRAS
– Dept. of Statistics and Operations Research, UNC at Chapel Hill,
CB#3260, Hanes Hall, Chapel Hill, NC 27599, USA
dglotzer@email.unc.edu
pipiras@email.unc.edu

VADIM BELENKY, BRADLEY CAMPBELL, TIMOTHY SMITH
– Naval Surface Warfare Center Carderock Division (NSWCCD),
9500 MacArthur Blvd, W. Bethesda, MD 20817, USA
vadim.belenky@navy.mil
bradley.campbell@navy.mil
timothy.c.smith1@navy.mil

Received: May 2015

Revised: January 2016

Accepted: February 2016

Abstract:

- Statistical inference of a probability of exceeding a large critical value is studied in the peaks-over-threshold (POT) approach. The focus is on assessing the performance of the various confidence intervals for the exceedance probability, both for the generalized Pareto distribution used above a selected threshold and in the POT setting for general distributions. The developed confidence intervals perform well in an application to extreme ship motion data. Finally, several approaches to uncertainty reduction are also considered.

Key-Words:

- *exceedance probability; quantiles; confidence intervals; peaks over threshold; generalized Pareto distribution; threshold selection; uncertainty reduction.*

AMS Subject Classification:

- 62G32, 62P30, 62F25, 62G15.

1. INTRODUCTION

We describe first the real-life application which sets the directions and frames the questions pursued in this work (Section 1.1). We then outline the contributions and the structure of this work (Section 1.2).

1.1. Motivation

This work is motivated by applications to ship motions and, more specifically, their stability in irregular seas. See, for example, Lewis [25], Benford [7], Belenky and Sevastianov [4], Neves *et al.* [28] for more information on this research area. When it comes to ship stability, the focus is on several variables characterizing the ship motion including *roll* and *pitch* angles, which are, respectively, the rotational movements around longitudinal (stern-to-bow) and lateral (starboard-to-port side or right-to-left side) axes, as well as vertical and lateral *accelerations* in various locations on the ship. See Figure 1. The ship stability failures are related directly to the exceedance of certain values by these variables. For example, the exceedance of a certain roll angle can lead to a cargo shift (which then can lead to capsizing), loss or damage of cargo in containers on deck, or down-flooding internal volumes of a ship. A large enough acceleration can lead to serious injuries or even death of a crew and passengers, as well as cargo damage. Such occurrences are well known not only among the researchers working in the area but also often make it to the popular media.¹

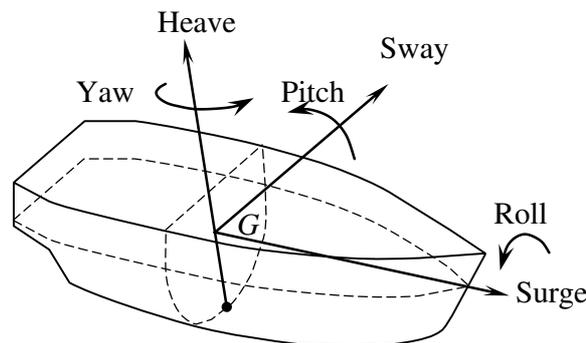


Figure 1: The motions of the ship.

¹Recent examples of accidents related to intact stability failures include: Ro/Ro Ferry *Aratere* on 3rd March 2006 (Maritime New Zealand, 2007), Cruise ship *Pacific Sun* on 30 July 2008 (Marine Accident Investigation Branch, 2009), Ferry *Ariake* on 13 November, 2009 (Transportation Safety Board, 2011), to name but a few.

The measured variables of interest to stability are understandably affected by the *geometry* and *loading* of the ship, the *operational parameters* and the surrounding *sea*. The operational side includes the *heading* (the angle between the vector of ship speed and predominant direction of wave propagation) and the value of *speed* of the ship. The state of the sea is usually described by a spectrum of wave elevations. Note that a wide range of *conditions* (the values of the above descriptors) are possible. What can be expected under a particular condition is often suggested from the understanding of the dynamics governing the ship motion.

An appealing but also critical feature of the research area is the availability of computer programs simulating ship motions, see the recent state-of-the-art review by Reed *et al.* [30]. In this work, we use a fast volume-based ship motion simulation algorithm developed in Weems and Wundrow [37]. The developed code does not incorporate finer hydrodynamics features of ship motions such as the influence of a ship motion on wave pressure field (i.e. wave diffraction and radiation; cf. Large Amplitude Motion Program or LAMP, see Lin and Yue [26]). But it is considered qualitatively representative of ship motions and their extremes. Moreover, the code is fast enough (in fact, the only such realistic method available) to be used in validation, where very long time histories of ship motions are necessary (see Section 3 below).

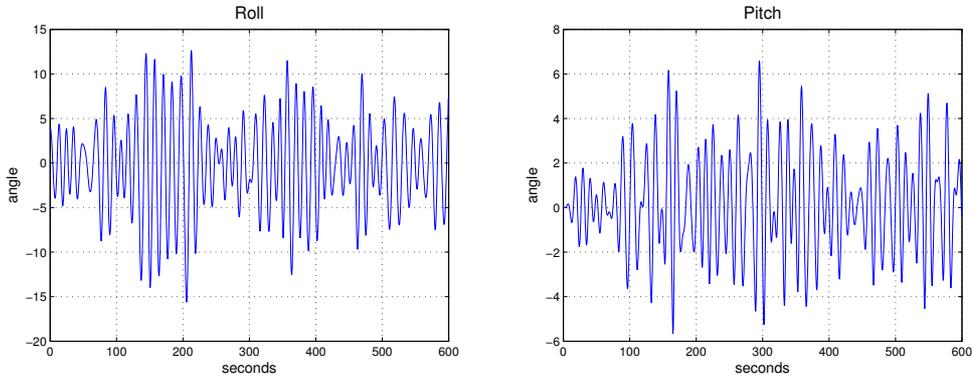


Figure 2: The roll and pitch angle series for 10 minutes.

Figure 2 depicts the time series of roll and pitch angles obtained by the above referenced code for a 10 minute time window at 0.5 second measurement intervals. The ship geometry is that of the ONR tumblehome top (Bishop *et al.* [8]). The heading is at 45 degrees, the speed is 6 knots, the waves are characterized by significant height of 9m and mean zero-crossing period of 10.65s which corresponds to 15s of the modal period, using Bretschneider spectrum in open ocean (Lewis [25]).

A basic problem is to estimate the probability of roll, pitch or other variable of interest exceeding a critical value, as well as to provide a confidence interval. For example, in the condition of Figure 2, one could be interested in the roll angle exceeding 60 degrees (in either positive or negative direction). Inference would have to be made from the roll series of, for example, 100 hours, which would typically not contain such extreme occurrences. Again, the critical angle is often suggested from real-life considerations.

A method suggested for the problem above (and, more specifically, the associated confidence intervals) can be assessed through a validation procedure. The computer code mentioned above can be used to generate millions of hours of ship motion data which would contain exceedances of the target of interest. The “true” exceedance probability can then be estimated directly from this long history of the ship motion. In the validation procedure, the performance of the suggested method could be checked against the “true” exceedance probability at hand. See Section 3 for further details and a solution to the estimation problem.

1.2. Description of work and contributions

A natural mathematical framework to address the problem of estimating exceedance probabilities described above is the peaks-over-threshold (POT) approach (see, for example, Embrechts *et al.* [18], Coles [9], Beirlant *et al.* [2], as well as de Carvalho *et al.* [11], Ferreira and de Haan [19] for more recent related work). According to this approach, the probability of exceeding a given target of interest is computed as the product of the probability of exceeding a smaller threshold and the (conditional) probability of exceeding the target above the threshold. The former probability is estimated simply as the proportion of data above the chosen threshold. The peaks over the threshold are modeled using the *generalized Pareto distribution* (GPD), whose complementary distribution function has the form

$$(1.1) \quad \bar{F}_{\mu,\xi,\sigma}(x) := \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)_+^{-1/\xi} := \begin{cases} \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}, & \mu < x, & \text{if } \xi > 0, \\ e^{-\frac{x-\mu}{\sigma}}, & \mu < x, & \text{if } \xi = 0, \\ \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}, & \mu < x < \mu - \frac{\sigma}{\xi}, & \text{if } \xi < 0, \end{cases}$$

where ξ is the *shape* parameter, σ is the *scale* parameter and μ is a *threshold*. Note that the GPD has an upper bound $(-\sigma/\xi)$ (above the threshold) for a negative shape parameter $\xi < 0$. When $\xi = 0$, the GPD is the usual exponential distribution.

We are interested here in what confidence intervals should be used for an exceedance probability. As indicated above, in the POT approach, this exceedance

probability is the product of two probabilities, one of them being the exceedance probability for GPD. The question then is what confidence intervals should be used for the exceedance probability in the GPD framework. The probability of the GPD exceeding a fixed *target* c (above the threshold), and its estimator are given by:

$$(1.2) \quad p_c = p_c(\xi, \sigma) = \left(1 + \frac{\xi c}{\sigma}\right)^{-1/\xi}, \quad \hat{p}_c = p_c(\hat{\xi}, \hat{\sigma}) = \left(1 + \frac{\hat{\xi} c}{\hat{\sigma}}\right)^{-1/\hat{\xi}},$$

where $\hat{\xi}$ and $\hat{\sigma}$ are some estimators of the shape and scale parameters, respectively. Somewhat surprisingly, the question of confidence intervals for the exceedance probability in (1.2) has apparently not been considered in much depth in the literature on extreme values. The paper by Smith [33], which laid the mathematical foundations for the ML estimators of the GPD, considers the problem of estimating the exceedance probability and provides the asymptotic normality result for the probability estimator (Section 8 of Smith [33]). This can in turn be used for confidence intervals but the normality assumption is not particularly appropriate (see Section 2 below).

Estimation of exceedance probabilities has also been considered by others but with different goals in mind. For example, Smith and Shively [34] are interested in trends for exceedance probabilities. Exceedance probabilities in the spatial context appear in Draghicescu and Ignaccolo [16]. Considerable interest in exceedance (also sometimes referred to as failure) probabilities is when working with multivariate extremes. See, for example, de Haan and Sinha [13], de Haan and de Ronde [12], Heffernan and Tawn [20], Drees and de Haan [17].

Much of the focus in the extreme value analysis, on the other hand, has been on the related inverse problem of quantile estimation (see, for example, Embrechts *et al.* [18], Coles [9], Beirlant *et al.* [2]). The quantiles have been of greater practical interest in many applications driving the extreme value analysis, including finance (VaR calculations), insurance and hydrology (1-in- T years event). A closer look at the confidence intervals for quantiles can be found in Hosking and Wallis [22], Tajvidi [36] and also Section 4.3.3 of Coles [9], Section 5.5 of Beirlant *et al.* [2].

In applications to ship motions, as indicated in Section 1.1, it is common to look at the probabilities of exceeding a particular target rather than quantiles. Though perhaps not surprisingly, the two perspectives are also complementary. In fact, one of our findings is that the confidence intervals for exceedance probabilities perform well if constructed from those for quantiles. Another reason to focus on probabilities rather than quantiles is that probabilities can be aggregated naturally into “lifetime” probabilities, when integrated over a set of conditions of interest (as discussed, for example, in Section 1 of Belenky and Sevastianov [4]).

We study a number of ways to construct confidence intervals for the ex-

ceedance probability of the GPD and, more generally, in the POT framework in Section 2. We consider both direct methods, which are based on the functional form of exceedance probability (1.2) and the sampling distribution of the underlying estimators $\hat{\xi}, \hat{\sigma}$, and indirect (inverse) methods, which construct confidence intervals from those for quantiles.

The application of the considered confidence intervals to ship motions can be found in Section 3. In the validation procedure, the performance of the confidence intervals is analogous to that found under the idealized GPD framework. In particular, the methods recommended under the GPD framework also perform well and best in the application to ship motions. It should also be noted that the proposed solution is the first to address satisfactorily the estimation problem of the exceedance probabilities in ship stability. Some earlier attempts include Benkeny and Campbell [5] who used the Weibull distribution (instead of the GPD) to fit peaks over threshold, and McTaggart [27].

Finally, in Section 4, we discuss the issue of uncertainty (the size of confidence intervals) and its reduction. Conclusions can be found in Section 5.

2. CONFIDENCE INTERVALS FOR EXCEEDANCE PROBABILITIES

2.1. Methods for GPD

We study and assess here several ways to construct confidence intervals for the exceedance probability p_c of the GPD given in (1.2). The probability is estimated through \hat{p}_c in (1.2) where we use the ML estimators $\hat{\xi}$ and $\hat{\sigma}$ computed from the sample y_1, \dots, y_n of size n . The large sample asymptotics of the ML estimators (Smith [33]) is the bivariate normal,

$$(2.1) \quad \sqrt{n} \begin{pmatrix} \hat{\xi} - \xi_0 \\ \hat{\sigma} - \sigma_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, W^{-1}),$$

where ξ_0, σ_0 denote the true values and

$$(2.2) \quad W^{-1} = \begin{pmatrix} 1 + \xi_0 & -\sigma_0 \\ -\sigma_0 & 2\sigma_0^2 \end{pmatrix}.$$

In practice, the limiting covariance matrix can be estimated by replacing ξ_0 and σ_0 with their respective estimators $\hat{\xi}$ and $\hat{\sigma}$. Another common choice is to approximate nW through the observed information matrix

$$(2.3) \quad n\widehat{W} = \begin{pmatrix} -\frac{\partial^2}{\partial \xi^2} l(\xi, \sigma) & -\frac{\partial^2}{\partial \xi \partial \sigma} l(\xi, \sigma) \\ -\frac{\partial^2}{\partial \xi \partial \sigma} l(\xi, \sigma) & -\frac{\partial^2}{\partial \sigma^2} l(\xi, \sigma) \end{pmatrix}_{(\xi, \sigma) = (\hat{\xi}, \hat{\sigma})},$$

where $l(\xi, \sigma) = \sum_{i=1}^n \ln f_{\xi, \sigma}(y_i)$ is the log-likelihood and $f_{\xi, \sigma}(y)$ denotes the density of the GPD. Strictly speaking, the asymptotic result (2.1) holds for $\xi > -1/2$ only (Smith [33]). It should also be noted that other estimation methods than the MLE are possible for ξ_0 and σ_0 . See, for example, a review paper by de Zea Bermudez and Kotz [14, 15] and references therein. Some of these estimators outperform the ML estimators for small samples. For the sample sizes relevant to our problem of interest, the ML estimators seem to perform quite well and, in particular, to be approximately normal as stated in (2.1), and will be used throughout this work.

The exceedance probability $p_c = p_c(\xi, \sigma)$ in (1.2) is a function of ξ and σ , and is estimated through (1.2) by replacing the two parameters ξ and σ by their ML estimates. A confidence interval for p_c can then naturally be obtained through the standard delta method, using the asymptotic result (2.1). This is the approach seemingly adopted by Smith [33], Section 8. However, we found the delta method to perform poorly, in part because p_c can be very small and the normal approximation of \hat{p}_c may be sufficiently wide to include negative values. We have also tried the delta method for $\log p_c$ but the normal approximation did not appear to provide a good fit to $\log \hat{p}_c$. Consequently, we consider below several, potentially more accurate ways to construct confidence intervals for the exceedance probabilities: the normal and lognormal methods, the boundary method, the bootstrap method, the profile likelihood method and the quantile method. The terminology behind the normal, lognormal, boundary and quantile methods are ours.

Normal method: The idea behind the normal method is still to use (2.1), which as mentioned earlier provides a good approximation in practice, but not to linearize the function $p_c(\xi, \sigma)$ (or $\log p_c(\xi, \sigma)$) as in the unsatisfactory delta method. In fact, assuming the bivariate normal approximation for $\hat{\xi}$ and $\hat{\sigma}$ according to (2.1), we can derive the exact distribution of \hat{p}_c as follows. Observe that the distribution function of \hat{p}_c is: for $0 \leq z \leq 1$,

$$\begin{aligned} F_{\hat{p}_c}(z) &= P\left(\left(1 + \frac{\hat{\xi}c}{\hat{\sigma}}\right)^{-1/\hat{\xi}} \leq z\right) \\ &= P\left(\left(1 + \frac{\hat{\xi}c}{\hat{\sigma}}\right)^{-1/\hat{\xi}} \leq z, 1 + \frac{\hat{\xi}c}{\hat{\sigma}} > 0\right) + P\left(1 + \frac{\hat{\xi}c}{\hat{\sigma}} \leq 0\right), \end{aligned}$$

where we use the fact that $\hat{p}_c = 0$ if $1 + \hat{\xi}c/\hat{\sigma} \leq 0$. This can further be expressed as

$$F_{\hat{p}_c}(z) = P\left(\hat{\sigma} \leq \frac{\hat{\xi}c}{z^{-\hat{\xi}} - 1}, \hat{\sigma} > -\hat{\xi}c\right) + P(\hat{\sigma} \leq -\hat{\xi}c),$$

if we assume that $\hat{\sigma}$ takes only positive values. (Note also that $\hat{\xi}/(z^{-\hat{\xi}} - 1) > 0$ for both $\hat{\xi} < 0$ and $\hat{\xi} > 0$.) Note, however, that it is not possible to have $\hat{\sigma} > \hat{\xi}c/(z^{-\hat{\xi}} - 1)$ and $\hat{\sigma} \leq -\hat{\xi}c$. Indeed, this is certainly not possible if $\hat{\xi} > 0$, since then $-\hat{\xi}c < 0$ and $\hat{\xi}c/(z^{-\hat{\xi}} - 1) > 0$. If $\hat{\xi} < 0$, on the other hand, this is not

possible since $-\widehat{\xi}c \leq \widehat{\xi}c/(z^{-\widehat{\xi}} - 1)$ or, equivalently, $z^{-\widehat{\xi}} < 1$. Hence, we also have

$$(2.4) \quad F_{\widehat{p}_c}(z) = P\left(\widehat{\sigma} \leq \frac{\widehat{\xi}c}{z^{-\widehat{\xi}} - 1}\right) = \int_{\sigma \leq \xi c/(z^{-\xi} - 1)} g_{\widehat{\xi}, \widehat{\sigma}}(\xi, \sigma) d\xi d\sigma,$$

where $g_{\widehat{\xi}, \widehat{\sigma}}(\xi, \sigma)$ denotes the bivariate normal density of the limit law (2.1) (replacing ξ_0 and σ_0 by $\widehat{\xi}$ and $\widehat{\sigma}$). In practice, the distribution function $F_{\widehat{p}_c}(z)$ is computed numerically and the $100(1 - \alpha)\%$ confidence interval is set as (z_1, z_2) where $z_j = \inf\{z : F_{\widehat{p}_c}(z) \geq \alpha_j\}$, $j = 1, 2$, where $\alpha_1 = \alpha/2$ and $\alpha_2 = 1 - \alpha/2$. We use the generalized inverse in the last expression since $F_{\widehat{p}_c}(z)$ can have a discontinuity (mass) at $z = 0$.

Lognormal method: In the normal method above, we assumed that $\widehat{\sigma}$ does not take negative values or that, from a practical perspective, the probability of $\widehat{\sigma}$ being negative according to (2.1) is negligible. This may not be the case for smaller values of σ and sample sizes n . A natural way to address this is by parameterizing the GPD through ξ and $\ln \sigma$, instead of σ . The difference is that $\ln \sigma$ now takes possibly negative values. The asymptotic normality result then becomes

$$(2.5) \quad \sqrt{n} \begin{pmatrix} \widehat{\xi} - \xi_0 \\ \widehat{\ln \sigma} - \ln \sigma_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, W_1^{-1}),$$

where

$$(2.6) \quad W_1^{-1} = \text{diag}\{1, \sigma_0^{-1}\} W^{-1} \text{diag}\{1, \sigma_0^{-1}\}.$$

Arguing as in the normal method above, we have

$$(2.7) \quad F_{\widehat{p}_c}(z) = P\left(\widehat{\ln \sigma} \leq \ln \frac{\widehat{\xi}c}{z^{-\widehat{\xi}} - 1}\right) = \int_{\ln \sigma \leq \ln(\xi c/(z^{-\xi} - 1))} g_{\widehat{\xi}, \widehat{\ln \sigma}}(\xi, \ln \sigma) d\xi d \ln \sigma,$$

where $g_{\widehat{\xi}, \widehat{\ln \sigma}}(\xi, \ln \sigma)$ denotes the bivariate normal density of the limit law (2.5). The confidence interval can then be computed as in the normal method above. We shall refer to this as the lognormal method. A nice feature of the normal and lognormal methods is that they provide confidence intervals even in the case when $\widehat{\xi} < 0$ and the target is beyond the estimated support bound $(-\widehat{\sigma}/\widehat{\xi})$.

Boundary method: The normal and lognormal methods described above involve a relatively intensive numerical computation of the integrals (2.4) and (2.7). An approximate confidence interval which is fast to compute and easy to implement, can be constructed through the following boundary method. That is, take the confidence interval as

$$(2.8) \quad \left(\min_{j,k=1,2} p_c(\xi_j, \sigma_k), \max_{j,k=1,2} p_c(\xi_j, \sigma_k) \right),$$

where ξ_1, ξ_2 and σ_1, σ_2 are suitable critical values of the distributions of $\widehat{\xi}$ and $\widehat{\sigma}$, respectively. If $\widehat{\xi}$ and $\widehat{\sigma}$ were asymptotically uncorrelated, it would be natural

to consider $\xi_j = \hat{\xi} \pm C_{\sqrt{\alpha}} \text{se}_{\hat{\xi}}$ and $\sigma_k = \hat{\sigma} \pm C_{\sqrt{\alpha}} \text{se}_{\hat{\sigma}}$, where se stands for standard error, C_{β} denotes the $100(\beta/2)\%$ quantile of the standard normal distribution and $(1 - \alpha)\%$ is the confidence level sought. To account for the correlation between $\hat{\xi}$ and $\hat{\sigma}$, we take

$$(2.9) \quad \begin{pmatrix} \xi_j \\ \sigma_k \end{pmatrix} = V \begin{pmatrix} \xi_{0,j} - \hat{\xi} \\ \sigma_{0,k} - \hat{\sigma} \end{pmatrix} + \begin{pmatrix} \hat{\xi} \\ \hat{\sigma} \end{pmatrix},$$

where $n^{-1}W^{-1} = VDV'$ with a diagonal $D = \text{diag}\{d_1, d_2\}$ and $\xi_{0,j} = \hat{\xi} \pm C_{\sqrt{\alpha}}\sqrt{d_1}$ and $\sigma_{0,k} = \hat{\sigma} \pm C_{\sqrt{\alpha}}\sqrt{d_2}$. Note that the confidence intervals obtained by the boundary method are expected to be conservative. Indeed, the region determined by the points (ξ_j, σ_k) can be thought as the $100(1 - \alpha)\%$ confidence region for the parameters ξ_0 and σ_0 . But since $p_c(\xi, \sigma)$ is not a one-to-one function, there are points (ξ, σ) outside the confidence region for which the value $p_c(\xi, \sigma)$ falls inside the confidence interval (2.8).

Bootstrap method: The bootstrap method is somewhat standard with the confidence interval determined by the $100(\alpha/2)\%$ and $100(1 - \alpha/2)\%$ quantiles of the bootstrap distribution of the exceedance probability.

Profile (likelihood) method: The profile (likelihood) method refers to another standard method to construct confidence intervals based on the profile likelihood. This is achieved by first expressing σ as a function of ξ and the exceedance probability p_c ,

$$\sigma = \frac{\xi c}{p_c^{-\xi} - 1},$$

then parameterizing the likelihood in terms of ξ and p_c (instead of σ), and finally constructing the confidence interval based on the profile likelihood in a standard way. (See Coles [9] for the same approach when estimating a return level, instead of an exceedance probability.) Since the exceedance probability is constrained to be nonnegative, the use of the profile likelihood may be questionable.

Quantile method: Finally, the quantile method actually refers to a set of methods. The basic idea is the following. Exceedance probabilities p (p_c above) are associated with respective return levels (quantiles) x_p (c above) of the GPD distribution. A return level x_p can be estimated with a confidence interval $\hat{x}_p \pm m_p$. Any of the methods discussed above (normal, lognormal, boundary, bootstrap and profile) can be adapted to construct a confidence interval for x_p – the difference being that the function (1.2) is now the return level

$$(2.10) \quad x_p = x_p(\xi, \sigma) = \frac{\sigma}{\xi} (p^{-\xi} - 1).$$

Moreover, the plot of $(-\ln p, \hat{x}_p)$ with added confidence intervals is known as a return level plot (e.g. Coles [9]). To indicate the underlying method used to set confidence intervals for return levels, we will refer to the quantile method as

quantile-boundary, quantile-lognormal, etc. A natural way to set a confidence interval for the exceedance probability p_c of the level c is then

$$(2.11) \quad (p_1, p_2),$$

where $p_1 = \inf\{p : \hat{x}_p + m_p \geq c\}$ and $p_2 = \inf\{p : \hat{x}_p - m_p \geq c\}$ (with $\inf\{\emptyset\} = 0$). See Figure 3. For the parameter values considered below, the functions $\hat{x}_p + m_p$ and $\hat{x}_p - m_p$ are increasing and continuous in the argument $(-\ln p)$. The quantile approach is appealing in that it makes estimation of exceedance probabilities and return levels consistent.

In the reliability context and for a location-scale family of distributions, the quantile approach was studied in Hong *et al.* [21] (see also Section I-C therein for earlier uses of connections between confidence intervals for quantiles and exceedance probabilities).

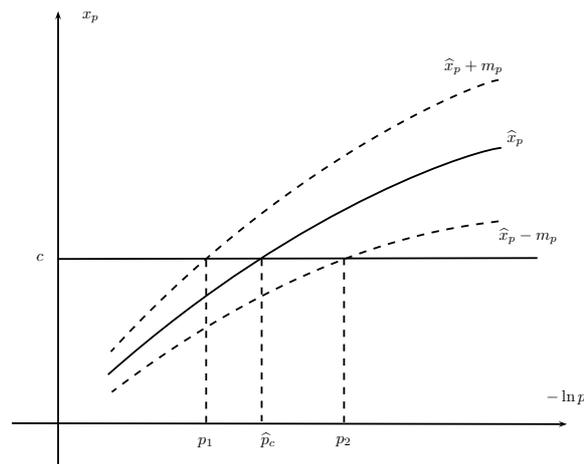


Figure 3: The quantile method to set confidence intervals for exceedance probability.

2.2. Simulation study for GPD

We examine here the confidence intervals proposed in Section 2.1 through a simulation study. The empirical coverage frequencies of the confidence intervals (based on 500 Monte Carlo replications) are reported in Tables 1 and 2 for the sample sizes $n = 100$ and $n = 50$, respectively. The sample size of approximately $n = 100$ is a typical value that we encounter in the application to ship motions described in Section 3 below. The results are also presented for the smaller sample size $n = 50$, since in practice, one does not expect many peaks over a threshold for which the GPD is used as a model.

The first four columns in the tables present the true values of the parameters ξ_0 , σ_0 , and also the target c and the corresponding exceedance probability p_c . The values of $\xi_0 = \pm 1$ are some of the typical values encountered in our application of interest. When $\xi_0 = .6$, the GPD has infinite variance but finite mean. σ_0 is just a scale parameter, which we fix at 1. For the other two true parameters, we fix the exceedance probability p_c and compute the respective target c .

The other columns of the tables correspond to the methods considered. The normal, lognormal and boundary methods use the limiting covariance matrix W^{-1}/n in (2.1). It is approximated by the inverse of the observed information matrix (2.3), which we found to yield better results than using, for example, the expression (2.2) (with ξ_0, σ_0 replaced by $\hat{\xi}, \hat{\sigma}$). The bootstrap method is based on 500 bootstrap replications. Finally, for the quantile methods, we consider three ways to construct confidence intervals for the return levels: lognormal, boundary and profile.

A number of observations can be drawn from Tables 1 and 2. The normal and lognormal methods are slightly anti-conservative, with the lognormal method preferred. The reason for the methods being anti-conservative is the estimation of the limiting covariance matrix W^{-1}/n in (2.1). The intervals have the expected coverage probability if the true covariance matrix (2.2) is taken (the exact coverage probabilities not reported here). As claimed in Section 2.1, the boundary method yields slightly conservative confidence intervals. The bootstrap and profile methods do not work well, especially for the value of ξ_0 close to zero or negative. Again, we suspect that this is due to the fact that the probability cannot be negative. Issues with bootstrap for the GPD were also reported and studied in Tajvidi [36].

Turning to the quantile methods, the quantile-lognormal method is slightly anti-conservative, as is the direct lognormal method. The quantile-boundary method is, on the other hand, slightly conservative. The quantile-profile method seems to perform best, with the coverage probabilities consistently close to the nominal level. Note that the profile-likelihood method for return levels does not have such pronounced limitation of the same method for exceedance probabilities – although it is true that a return level cannot be negative, the confidence interval would rarely reach zero. Note also that the results for $n = 100$ and $n = 50$ are comparable. One notable difference is that the quantile methods become slightly more anti-conservative when the sample size is reduced from $n = 100$ to $n = 50$.

In conclusion, the quantile method based on profile likelihood seems to perform best among the methods considered. The (log)normal and boundary methods, for both direct and indirect (quantile) approaches, can also be recommended but keeping in mind their (anti)conservative nature. Finally, we also note that the direct (log)normal and boundary methods are computationally less intensive compared to the indirect (quantile) methods.

Table 1: Coverage frequencies for confidence intervals when $n = 100$.

true values				direct methods					quantile methods		
ξ_0	σ_0	c	p_c	norm	logn	bound	boot	profl	logn	bound	profl
-.1	1	6.02	10^{-4}	90.4	90.8	96.2	68.2	76.2	92	97	95
		6.84	10^{-5}	95.2	95.6	97.6	65.6	78.8	88.8	96.6	94.6
		7.49	10^{-6}	94	94.6	96	74.6	80.8	91.2	96.8	94.2
.1	1	15.12	10^{-4}	92.6	93.2	98	87	98.4	89.6	97	94.2
		21.62	10^{-5}	90.6	91.2	97.2	82.8	97.6	92.4	98.6	95.2
		29.81	10^{-6}	91.2	92.6	97.6	81.2	97.8	91.2	98.4	94.6
.3	1	49.5	10^{-4}	91.8	92.4	98	89	97.2	89.4	97.4	92.2
		102.08	10^{-5}	88.8	89.2	98.4	86.6	97.6	92.2	97.8	95
		206.99	10^{-6}	93.4	94.2	99	91.8	98.6	92.6	98.4	94.4
.6	1	416.98	10^{-4}	90.8	90.8	97.8	91	93.8	92.4	98.6	94.2
		1665	10^{-5}	92.8	93.2	98	92.6	95.6	92.6	98.4	94
		6633.45	10^{-6}	94	93.8	98.6	92	95.8	93.4	99.4	95.2

Table 2: Coverage frequencies for confidence intervals when $n = 50$.

true values				direct methods					quantile methods		
ξ_0	σ_0	c	p_c	norm	logn	bound	boot	profl	logn	bound	profl
-.1	1	6.02	10^{-4}	90.2	91.8	96.2	65.8	78	83.6	94	92.8
		6.84	10^{-5}	96	96	94.4	69.8	71.8	89.6	95.4	93.4
		7.49	10^{-6}	95.4	95.8	93.8	74.4	75.4	88.2	94.4	92.2
.1	1	15.12	10^{-4}	91.4	92.2	97.6	75.6	98.4	89.9	97.4	92.6
		21.62	10^{-5}	87.6	89	96.6	67	98.2	88.4	96.8	94
		29.81	10^{-6}	90.8	92.2	96.2	70	98.2	88.8	96.4	93
.3	1	49.5	10^{-4}	88.2	91	96	88	99	90	97.2	95.8
		102.08	10^{-5}	85.8	87.8	94.4	82.6	98.6	90.2	96.8	93.4
		206.99	10^{-6}	88.4	90	96.8	85.4	98.2	90.8	96.6	93.8
.6	1	416.98	10^{-4}	80.4	90.6	98.2	88	96.4	89.8	98.4	93.4
		1665	10^{-5}	80.8	91.4	97	89.8	98.2	89.4	96.8	93
		6633.45	10^{-6}	83.4	90.2	97.8	89.4	98	92.6	97.4	92.8

2.3. The POT framework

Suppose now that x_1, \dots, x_N are i.i.d. observations of a general (i.e. non-GPD) random variable X , and that we are interested in estimating the probability $P(X > x_{cr})$ of the variable X exceeding a critical value x_{cr} . Again, in the peaks-

over-threshold (POT) approach, the probability is written as

$$(2.12) \quad \begin{aligned} P(X > x_{cr}) &= P(X > u)P(X > x_{cr}|X > u) \\ &= P(X > u)P(X - u > x_{cr} - u|X > u) =: P_{nr} \cdot P_r, \end{aligned}$$

where u stands for an intermediate threshold, and the subscripts nr and r refer to the non-rare and rare problems, respectively. The non-rare probability is estimated directly from the data as the proportion of data above the threshold u , $\hat{P}_{nr} = \sum_{j=1}^N 1_{\{x_j > u\}}/N$, with the respective confidence interval based on standard binomial calculations. The rare probability is estimated supposing that the peaks over threshold $Y = X - u$ follow a GPD, and setting

$$\hat{P}_r = \hat{p}_{x_{cr}-u},$$

where \hat{p}_c is the exceedance probability (1.2) in the GPD framework, estimated from the data $y_i = x_{i'} - u$ of the peaks exceeding the threshold. The confidence intervals for $P_r = p_{x_{cr}-u}$ are constructed by one of the methods of Section 2.1. The confidence interval for the original exceedance probability $P(X > x_{cr})$ is obtained by multiplying the respective endpoints of the confidence intervals of P_{nr} and P_r .

Threshold selection has been discussed and studied by many authors (for example, a review is given in Scarrott and MacDonald [32]) and is not the focus here. A special feature of the application to ship motions discussed in Section 3 is that the threshold selection should be automated, but with the possibility of closer examination if needed. The automatic selection is naturally sought in the ship motion application because multiple records need to be analyzed for the accuracy that is meaningful for practical applications.

In the automatic selection that we use, the threshold u is selected as the maximum of the thresholds u_{sh} , u_{ms} , u_{me} and u_{rt} chosen by the following four automatic procedures. The thresholds u_{sh} , u_{ms} and u_{me} are selected automatically from the commonly used shape parameter, modified scale parameter and mean excess plots, respectively. For example, the plot of the estimated shape parameters with confidence intervals (against thresholds) should be about constant over the range where GPD fit is appropriate. The threshold u_{sh} is chosen as the smallest threshold for which the horizontal line drawn from the corresponding estimate passes through the confidence intervals of the shape parameter for all the larger thresholds. The thresholds u_{ms} and u_{me} are chosen similarly except that the line in the mean excess plot does not need to be horizontal. The choice of the three thresholds is illustrated in Figure 4, for one of the data sets considered in Section 3 below.

The threshold u_{rt} , on the other hand, is selected following the Reiss and Thomas [31], p. 137, automatic procedure (see also Neves and Fraga Alves [29]). Let $\xi_{k,n}$ be the estimates of the shape parameter ξ based on the k largest values

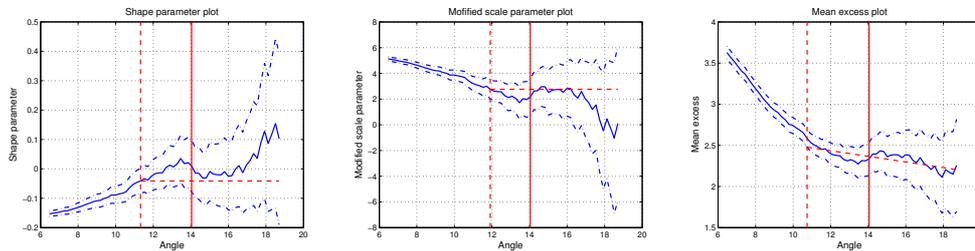


Figure 4: Shape parameter, modified scale parameter and mean excess plots. The vertical dashed line indicates the thresholds chosen with the corresponding (horizontal or arbitrary) lines passing through the confidence intervals for the larger thresholds. The vertical solid line indicates the threshold choice using the Reiss and Thomas method.

of y_i (by using the moment estimation for computational efficiency). Choose k^* as the value that minimizes

$$\frac{1}{k} \sum_{i \leq k} i^\beta |\xi_{k,n} - \text{med}(\xi_{1,n}, \dots, \xi_{k,n})|,$$

where $\beta = 1/2$ (though other values of $\beta < 1/2$ can be considered as well) and med denotes the median. In practice, after the suggestion of Reiss and Thomas, the function above is slightly smoothed. The threshold u_{rt} is then chosen as the k^* largest value of y_i . It is depicted as a vertical solid line in Figure 4 and probably better corresponds to a visually desired choice of threshold. In our experience, the Reiss and Thomas choice most often provides the largest (most conservative) value among the methods considered.

Table 3 presents the empirical coverage frequencies of the confidence intervals constructed through the above POT approach for several non-GPDs. The distributions considered are: the Weibull distribution with the CDF

$$F(x) = 1 - e^{-\lambda x^\tau}, \quad x > 0,$$

with parameters $\lambda > 0, \tau > 0$; the Burr distribution with the CDF

$$F(x) = 1 - \left(\frac{\beta}{\beta + x^\tau}\right)^\lambda, \quad x > 0,$$

with parameters $\lambda > 0, \tau > 0, \beta > 0$; and the reverse Burr distribution with the CDF

$$F(x) = 1 - \left(\frac{\beta}{\beta + (x_+ - x)^{-\tau}}\right)^\lambda, \quad x < x_+,$$

with parameters $\lambda > 0, \tau > 0, \beta > 0$. Two choices of the parameter τ are considered for the Weibull distribution, with $\tau = 1/2$ ($\tau = 2$, resp.) providing heavier

(lighter, resp.) tails than exponential (but both associated with the shape parameter $\xi = 0$ in the POT framework). The Burr distribution has a power-law tail, corresponding to the shape parameter $\xi = 1/(\tau\lambda)$ in the POT framework. Similarly, the reverse Burr distribution has a finite upper bound x_+ , and corresponds to the negative shape parameter $\xi = -1/(\tau\lambda)$ in the POT framework.

Table 3: Empirical coverage frequencies in the non-GPD context using the POT approach.

non-GPD						direct		quantile		
model	parameters	N	n	c	p_c	logn	bound	logn	bound	profl
Weibull	$(\lambda, \tau) = (1, 1/2)$	2000	126	132.5	10^{-5}	90.0	99.2	97.0	99.6	95.0
			123	190.9	10^{-6}	92.2	99.4	94.8	98.8	93.4
	$(\lambda, \tau) = (1, 2)$	2000	194	3.4	10^{-5}	94.6	96.4	90.0	96.4	94.4
			195	3.7	10^{-6}	94.4	97.2	86.8	94.2	93.2
Burr	$(\beta, \tau, \lambda) = (1, 2, 2)$	2000	221	17.8	10^{-5}	95.6	99.2	90.4	97.2	92.8
			210	31.6	10^{-6}	96.4	99.6	87.4	94.8	92.4
Reverse Burr	$(\beta, x_+) = (0.1, 10)$ $(\tau, \lambda) = (2, 2)$	2000	156.5	9.8	10^{-5}	96.8	93	83.4	92.4	90.8
			150	9.9	10^{-6}	98.2	92.2	80.4	90.2	89.0

Under the direct approach in Table 3, the coverage probabilities are reported only for the lognormal and boundary methods. The quantile methods use the proportion of data above the threshold to estimate P_{nr} but do not take the estimation uncertainty of P_{nr} into account. Two of the columns also give the sample size N and the average number of peaks over threshold n . As before, p_c is the exceedance probability and c is the corresponding critical target.

Our goal with Table 3 is not to provide an exhaustive study of the POT approach in the non-GPD framework, but rather to make a few general comments. First, note from the table that the approach works quite well. Second, note that the performance of the considered methods is not as uniformly good as in the GPD context. Thus, the performance of the methods for non-GPDs depends not only on the way to produce confidence intervals above a threshold but also on the non-GPD itself, as well as the (automatic) choice of the threshold.

3. APPLICATION TO EXTREME SHIP MOTIONS

We shall use the POT approach outlined in Section 2.3 to estimate the probability of roll and pitch angle exceeding a critical value. Several issues need to be addressed before we can apply the methods for constructing confidence

intervals discussed in Section 2.3. An important and pressing issue is the presence of temporal dependence as clearly seen from Figure 2. A related issue is also what is meant by an exceedance probability and how it relates to time.

The issue of temporal dependence is addressed through the following envelope approach. Motivated by the periodic nature of a ship motion, the maxima and minima are first found between consecutive zero crossings of the series. These are the positive and negative peaks in the series of interest. The absolute values of the peaks are then connected by a piecewise linear function producing an *envelope* of the series. This is depicted in Figure 5. The left plot includes the original roll series for 5 minutes, with the positive and negative envelope. The right plot depicts the absolute values of the roll and the positive envelope connecting linearly the absolute values of the peaks.

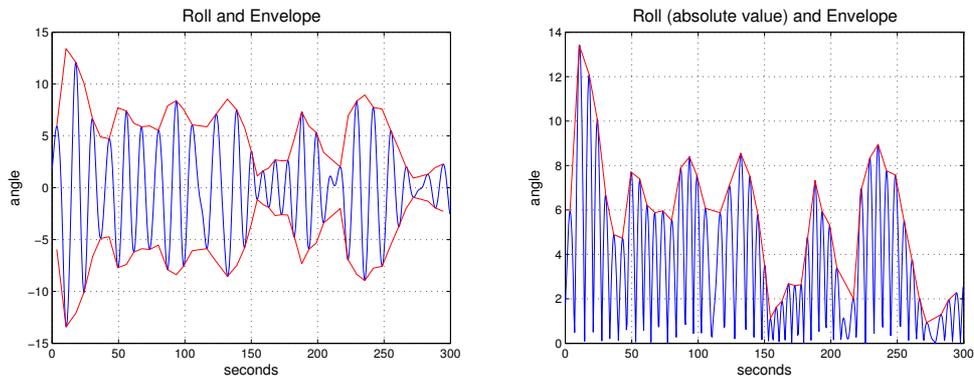


Figure 5: The roll angle series with envelope for 5 minutes. Left: original roll series. Right: roll series in absolute value.

After the envelope is found for the whole roll time series (not just the 5 minutes shown), its average value is computed. Next, the maxima and minima are found in the envelope between consecutive crossings of the average envelope value. These are the envelope peaks above/below the envelope average. This is illustrated in Figure 6, where the envelope average is plotted as a horizontal line and the envelope peaks above/below the envelope average are indicated by small black marks.

Note from Figure 6 that focusing on the envelope peaks (above the average) deals, at least qualitatively, with temporal dependence. That is, the larger values close in time are “clustered” and only the largest values in “clusters” are recorded as envelope peaks. (A closer look at the decorrelation properties of the envelope peak series can be found in a report by Belenky and Campbell [5].) In what follows, we shall work only with the envelope peaks. It is also important to note that the envelope approach is automated. This is particularly convenient when

dealing with multiple conditions and many records.

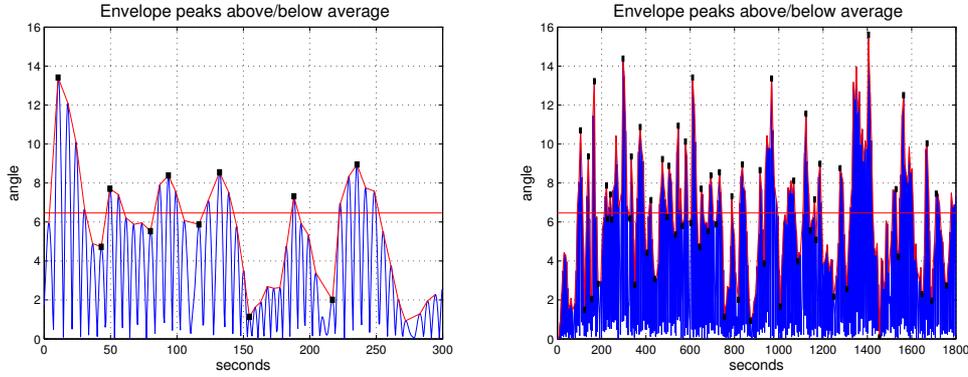


Figure 6: Envelope peaks above/below envelope average.
Left: 5 minutes. Right: 30 minutes.

Focusing on the envelope peaks also simplifies the notion of exceedance and the associated exceedance probabilities. Note that the series of interest will exceed a large target when an envelope peak will exceed the target. It is then natural to think of an exceedance probability as that for the envelope peaks. This is the perspective adopted throughout the paper.

We should also clarify what we mean by probabilities, which are now related to the envelope peaks. Suppose a series contains 1,000 envelope peaks of which 45 exceed a given threshold. Then, the estimated probability is $45/1000 = .045$ of exceeding the threshold. This probability is not informative without a reference to time. Suppose the series is actually recorded over 15 minutes or $15 \cdot 60 = 900$ seconds. It is then more informative to consider the (probability) rate of $45/900 = .05$ envelope peaks (over the threshold) per second. Though we will continue referring to probabilities below, the results will be reported in terms of (probability) rates, rather than probabilities themselves.

If x_1, \dots, x_N are the envelope peaks of the series at hand, the exceedance probability is then estimated with a confidence interval as explained in Section 2.3. The performance of the confidence intervals can be assessed through a validation procedure as follows. The computer code (discussed in Section 1) can be used to generate significantly more series of ship motions, which contain rare events of interest and from which exceedance probabilities can be estimated by direct counting. More specifically, for the same condition used in Figures 2–6, the code was used to generate 115,000 hours of the ship motion. With the target roll angle of $x_{cr} = 60$ degrees, the probability rate of exceedance obtained by direct counting based on rare events from the available records is 7.25×10^{-8} envelope peaks per second (that is, 30 envelope peaks above 60 degrees in 115,000 hours).

This “true” rate estimate can be supplemented by the confidence interval obtained by a standard binomial argument.

A typical given series (record) to make inference from covers only 100 hours and would not contain rare events of interest. For each record, confidence intervals for exceedance probabilities can be computed as in Section 2.3. The confidence intervals can then be assessed by their coverage frequencies of the “true” exceedance probability. This could be examined graphically as in Figure 7 where the lognormal, boundary, quantile-lognormal and quantile-profile confidence intervals are presented for 100 records of the total length of 100 hours. The critical value of interest is the roll of 60 degrees as above. Note that the vertical axis for the probability rate is in the log scale, and that we truncated the confidence intervals and the probability (rate) estimates at a practically negligible probability rate of 10^{-15} . The horizontal dashed lines indicate the confidence bounds for the “true” probability. The small circles are the probability rate estimates.

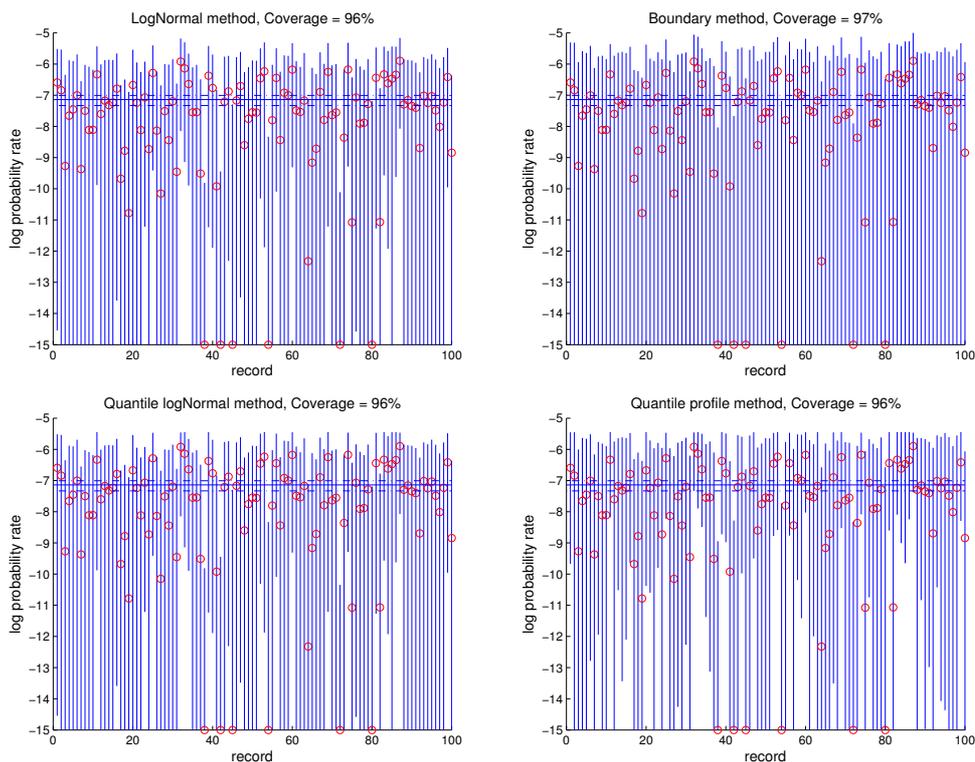


Figure 7: Confidence intervals for 100 records of the length of 100 hours. Roll series for 45° heading and with critical roll angle of 60° . Top left: lognormal method. Top right: boundary method. Bottom left: quantile-lognormal method. Bottom right: quantile-profile method.

For the roll and pitch motion at 45 and 30 degree headings, we also report the coverage frequencies for the methods of Section 2 in Table 4, based on the results in 100 records. The columns under $\hat{\xi}$ and n provide the average estimates of the shape parameter and the number of peaks over threshold. The standard errors are given in parentheses. In the parentheses under the coverage probabilities, we provide the average of the sizes of the suggested confidence intervals above the true value (supposing it is contained), which will be discussed further in Section 4 below.

Table 4: Headings of 30 and 45 degrees. Roll: target is 60 at 45° and 35 at 30°. Pitch: target is 10.

series				direct methods				quantile methods		
varble	head	$\hat{\xi}$	n	logn	bound	boot	profl	logn	bound	profl
roll	45	0.19 (0.13)	96.06 (31.76)	96 (1.19)	97 (1.46)	92 (1.15)	100 (1.52)	96 (1.18)	97 (1.41)	96 (1.40)
	30	0.04 (0.13)	105.03 (46.96)	84 (0.82)	91 (1.13)	76 (0.87)	99 (1.37)	84 (0.82)	91 (1.12)	89 (1.07)
pitch	45	-0.06 (0.11)	107.06 (50.92)	99 (0.62)	100 (0.73)	95 (0.57)	98 (0.73)	99 (0.62)	100 (0.73)	100 (0.74)
	30	-0.08 (0.11)	107.63 (46.61)	97 (0.43)	98 (0.49)	94 (0.41)	96 (0.51)	97 (0.43)	98 (0.48)	98 (0.51)

Note from Table 4 that the performance of the confidence intervals is similar to those in Sections 2.2 and 2.3. Target values are chosen based on available rare events in the large set of records. The performance seems also satisfactory, validating the approach from a practical perspective. The point of using such validation is to show that the approach works on the ship motion data generated by a qualitatively correct computer code, before applying the methods to real or experimental data (where a large number of records are naturally not available). Or, put differently, had the methods not passed the validation, no applied researcher would be confident in using them.

The approach to estimate the exceedance probabilities certainly works in part because of the mathematical justification as outlined in Section 2.3. But this is not the whole story! Another important component to success is related to the length of the record and the physics of the ship motion. The 100-hour records are typical for Naval Architecture purposes. Our results show that these records have sufficiently enough physics to allow one to extrapolate into the tail using the POT framework.

4. UNCERTAINTY REDUCTION

An interesting but also practically important question is whether the uncertainty of the estimators or, equivalently, the size of the confidence intervals can be reduced. For example, in Figure 7, the right (top) endpoints of the confidence intervals are about one order of magnitude above the true value. One order seems acceptable from a practical perspective. But we also encounter conditions where the uncertainty could be as high as two or three orders of magnitude.

Can the uncertainty (or the size of confidence intervals) be reduced? It surely depends on the approach and model used (that is, the POT approach with the two parameter GPD above threshold), the sample size (that is, the number of exceedances above threshold), and the efficiency of the estimation method used. Efficiency cannot be improved considerably since the ML estimators of the GPD parameters are used. But several directions could be explored when it comes to the first two points.

More specifically, in Section 4.1, we study the situation where it may be meaningful to fix a right upper bound when a negative shape parameter is expected. A substantial uncertainty reduction is achieved with this approach but it may not be promising to search for extensions to positive shape parameters, or ways of fixing a bound. Section 4.2 contains a short and, in our view, informative account of several other possibilities that we tried but which did not lead to much of the uncertainty reduction.

4.1. Fixing upper bound

When the shape parameter of a GP distribution is negative, the distribution has a finite upper bound. One direction for uncertainty reduction is to fix this upper bound before estimation based on some physical considerations, e.g. limiting angle for roll after which ship capsizes. Fixing the bound reduces the number of parameters from 2 to 1, so that the reduction of uncertainty is expected.

In applications to ship stability, the pitch motion typically yields a negative shape parameter, as can already be seen from Table 4 (3rd column). There are physical reasons for this phenomenon which, in technical terms, have to do with the form of the stiffness of the pitch motion. Moreover, again for physical reasons, an upper bound for the pitch motion may be expected at about 15° – 20° , as roll stiffness of ONR Tumblehome becomes flat and does not support any resonance excitation. Details of the physics of the pitch motion go beyond the scope of this paper.

From a statistical standpoint, deriving the GPD framework with a fixed upper bound is straightforward. Suppose for notational simplicity that the threshold μ is 0, and denote a fixed upper bound by y_{\max} . When the shape parameter ξ of the GPD (1.1) is negative, the upper bound is given by $(-\sigma/\xi)$. Setting $y_{\max} = -\sigma/\xi$, solving for $\xi = -\sigma/y_{\max}$ and substituting this into (1.1) when $\xi < 0$, we obtain the complementary GPD function with the upper bound y_{\max} ,

$$(4.1) \quad \bar{F}_\sigma(y) = \left(1 - \frac{y}{y_{\max}}\right)^{y_{\max}/\sigma}, \quad 0 < y < y_{\max}.$$

Note that the function (4.1) depends only on the scale parameter σ (with the shape parameter of the GPD being $\xi = -\sigma/y_{\max}$).

The parameter σ in (4.1) can be estimated using ML. Given observations y_1, \dots, y_n (all smaller than y_{\max}), optimizing the log-likelihood

$$\ell(\sigma) = \sum_{i=1}^n \log \left(\frac{1}{\sigma} \left(1 - \frac{y_i}{y_{\max}}\right)^{y_{\max}/\sigma - 1} \right)$$

leads to the ML estimator

$$(4.2) \quad \hat{\sigma} = -\frac{y_{\max}}{n} \sum_{i=1}^n \log \left(1 - \frac{y_i}{y_{\max}}\right).$$

The inverse of the observed information matrix can easily be checked to be

$$(4.3) \quad \left(-\frac{\partial^2 \ell}{\partial \sigma^2} \right)^{-1} \Big|_{\sigma=\hat{\sigma}} = \frac{\hat{\sigma}^2}{n}.$$

A confidence interval for an exceedance probability $p_c = \bar{F}_{\sigma_0}(c)$ can then be given by the boundary method as $(\bar{F}_{\sigma_1}(c), \bar{F}_{\sigma_2}(c))$, where $\sigma_1 = \hat{\sigma} - C_\alpha \hat{\sigma}/\sqrt{n}$ and $\sigma_2 = \hat{\sigma} + C_\alpha \hat{\sigma}/\sqrt{n}$ are two critical values for the distribution of $\hat{\sigma}$ based on (4.3) (with as before, C_α denoting the $100(\alpha/2)\%$ quantile of the standard normal distribution).

Figure 8 compares the confidence intervals for the exceedance probability of the pitch motion at the 30° heading (under the same condition as earlier) obtained through the lognormal method as in Section 3, and the boundary method with the upper bound fixed at 15° as explained above. The left plot in Figure 8 corresponds to the entry of Table 4 under “pitch”, “30” degree heading and “logn” method, with the uncertainty measure of 0.43 in the parentheses. The same measure for the right-plot of Figure 8 is 0.34. The reduction of uncertainty is also evident from Figure 8 itself, with smaller variability of the estimators (red circles) and the sizes of confidence intervals in the right plot.

It should also be noted that the results with the fixed upper bound are not sensitive to the choice of the bound (suggested by physical considerations). For example, fixing the bound at 17° and 20° leads to the same coverage frequency of 99%, with the exception that the uncertainty measure above becomes slightly

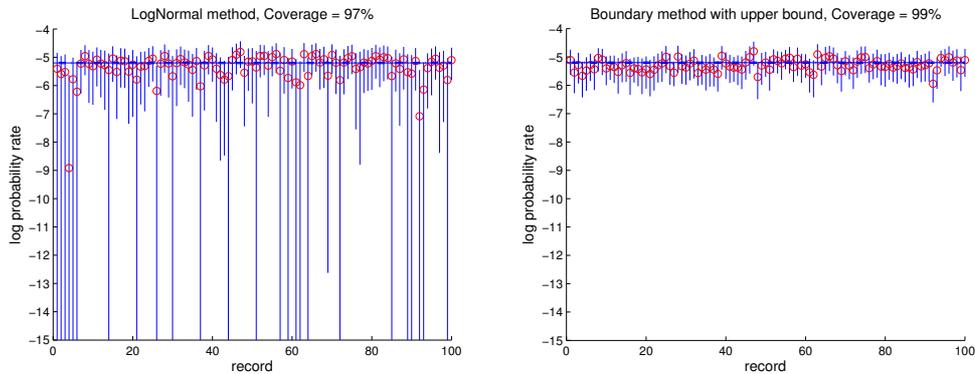


Figure 8: Confidence intervals for 100 records of the length of 100 hours. Pitch 30° . Left: lognormal method. Right: boundary method with fixed upper bound at 15° .

larger, 0.36 and 0.38, respectively. The conclusions are the same for the pitch motion at the 30° heading (not reported here).

Remark 4.1. Whether a similar approach can be developed for a positive shape parameter remains an open question. One idea we entertained was to experiment with truncated GPD models in the spirit of, for example, Aban *et al.* [1], Beirlant *et al.* [3]. (Truncation seems natural because, for example, the roll and pitch angles are bounded by 180 degrees.) But the truncated GPD models did not appear to fit the data well. In Belenky *et al.* [6], we study extreme value methods on mathematically tractable physical models mimicking ship motion dynamics, and expect to gain further insight into the above issues from this approach.

4.2. Other possibilities

We explored or thought about several other possibilities for uncertainty reduction. One natural possibility would be to view the variables describing different conditions as covariates and then pool the data across different conditions by modeling covariates to reduce uncertainty. This idea is particularly relevant in the application of interest here since naval engineers have to take measurements regularly across a range of conditions. The idea also has a sound statistical footing, as developed in Davison and Smith [10] and described, for example, in Chapter 6 of Coles [9].

Following this approach, we have modeled records across a number of head-

ings (e.g. 15° , 22.5° , 30° , 37.5° , 45° degrees). But we generally found the reduction in uncertainty small if any. Some of this is due to a small reduction of uncertainty even under ideal situations (when the model incorporating the covariates is known). The uncertainty in the underlying model for the covariates (entering the POT framework) also plays a role.

Finally, another possibility might be to use some of the more advanced approaches in modeling dependent peaks over threshold, as in e.g. Smith *et al.* [35]. The idea here is that this would seemingly allow for a larger sample size to be considered. Even if the dependence structure is captured correctly by these approaches, we also expect them to lead to little uncertainty reduction. As with the covariates above, we view these approaches as serving different purposes and used to answer different questions.

5. CONCLUSIONS

In this work, we studied the various methods to construct confidence intervals for exceedance probabilities in the peaks-over-threshold approach. The performance of the confidence intervals was assessed through several simulation studies, pointing to the superior performance of some of the considered methods. The developed methods were applied to build confidence intervals for the probabilities of extreme ship motions, leading to satisfactory results overall. Finally, several uncertainty reduction approaches were considered, with a promising solution when a negative shape parameter is expected. Whether uncertainty reduction can be achieved in the case of a positive shape parameter remains an open question.

ACKNOWLEDGMENTS

The authors thank the anonymous Reviewer whose comments helped improve the paper considerably. The authors would also like to thank Prof. Ross Leadbetter (University of North Carolina, Chapel Hill) for valuable discussions about the paper. The work described in this paper has been funded by the Office of Naval Research (ONR) under Dr. Patrick Purtell and Dr. Ki-Han Kim as well as NSWCCD under Dr. Jack Price. Participation of the first author was facilitated by the Summer Faculty Program supported by ONR and managed by Dr. Jack Price. The hospitality of the NSWCCD and the ONR support are gratefully acknowledged.

REFERENCES

- [1] ABAN, I.B.; MEERSCHAERT, M.M. and PANORSKA, A.K., (2006). Parameter estimation for the truncated Pareto distribution, *Journal of the American Statistical Association*, **101**(473), 270–277.
- [2] BEIRLANT, J.; GOEGEBEUR, Y.; TEUGELS, J. and SEGERS, JO. (2004). *Statistics of Extremes*, John Wiley & Sons, Ltd., Chichester.
- [3] BEIRLANT, J.; FRAGA ALVES, I. and GOMES, I. (2015). *Tail fitting for truncated and non-truncated Pareto-type distributions*, Preprint.
- [4] BELENKY, V.L. and SEVASTIANOV, N.B. (2007). *Stability and Safety of Ships: Risk of Capsizing, 2nd edition*, Society of Naval Architects and Marine Engineers.
- [5] BELENKY, V. and CAMPBELL, B.L. (2011). Valuation of the exceedance rate of a stationary stochastic process by statistical extrapolation using the envelope peaks over threshold (EPOT) method, *Naval Surface Warfare Center, Carderock Division*, Hydromechanics Department Report, NSWCCD-50-TR-2011/032.
- [6] BELENKY, V.; GLOTZER, D.; PIPIRAS, V. and SAPSIS, T.P. (2016). *Distribution tail structure and extreme value analysis of constrained piecewise linear oscillators*, Preprint.
- [7] BENFORD, H. (1991). *Naval Architecture for Non-Naval Architects*, Society of Naval Architects and Marine Engineers.
- [8] BISHOP, R.C.; BELKNAP, W.; TURNER, C.; SIMON, B. and KIM, J.H. (2005). Parametric investigation on the influence of GM, roll damping, and above-water form on the roll response of model 5613, *David Taylor Model Basin – NSWCCD*, Report, NSWCCD-50-TR-2005/027.
- [9] COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag London Ltd., London.
- [10] DAVISON, A.C. and SMITH, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society. Series B. Methodological*, **52**(3), with discussion and a reply by the authors.
- [11] DE CARVALHO, M.; TURKMAN, K.F. and RUA, A. (2013). Dynamic threshold modelling and the US business cycle, *Journal of the Royal Statistical Society. Series C. Applied Statistics*, **62**(4), 535–550.
- [12] DE HAAN, L. and DE RONDE, J. (1998). Sea and wind: multivariate extremes at work, *Extremes*, **1**(1), 7–45.
- [13] DE HAAN, L. and SINHA, A.K. (1999). Estimating the probability of a rare event, *The Annals of Statistics*, **27**(2), 732–759.
- [14] DE ZEA BERMUDEZ, P. and KOTZ, S. (2010). Parameter estimation of the generalized Pareto distribution, I, *Journal of Statistical Planning and Inference*, **140**(6), 1353–1373.
- [15] DE ZEA BERMUDEZ, P. and KOTZ, S. (2010). Parameter estimation of the generalized Pareto distribution, II, *Journal of Statistical Planning and Inference*, **140**(6), 1374–1388.

- [16] DRAGHICESCU, D. and IGNACCOLO, R. (2009). Modeling threshold exceedance probabilities of spatially correlated time series, *Electronic Journal of Statistics*, **3**, 149–164.
- [17] DREES, H. and DE HAAN, L. (2014). *Estimating failure probabilities*, Preprint.
- [18] EMBRECHTS, P.; KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extreme Events*, Springer-Verlag, New York.
- [19] FERREIRA, A. and DE HAAN, L. (2014). The generalized Pareto process; with a view towards application and simulation, *Bernoulli*, **20**(4), 1717–1737.
- [20] HEFFERNAN, J.E. and TAWN, J.A. (2010). A conditional approach for multivariate extreme values, *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **66**(3), 497–546, with discussions and reply by the authors.
- [21] HONG, Y.; MEEKER, W.Q. and ESCOBAR, L.A. (2008). The relationship between confidence intervals for failure probabilities and life time quantiles, *Reliability, IEEE Transactions on*, **57**(2), 260–266.
- [22] HOSKING, J.R.M. and WALLIS, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**(3), 339–349.
- [23] KLEIN, J.P. and MOESCHBERGER, M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, Springer.
- [24] LAWLESS, J.F. (1978). Confidence interval estimation for the Weibull and extreme value distributions, *Technometrics*, **20**(4, part 1), 355–365, with a discussion by R.G. Easterling.
- [25] LEWIS, E.V. and SOCIETY OF NAVAL ARCHITECTS AND MARINE ENGINEERS (U.S.) (1990). *Principles of Naval Architecture: Volume III Motions in Waves and Controllability*, Society of Naval Architects and Marine Engineers.
- [26] LIN, W.M. and YUE, D.K.P. (1991). Numerical solutions for large amplitude ship motions in the time-domain, *Proceedings of the 18th Symposium on Naval Hydrodynamics, Ann Arbor*, 41–66.
- [27] MCTAGGART, K.A. (2000). Ship capsize risk in a seaway using fitted distributions to roll maxima, *Journal of Offshore Mechanics and Arctic Engineering*, **122**(2), 141–146.
- [28] NEVES, M.A.S.; BELENKY, V.L.; DE KAT, J.O.; SPYROU, K. and UMEDA, N. (2011). *Contemporary Ideas on Ship Stability and Capsizing in Waves*, Springer.
- [29] NEVES, C. and FRAGA ALVES, M.I. (2004). Reiss and Thomas’ automatic selection of the number of extremes, *Computational Statistics & Data Analysis*, **47**(4), 689–704.
- [30] REED, A.; BECK, R. and BELKNAP, W. (2014). Advances in predictive capability of ship dynamics in waves, *Proceedings of the 30th Symposium of Naval Hydrodynamics, Hobart*.
- [31] REISS, R.D. and THOMAS, M. (2007). *Statistical Analysis of Extreme Values With Applications To Insurance, Finance, Hydrology and Other Fields*, Birkhäuser Verlag, Basel.
- [32] SCARROTT, C. and MACDONALD, A. (2012). A review of extreme value threshold estimation and uncertainty quantification, *REVSTAT*, **10**(1), 33–60.
- [33] SMITH, R.L. (1987). Estimating tails of probability distributions, *The Annals of Statistics*, **15**, 1174–1207.

- [34] SMITH, R.L. and SHIVELY, T.S. (1995). Point process approach to modeling trends in tropospheric ozone based on exceedances of a high threshold, *Atmospheric Environment*, **29**(23), 3489–3499.
- [35] SMITH, R.L.; TAWN, J.A. and COLES, S.G. (1997). Markov chain models for threshold exceedances, *Biometrika*, **84**(2), 249–268.
- [36] TAJVIDI, N. (2003). Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions, *Extremes*, **6**(2), 111–123.
- [37] WEEMS, K. and WUNDROW, D. (2013). Hybrid models for fast time-domain simulation of stability failures in irregular waves with volume-based calculations for Froude-Krylov and hydrostatic forces, *Proceedings of the 13th International Ship Stability Workshop, Brest*.
- [38] ZHAO, W.; PAN, R.; ARON, A. and METTAS, A. (2006). Some properties of confidence bounds on reliability estimation for parts under varying stresses, *Reliability, IEEE Transactions on*, **55**(1), 7–17.