NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

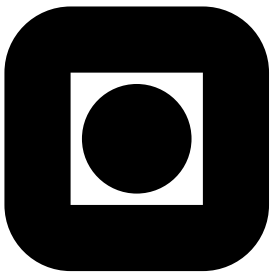
Working paper no. N2-2011
Department of Mathematical Sciences, NTNU

Prediction of Extreme Values by the Average Conditional Exceedance Rate Method

by

Arvid Naess and Oleg Gaidai

PREPRINT
STATISTICS NO. 8/2011



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL
http://www.math.ntnu.no/preprint/statistics/2011/S8-2011.pdf
Arvid Naess has homepage: http://www.math.ntnu.no/∼arvidn
E-mail: arvidn@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway.

Prediction of Extreme Values by the Average Conditional Exceedance Rate Method

Arvid Naess*, Oleg Gaidai†

Abstract

This paper details a new method for extreme value prediction on the basis of a sampled time series. The method is specifically designed to account for statistical dependence between the sampled data points in a precise manner. In fact, if properly used, the new method will provide estimates of the exact extreme value distribution provided by the data. It avoids the problem of having to decluster the data to ensure independence, which is a requisite component in the application of e.g. the standard peaks-over-threshold method. The proposed method also targets the use of sub-asymptotic data to improve prediction accuracy. The method will be demonstrated by application to both synthetic and real data. From a practical point of view, it seems to perform better than the POT and block extremes methods, and, with an appropriate modification, it is directly applicable to nonstationary time series.

Keywords: Extreme value estimation, Sampled time series, Approximation by conditioning, Mean exceedance rate, Monte Carlo simulation.

1 Introduction

Extreme value statistics, even in applications, is generally based on asymptotic results. This is done either by assuming that the epochal extremes, for example yearly extreme wind speeds at a given location, are distributed according to the generalized (asymptotic) extreme value distribution with unknown parameters to be estimated on the basis of the observed data (Coles, 2001; Beirlant et al., 2004). Or it is assumed that the exceedances above high thresholds follow a generalized (asymptotic) Pareto distribution with parameters that are estimated from the data (Coles, 2001; Beirlant et al., 2004; Davison and Smith, 1990; Reiss and Thomas, 2007). The major problem with both of these approaches is that the asymptotic extreme value theory itself cannot be used in practice to decide to what extent it is applicable for the observed data. And

*Department of Mathematical Sciences, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway, e-mail: arvidn@math.ntnu.no

†Marintek A/S, NO-7491 Trondheim, Norway, email: oleg.gaidai@marintek.sintef.no

since statistical tests to decide this issue are rarely precise enough to completely settle this problem, the assumption that a specific asymptotic extreme value distribution is the appropriate distribution for the observed data is based more or less on faith or convenience.

On the other hand, one can reasonably assume that in most cases long time series obtained from practical measurements do contain values that are large enough to provide useful information about extreme events that are truly asymptotic. This cannot be strictly proved in general, of course, but the accumulated experience indicates that asymptotic extreme value distributions provide reasonable, if not always very accurate, predictions when based on measured data. This is amply documented in the vast literature on the subject, and good references to this literature are (Beirlant et al., 2004; Embrechts et al., 1997; Falk et al., 2004). In an effort to improve on the current situation, we have tried to develop an approach to the extreme value prediction problem that is less restrictive and more flexible than the ones based on asymptotic theory. In particular, it is designed to improve on two important aspects of extreme value prediction based on observed data. Firstly, it has the capability to accurately capture the effect of statistical dependence in the data, which opens for the possibility to use all the available data in the analysis. Secondly, it makes it possible to incorporate to a certain extent also the sub-asymptotic part of the data into the extreme value prediction, which is of some importance for accurate prediction. We have used the proposed methods on a wide variety of prediction problems, and our experience is that they represent a viable addition to the toolbox of methods for extreme value prediction.

2 Cascade of Conditioning Approximations

Consider a stochastic process $Z(t)$, which has been observed over a time interval, $(0, T)$ say. Assume that values $X_1, \ldots, X_N$, which have been derived from the observed process, are allocated to the discrete times $t_1, \ldots, t_N$ in $(0, T)$. This could be simply the observed values of $Z(t)$ at each t_j, $j = 1, \ldots, N$, or it could be average values or peak values over smaller time intervals centered at the t_j's. Our goal in this paper is to accurately determine the distribution function of the extreme value $M_N = \max\{X_j \, ; j = 1, \ldots, N\}$. Specifically, we want to estimate $P(\eta) = \text{Prob}(M_N \leq \eta)$ accurately for large values of η. An underlying premise for the development in this paper is that a rational approach to the study of the extreme values of the sampled time series is to consider exceedances of the individual random variables X_j above given thresholds, as in classical extreme value theory. The alternative approach of considering the exceedances by upcrossing of given thresholds by a continuous stochastic process has been developed in (Naess and Gaidai, 2008; Naess et al., 2007) along lines similar to that adopted here. The approach taken in the present paper seems to be the

appropriate way to deal with the recorded data time series of, for example, the hourly or daily largest wind speeds observed at a given location.

From the definition of $P(\eta)$ it follows that

$$\begin{aligned} P(\eta) &= \text{Prob}(M_N \leq \eta) = \text{Prob}\{X_1 \leq \eta, \ldots, X_N \leq \eta\} \\ &= \text{Prob}\{X_N \leq \eta | X_1 \leq \eta, \ldots, X_{N-1} \leq \eta\} \cdot \text{Prob}\{X_1 \leq \eta, \ldots, X_{N-1} \leq \eta\} \\ &= \prod_{j=2}^{N} \text{Prob}\{X_j \leq \eta | X_1 \leq \eta, \ldots, X_{j-1} \leq \eta\} \cdot \text{Prob}(X_1 \leq \eta) \end{aligned} \tag{1}$$

In general, the variables X_j are statistically dependent. Hence, instead of assuming that all the X_j are statistically independent, which leads to the classical approximation

$$P(\eta) \approx \prod_{j=1}^{N} \text{Prob}(X_j \leq \eta), \tag{2}$$

the following one-step memory approximation will to a certain extent account for the dependence between the X_j's,

$$\text{Prob}\{X_j \leq \eta | X_1 \leq \eta, \ldots, X_{j-1} \leq \eta\} \approx \text{Prob}\{X_j \leq \eta | X_{j-1} \leq \eta\}, \tag{3}$$

for $2 \leq j \leq N$. This approximation can be extended to

$$\text{Prob}\{X_j \leq \eta | X_1 \leq \eta, \ldots, X_{j-1} \leq \eta\} \approx \text{Prob}\{X_j \leq \eta | X_{j-2} \leq \eta, X_{j-1} \leq \eta\}, \tag{4}$$

where $3 \leq j \leq N$, and so on. It should be noted that the one-step memory approximation adopted above is not a Markov chain approximation (Smith, 1992; Coles, 1994), nor do the k-step memory approximations lead to kth-order Markov chains (Yun, 1998).

Eqs. (3) and (4) represent refinements of the independence assumption. One would expect that such approximations would be able to capture the effect of statistical dependence between neighboring data in the time series with increasing accuracy. As will be seen in sections 8 and 9, $P(\eta)$ computed using Eq. (4) is often quite close to the value obtained using Eq. (3). This indicates that in practice, Eq. (3) is oftentimes able to capture the effect of statistical dependence in e.g. wind speed data with good accuracy. This approximation was introduced in (Naess, 1985, 1990). However, there is no noticeable increase of numerical effort by using Eq. (4), or its further refinements by including three or more preceding data. And, as demonstrated below, this cascade of refinements will provide a very useful diagnostic tool to highlight the importance of statistical dependence on the extreme value predictions.

Combining Eq. (1) with Eq. (3), the following relation is obtained

$$P(\eta) \approx \frac{\prod_{j=2}^{N} p_{2j}(\eta)}{\prod_{j=2}^{N-1} p_{1j}(\eta)} \tag{5}$$

where we have introduced the notation $p_{kj}(\eta) = \text{Prob}\{X_{j-k+1} \leq \eta, \ldots, X_j \leq \eta\}$ for $j \geq k$.

It is of interest to have a closer look at the values for $P(\eta)$ obtained by using Eq. (5) as compared to Eq. (2). Now, Eq. (2) can be rewritten in the form

$$P(\eta) \approx \prod_{j=1}^{N} \big(1 - \alpha_{1j}(\eta)\big), \tag{6}$$

where

$$\alpha_{1j}(\eta) = \text{Prob}\{X_j > \eta\} = 1 - p_{1j}(\eta). \tag{7}$$

Then

$$P(\eta) \approx P_1(\eta) = \exp\Big(-\sum_{j=1}^{N} \alpha_{1j}(\eta)\Big), \tag{8}$$

where $P_1(\eta)$ is defined by the last equality in Eq. (8).

Alternatively, Eq. (5) gives

$$P(\eta) \approx \prod_{j=2}^{N} \big(1 - \alpha_{2j}(\eta)\big)\, p_{11}(\eta), \tag{9}$$

where $\alpha_{kj}(\eta) = 1 - p_{kj}(\eta)/p_{k-1,j-1}(\eta)$, for $j \geq k \geq 2$. That is

$$\alpha_{kj}(\eta) = \text{Prob}\{X_j > \eta \,|\, X_{j-k+1} \leq \eta, \ldots, X_{j-1} \leq \eta\} \tag{10}$$

denotes the exceedance probability conditional on $k-1$ previous non-exceedances. From Eq. (9) it is obtained that,

$$P(\eta) \approx P_2(\eta) = \exp\Big(-\sum_{j=2}^{N} \alpha_{2j}(\eta) - \alpha_{11}(\eta)\Big), \tag{11}$$

since $p_{11}(\eta) \approx \exp(-\alpha_{11}(\eta))$.

Conditioning on the two previous observations X_{j-2}, X_{j-1} preceding X_j gives

$$P(\eta) \approx P_3(\eta) = \exp\Big(-\sum_{j=3}^{N} \alpha_{3j}(\eta) - \alpha_{22}(\eta) - \alpha_{11}(\eta)\Big), \tag{12}$$

while conditioning on three prior observations leads to the equation

$$P(\eta) \approx P_4(\eta) = \exp\Big(-\sum_{j=4}^{N} \alpha_{4j}(\eta) - \alpha_{33}(\eta) - \alpha_{22}(\eta) - \alpha_{11}(\eta)\Big), \tag{13}$$

and so on. Therefore, extreme value prediction by the conditioning approach described above reduces to estimation of (combinations) of the $\alpha_{kj}(\eta)$ functions. For most practical applications $N >> k$, so that $\sum_{j=1}^{k-1} \alpha_{jj}(\eta)$ is effectively

negligible compared to $\sum_{j=k}^{N} \alpha_{kj}(\eta)$. Hence, for simplicity, we shall adopt the approximation,

$$P_k(\eta) = \exp\Big(-\sum_{j=k}^{N} \alpha_{kj}(\eta) \Big) \ , \ \ k \geq 1 \, . \tag{14}$$

Going back to Eq. (8), and the definition of $\alpha_{1j}(\eta)$, it follows that $\sum_{j=1}^{N} \alpha_{1j}(\eta)$ is equal to the expected number of exceedances of the threshold η during the time interval $(0, T)$. Eq. (8) therefore expresses the approximation that the stream of exceedance events constitute a (non-stationary) Poisson process. This opens for an understanding of Eq. (11) and subsequent approximations by interpreting the expressions $\sum_{j=k}^{N} \alpha_{kj}(\eta)$ as the expected effective number of independent exceedance events provided by conditioning on $k-1$ previous observations.

3 Empirical Estimation of the Average Conditional Exceedance Rates

It is expedient to introduce the concept of average conditional exceedance rate (ACER) of order k as follows,

$$\varepsilon_k(\eta) = \frac{1}{N-k+1} \sum_{j=k}^{N} \alpha_{kj}(\eta) \, , \ k = 1, 2, \ldots \tag{15}$$

In general, this ACER function also depends on the number of data points N.

In practice there are typically two scenarios for the underlying process $Z(t)$. Either we may consider it to be a stationary process, or, in fact, even an ergodic process. The alternative is to view $Z(t)$ as a process that depends on certain parameters whose variation in time may be modelled as an ergodic process in its own right. For each set of values of the parameters, the premise is that $Z(t)$ can then be modelled as an ergodic process. This would be the scenario that can be used to model long-term statistics (Naess, 1984; Schall et al., 1991).

For both these scenarios, the empirical estimation of the ACER function $\varepsilon_k(\eta)$ proceeds in a completely analogous way by counting the total number of favourable incidents, that is, exceedances combined with the requisite number of preceding non-exceedances, for the total data time series and then finally dividing by $N-k+1 \approx N$. This can be shown to apply for the long-term situation.

A few more details on the numerical estimation of $\varepsilon_k(\eta)$ for $k \geq 2$ may be appropriate. We start by introducing the following random functions,

$$A_{kj}(\eta) = \mathbf{1}\{X_j > \eta, X_{j-1} \leq \eta, \ldots, X_{j-k+1} \leq \eta\} \, , \ j = k, \ldots, N, \ k = 2, 3, \ldots \tag{16}$$

and

$$B_{kj}(\eta) = \mathbf{1}\{X_{j-1} \le \eta, \ldots, X_{j-k+1} \le \eta\}\,,\ j = k, \ldots, N,\ k = 2, \ldots\,, \tag{17}$$

where $\mathbf{1}\{\mathcal{A}\}$ denotes the indicator function of some event $\mathcal{A}$. Then

$$\alpha_{kj}(\eta) = \frac{\mathbb{E}[A_{kj}(\eta)]}{\mathbb{E}[B_{kj}(\eta)]}\,,\ j = k, \ldots, N,\ k = 2, \ldots\,, \tag{18}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator. Assuming an ergodic process, then obviously $\varepsilon_k(\eta) = \alpha_{kk}(\eta) = \ldots = \alpha_{kN}(\eta)$, and by replacing ensemble means with corresponding time averages, it may be assumed that for the time series at hand

$$\varepsilon_k(\eta) = \lim_{N\to\infty} \frac{\sum_{j=k}^{N} a_{kj}(\eta)}{\sum_{j=k}^{N} b_{kj}(\eta)}\,, \tag{19}$$

where $a_{kj}(\eta)$ and $b_{kj}(\eta)$ are the realized values of $A_{kj}(\eta)$ and $B_{kj}(\eta)$, respectively, for the observed time series.

For multiple recorded stationary time series, the sample estimate of $\varepsilon_k(\eta)$ would be,

$$\hat{\varepsilon}_k(\eta) = \frac{1}{R}\sum_{r=1}^{R} \hat{\varepsilon}_k^{(r)}(\eta)\,, \tag{20}$$

where R is the number of realizations (samples), and

$$\hat{\varepsilon}_k^{(r)}(\eta) = \frac{\sum_{j=k}^{N} a_{kj}^{(r)}(\eta)}{\sum_{j=k}^{N} b_{kj}^{(r)}(\eta)}\,, \tag{21}$$

where the index (r) refers to realization no. r.

Clearly, $\lim_{\eta\to\infty} \mathbb{E}[B_{kj}(\eta)] = 1$. Hence, $\lim_{\eta\to\infty} \tilde{\varepsilon}_k(\eta)/\varepsilon_k(\eta) = 1$, where

$$\tilde{\varepsilon}_k(\eta) = \frac{\sum_{j=k}^{N} \mathbb{E}[A_{kj}(\eta)]}{N-k+1}\,. \tag{22}$$

The advantage of using the modified ACER function $\tilde{\varepsilon}_k(\eta)$ for $k \ge 2$ is that it is easier to use for non-stationary or long-term statistics than $\varepsilon_k(\eta)$. Since our focus is on the values of the ACER functions at the extreme levels, we may use any function that provides correct predictions of the appropriate ACER function at these extreme levels.

To see why Eq. (22) may be applicable for nonstationary time series, it is recognized that

$$\begin{aligned} P(\eta) &\approx \exp\Big(-\sum_{j=k}^{N} \alpha_{kj}(\eta)\Big) = \exp\Big(-\sum_{j=k}^{N} \frac{\mathbb{E}[A_{kj}(\eta)]}{\mathbb{E}[B_{kj}(\eta)]}\Big) \\ &\underset{\eta\to\infty}{\simeq} \exp\Big(-\sum_{j=k}^{N} \mathbb{E}[A_{kj}(\eta)]\Big)\,. \end{aligned} \tag{23}$$

If the time series can be segmented into K blocks such that $\mathbb{E}[A_{kj}(\eta)]$ remains approximately constant within each block and such that $\sum_{j\in C_i}\mathbb{E}[A_{kj}(\eta)] \approx \sum_{j\in C_i} a_{kj}(\eta)$ for a sufficient range of η-values, where C_i denotes the set of indices for block no. i, $i = 1,\ldots,K$, then $\sum_{j=k}^{N}\mathbb{E}[A_{kj}(\eta)] \approx \sum_{j=k}^{N} a_{kj}(\eta)$. Hence,

$$P(\eta) \approx \exp\big(-(N-k+1)\hat{\varepsilon}_k(\eta)\big)\,, \tag{24}$$

where

$$\hat{\varepsilon}_k(\eta) = \frac{1}{N-k+1}\sum_{j=k}^{N} a_{kj}(\eta)\,. \tag{25}$$

It is of interest to note what events are actually counted for the estimation of the various $\varepsilon_k(\eta)$, $k \geq 2$. Let us start with $\varepsilon_2(\eta)$. It follows from the definition of $\varepsilon_2(\eta)$ that $\varepsilon_2(\eta)\,(N-1)$ can be interpreted as the expected number of exceedances above the level η satisfying the condition that an exceedance is counted only if it is immediately preceded by a non-exceedance. A reinterpretation of this is that $\varepsilon_2(\eta)\,(N-1)$ equals the average number of clumps of exceedances above η for the realizations considered, where a clump of exceedances is defined as a maximum number of consecutive exceedances above η. In general, $\varepsilon_k(\eta)\,(N-k+1)$ then equals the average number of clumps of exceedances above η separated by at least $k-1$ non-exceedances. If the time series analysed is obtained by extracting local peak values from a narrow band response process, it is interesting to note the similarity between the ACER approximations and the envelope approximations for extreme value prediction (Naess and Gaidai, 2008; Vanmarcke, 1975).

Now, let us look at the problem of estimating a confidence interval for $\varepsilon_k(\eta)$, assuming a stationary time series. The sample standard deviation $\hat{s}_k(\eta)$ can be estimated by the standard formula,

$$\hat{s}_k(\eta)^2 = \frac{1}{R-1}\sum_{r=1}^{R}\Big(\hat{\varepsilon}_k^{(r)}(\eta) - \hat{\varepsilon}_k(\eta)\Big)^2\,. \tag{26}$$

Assuming that realizations are independent, for a suitable number R, e.g. $R \geq 20$, Eq. (26) leads to a good approximation of the 95 % confidence interval $\mathrm{CI} = \big(\mathrm{CI}^-(\eta), \mathrm{CI}^+(\eta)\big)$ for the value $\varepsilon_k(\eta)$, where

$$\mathrm{CI}^{\pm}(\eta) = \hat{\varepsilon}_k(\eta) \pm 1.96\,\hat{s}_k(\eta)/\sqrt{R}\,. \tag{27}$$

Alternatively, and which also applies to the non-stationary case, it is consistent with the adopted approach to assume that the stream of conditional exceedances over a threshold η constitute a Poisson process, possibly non-homogeneous. Hence, the variance of the estimator $\hat{E}_k(\eta)$ of $\tilde{\varepsilon}_k(\eta)$, where

$$\hat{E}_k(\eta) = \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{N-k+1}\,, \tag{28}$$

is $\mathrm{Var}[\hat{E}_k(\eta)] = \tilde{\varepsilon}_k(\eta)$. Therefore, for high levels η, the approximate limits of a 95 % confidence interval of $\tilde{\varepsilon}_k(\eta)$, and also $\varepsilon_k(\eta)$, can be written as,

$$\mathrm{CI}^{\pm}(\eta) = \hat{\varepsilon}_k(\eta)\big(1 \pm \frac{1.96}{\sqrt{(N-k+1)\hat{\varepsilon}_k(\eta)}}\big)\,. \tag{29}$$

4 Prediction of Extremes for the Asymptotic Gumbel Case

Part of the approach to extreme value prediction presented in this paper was originally derived for a time series with an asymptotic extreme value distribution which could be assumed to be of the Gumbel type, cf. (Naess and Gaidai, 2009). The implication of this assumption on the possible sub-asymptotic functional forms of $\varepsilon_k(\eta)$ cannot easily be decided in any detail. However, using the asymptotic form as a guide, it is assumed that the behaviour of the mean exceedance rate in the tail is dominated by a function of the form $\exp\{-a(\eta-b)^c\}$ $(\eta \geq \eta_1 \geq b)$ where a, b and c are suitable constants, and η_1 is an appropriately chosen tail marker. Hence, it will be assumed that,

$$\varepsilon_k(\eta) = q_k(\eta)\, \exp\{-a_k(\eta - b_k)^{c_k}\}\,,\ \ \eta \geq \eta_1, \tag{30}$$

where the function $q_k(\eta)$ is slowly varying compared with the exponential function $\exp\{-a_k(\eta-b_k)^{c_k}\}$ and a_k, b_k, and c_k are suitable constants, that in general will be dependent on k. Note that the value $c_k = 1$ and $q_k(\eta) =$ constant corresponds to the asymptotic Gumbel case.

From Eq. (30) it follows that,

$$-\log\big|\log\big(\varepsilon_k(\eta)/q_k(\eta)\big)\big| = -\,c_k \log(\eta - b_k) - \log(a_k). \tag{31}$$

Therefore, under the assumptions made, a plot of $-\log\big|\log\big(\varepsilon_k(\eta)/q_k(\eta)\big)\big|$ versus $\log(\eta - b_k)$ will exhibit a perfectly linear tail behaviour.

It is realized that if the function $q_k(\eta)$ could be replaced by a constant value, q_k say, one would immediately be in a position to apply a linear extrapolation strategy for deep tail prediction problems. In general, $q_k(\eta)$ is not constant, but its variation in the tail region is often sufficiently slow to allow for its replacement by a constant, possibly by adjusting the tail marker η_1. The proposed statistical approach to the prediction of extreme values is therefore based on the assumption that we can write,

$$\varepsilon_k(\eta) = q_k\, \exp\{-a_k(\eta - b_k)^{c_k}\}\,,\ \ \eta \geq \eta_1, \tag{32}$$

where a_k, b_k, c_k and q_k are appropriately chosen constants. In a certain sense this is a minimal class of parametric functions that can be used for this purpose which makes it possible to achieve three important goals. Firstly, the

parametric class contains the asymptotic form given by $c_k = q_k = 1$ as a special case. Secondly, the class is flexible enough to capture to a certain extent sub-asymptotic behaviour of any extreme value distribution that is asymptotically Gumbel. Thirdly, the parametric functions agree with a wide range of known special cases, of which a very important example is the extreme value distribution for a stationary Gaussian process, which has $c_k = 2$.

The viability of this approach has been successfully demonstrated by the authors for mean up-crossing rate estimation for extreme value statistics of the response processes related to a wide range of different dynamical systems, cf. (Naess and Gaidai, 2008; Naess et al., 2007).

Since the linearity in the plotting procedure described above is dependent on an appropriate choice of the parameters (b_k, q_k), it is of interest to discuss this issue in some detail.

To avoid problems with our definition of weight factors to be introduced below, we cut from consideration the very tail of the data, where uncertainty is too high according to the following criterion. As a practical procedure we suggest to neglect data points, where the relative confidence band width is greater than some constant δ, that is,

$$\frac{1.96\hat{s}_k(\xi)/\sqrt{R}}{\varepsilon_k(\eta)} > \delta \tag{33}$$

where the value chosen for δ is dependent on the actual 'roughness' of the data tail, but its value would typically be in the interval $(0.5, 1)$. Next, we come to the estimation of the predicted response level and its 95% confidence interval.

First, the tail marker η_1 is identified from visual inspection of the log plot $(\eta, \log \varepsilon_k(\eta))$. The value chosen for η_1 corresponds to the beginning of regular tail behaviour in a sense to be discussed below. Next, initial estimates for b and q are found by the procedure to linearize the tail on the transformed scale. Assuming that initial values of (b, q) have been identified, the initial value of the parameters a and c would then generally be determined from the initial 'optimal' straight line $-cx - \log a$ $(x = \log(\eta - b))$ approximating the data tail on the transformed plot (31).

Instead of doing the optimization directly from the loglog-log plot, which is appealing in the sense that it directly involves only two parameters, a more robust optimization may in fact be obtained by doing it on the log level even if the optimization has to be carried out with respect to all four parameters a, b, c, q. The optimal choice of parameters would then be obtained by minimizing the following mean square error function with respect to all four arguments (the subscript k, if it applies, is suppressed),

$$F(a, b, c, q) = \sum_{j=1}^{J} w_j \left| \log \hat{\varepsilon}(\eta_j) - \log q + a(\eta_j - b)^c \right|^2, \tag{34}$$

where $\eta_1 < \ldots < \eta_J$ denotes the levels where the ACER function has been estimated, w_j denotes a weight factor that puts more emphasis on the more reliably estimated $\hat{\varepsilon}(\eta_j)$. The choice of weight factor is to some extent arbitrary. We have previously used $w_j = \big(\log C^+(\eta_j) - \log C^-(\eta_j)\big)^{-\theta}$ with $\theta = 1$ and 2, combined with a Levenberg-Marquardt least squares optimization method (Gill et al., 1981). This has usually worked well provided reasonable, initial values for the parameters were chosen. Note that the form of w_j puts some restriction on the use of the data. Usually, there is a level η_j beyond which w_j is no longer defined, that is, $C^-(\eta_j)$ becomes negative. Hence, the summation in Eq. (34) has to stop before that happens. Also, the data should be preconditioned by establishing the tail marker η_1 in a sensible way.

A note of caution: When the parameter c is equal to 1.0 or close to it, the optimization problem becomes ill-defined or close to ill-defined. It is seen that when $c = 1.0$, there is an infinity of (b, q) values that gives exactly the same value of $F(a, b, c, q)$. Hence, there is no well defined optimum in parameter space. There are simply too many parameters. This problem is alleviated by fixing the q-value, and the obvious choice is $q = 1$.

Although the Levenberg-Marquardt method generally works well with four or, when appropriate, three parameters, we have also developed a more direct and transparent optimization method for the problem at hand. It is realized by scrutinizing Eq. (34) that if b and c are fixed, the optimization problem reduces to a standard weighted linear regression problem. That is, with both b and c fixed, the optimal values of a and $\log q$ are found using closed form weighted linear regression formulas in terms of w_j, $y_j = \log \varepsilon(\eta_j)$ and $x_j = (\eta_j - b)^c$. In that light, it can also be concluded that the best linear unbiased estimators (BLUE) are obtained for $w_j = \sigma_{yj}^{-2}$, where $\sigma_{yj}^2 = \mathrm{Var}[y_j]$ (empirical) (Draper and Smith, 1998; Montgomery et al., 2002). Unfortunately, this is not a very practical weight factor for the kind of problem we have here because the summation in Eq. (34) then typically would have to stop at undesirably small values of η_j.

It is obtained that the optimal values of a and q are given by the relations,

$$a^*(b, c) = -\frac{\sum_{j=1}^N w_j (x_j - \overline{x})(y_j - \overline{y})}{\sum_{j=1}^N w_j (x_j - \overline{x})^2}\,, \tag{35}$$

and

$$\log q^*(b, c) = \overline{y} + a^*(b, c)\overline{x}\,, \tag{36}$$

where $\overline{x} = \sum_{j=1}^N w_j x_j / \sum_{j=1}^N w_j$, with a similar definition of $\overline{y}$.

To calculate the final optimal set of parameters, one may use the Levenberg-Marquardt method on the function $\tilde{F}(b, c) = F(a^*(b, c), b, c, q^*(b, c))$ to find the optimal values b^* and c^*, and then use Eqs. (35) and (36) to calculate the corresponding a^* and q^*.

A practical approach that could be adopted is to get a first idea of the values of the parameters a, b, c, q by having a look at the loglog-log plot. These values may then be used as starting values for the Levenberg-Marquardt algorithm.

For construction of a confidence interval for the predicted, deep tail extreme value given by a particular ACER function as provided by the fitted parametric curve, the empirical confidence band is reanchored to the fitted curve by centering the individual confidence intervals for the point estimates of the ACER function on the fitted curve. Under the premise that the specified class of parametric curves fully describes the behaviour of the ACER functions in the tail, parametric curves are fitted as described above to the boundaries of the reanchored confidence band. These curves are used to determine a confidence interval of the predicted extreme value. As a final point, it has been observed that the predicted value is not very sensitive to the choice of η_1, provided it is chosen with some care.

5 Prediction of Extremes for the General Case

For independent data in the general case, the ACER function $\varepsilon_1(\eta)$ can be expressed asymptotically as,

$$\varepsilon_1(\eta) \underset{\eta\to\infty}{\simeq} \left[1+\xi\big(a(\eta-b)\big)\right]^{-\frac{1}{\xi}} , \tag{37}$$

where $a > 0$, b, ξ are constants.

Again, the implication of this assumption on the possible sub-asymptotic functional forms of $\varepsilon_k(\eta)$ in the general case is not a trivial matter. The approach we have chosen is to assume that the class of parametric functions needed for the prediction of extreme values for the general case can be modelled on the relation between the Gumbel distribution and the general extreme value distribution. The behaviour of the mean exceedance rate in the sub-asymptotic part of the tail is therefore assumed to follow a function largely of the form $\left[1+\xi\big(a(\eta-b)^c\big)\right]^{-\frac{1}{\xi}}$ $(\eta \geq \eta_1 \geq b)$ where $a > 0$, b, $c > 0$ and $\xi > 0$ are suitable constants, and η_1 is an appropriately chosen tail level. Hence, it will be assumed that (Naess, 2010),

$$\varepsilon_k(\eta) = q_k(\eta)\ \left[1+\xi_k\big(a_k(\eta-b_k)^{c_k}\big)\right]^{-\frac{1}{\xi_k}} ,\ \eta \geq \eta_1, \tag{38}$$

where the function $q_k(\eta)$ is weakly varying compared with the function $\left[1+\xi_k\big(a_k(\eta-b_k)^{c_k}\big)\right]^{-\frac{1}{\xi_k}}$ and $a_k > 0$, b_k, $c_k > 0$ and $\xi_k > 0$ are suitable constants, that in general will be dependent on k. Note that the values $c_k = 1$ and $q_k(\eta) = 1$ corresponds to the asymptotic limit, which is then a special case of the general expression given in Eq. (30).

An alternative form to Eq. (38) would be to assume that

$$\varepsilon_k(\eta) = \left[1+\xi_k\big(a_k(\eta-b_k)^{c_k}+d_k(\eta)\big)\right]^{-\frac{1}{\xi_k}} ,\ \eta \geq \eta_1, \tag{39}$$

where the function $d_k(\eta)$ is weakly varying compared with the function $a_k(\eta - b_k)^{c_k}$. However, for estimation purposes, it turns out that the form given by Eq. (30) is preferrable as it leads to simpler estimation procedures. This aspect will be discussed later in the paper.

For practical identification of the ACER functions given by Eq. (38), it expedient to assume that the unknown function $q_k(\eta)$ varies sufficiently slowly to be replaced by a constant. In general, $q_k(\eta)$ is not constant, but its variation in the tail region is assumed to be sufficiently slow to allow for its replacement by a constant. Hence, as in the Gumbel case, it is in effect assumed that $q_k(\eta)$ can be replaced by a constant for $\eta \geq \eta_1$, for an appropriate choice of tail marker η_1. For simplicity of notation, in the following we shall suppress the index k on the ACER functions, which will then be written as,

$$\varepsilon(\eta) = q\ [1 + \tilde{a}\,(\eta - b)^c]^{-\gamma}\ ,\ \ \eta \geq \eta_1, \tag{40}$$

where $\gamma = 1/\xi$, $\tilde{a} = a\xi$.

For the analysis of data, first the tail marker η_1 is provisionally identified from visual inspection of the log plot $(\eta, \ln \hat{\varepsilon}_k(\eta))$. The value chosen for η_1 corresponds to the beginning of regular tail behaviour in a sense to be discussed below.

The optimization process to estimate the parameters is done relative to the log plot, as for the Gumbel case. The mean square error function to be minimized is in the general case written as

$$F(\tilde{a}, b, c, q, \gamma) = \sum_{j=1}^{N} w_j \left| \log \hat{\varepsilon}(\eta_j) - \log q + \gamma\,[1 + \tilde{a}(\eta_j - b)^c] \right|^2, \tag{41}$$

where w_j is a weight factor as previously defined.

An option for estimating the five parameters $\tilde{a}, b, c, q, \gamma$ is again to use the Levenberg-Marquardt least squares optimization method, which can be simplified also in this case by observing that if $\tilde{a}$, b and c are fixed in Eq. (34), the optimization problem reduces to a standard weighted linear regression problem. That is, with $\tilde{a}$, b and c fixed, the optimal values of γ and $\log q$ are found using closed form weighted linear regression formulas in terms of w_j, $y_j = \log \hat{\varepsilon}(\eta_j)$ and $x_j = 1 + \tilde{a}(\eta_j - b)^c$.

It is obtained that the optimal values of γ and $\log q$ are given by relations similar to Eqs. (35) and (36). To calculate the final optimal set of parameters, the Levenberg-Marquardt method may then be used on the function $\tilde{F}(\tilde{a}, b, c) = F(\tilde{a}, b, c, q^*(\tilde{a}, b, c), \gamma^*(\tilde{a}, b, c))$ to find the optimal values $\tilde{a}^*$, b^* and c^*, and then the corresponding γ^* and q^* can be calculated. The optimal values of the parameters may e.g also be found by a sequential quadratic programming (SQP) method (Numerical Algorithms Group, 2010).

6 The Gumbel Method

To offer a comparison of the predictions obtained by the method proposed in this paper with those obtained by other methods, we shall use the predictions given by the two methods that seem to be most favored by practitioners, the Gumbel method and the peaks-over-threshold method.

The Gumbel method is based on recording epochal extreme values and fitting these values to a corresponding Gumbel distribution (Gumbel, 1958). By assuming that the recorded extreme value data are Gumbel distributed, then representing the obtained data set of extreme values as a Gumbel probability plot should ideally result in a straight line. In practice, one cannot expect this to happen, but on the premise that the data follow a Gumbel distribution, a straight line can be fitted to the data. Due to its simplicity, a popular method for fitting this straight line is the method of moments, which is also reasonably stable for limited sets of data. That is, writing the Gumbel distribution of the extreme value M_N as

$$\text{Prob}(M_N \leq \eta) = \exp\left\{-\exp\left(-a(\eta - b)\right)\right\}, \tag{42}$$

it is known that the parameters a and b are related to the mean value m_M and standard deviation σ_M of $M(T)$ as follows: $b = m_M - 0.57722\, a^{-1}$ and $a = 1.28255/\sigma_M$ (Bury, 1975). The estimates of m_M and σ_M obtained from the available sample therefore provides estimates of a and b, which leads to the fitted Gumbel distribution by the moment method.

Typically, a specified quantile value of the fitted Gumbel distribution is then extracted and used in a design consideration. To be specific, let us assume that the requested quantile value is the $100(1-\alpha)\%$ fractile, where α is usually a small number, for example $\alpha = 0.1$. To quantify the uncertainty associated with the obtained $100(1-\alpha)\%$ fractile value based on a sample of size $\tilde{N}$, the 95% confidence interval of this value is often used. A good estimate of this confidence interval can be obtained by using a parametric bootstrapping method (Efron and Tibshirani, 1993; Davison and Hinkley, 1997). Note that the assumption that the initial $\tilde{N}$ extreme values are actually generated with good approximation from a Gumbel distribution cannot easily be verified with any accuracy in general, which is a drawback of this method. Compared with the POT method, the Gumbel method would also seem to use much less of the information available in the data. This may explain why the POT method has become increasingly popular over the past years, but the Gumbel method is still widely used in practice.

7 The Peaks-Over-Threshold Method

The Generalized Pareto Distribution

The POT method is based on what is called the generalized Pareto (GP) distribution (defined below) in the following manner: It has been shown (Pickands, 1975) that asymptotically, the excess values above a high level will follow a GP distribution if and only if the parent distribution belongs to the domain of attraction of one of the extreme value distributions. The assumption of a Poisson process model for the exceedance times combined with GP distributed excesses can be shown to lead to the generalized extreme value (GEV) distribution for corresponding extremes, see below. The expression for the GP distribution is

$$G(y) = G(y; a, c) = \text{Prob}(Y \leq y) = 1 - \left(1 + c\,\frac{y}{a}\right)_+^{-1/c} . \tag{43}$$

Here $a > 0$ is a scale parameter and c ($-\infty < c < \infty$) determines the shape of the distribution. $(z)_+ = \max(0, z)$.

The asymptotic result referred to above implies that Eq. (43) can be used to represent the conditional cumulative distribution function of the excess $Y = X - u$ of the observed variate X over the threshold u, given that $X > u$ for u sufficiently large (Pickands, 1975). The cases $c > 0$, $c = 0$ and $c < 0$ correspond to Fréchet (Type II), Gumbel (Type I), and reverse Weibull (Type III) domains of attraction, respectively, cf. section below.

For $c = 0$, which corresponds to the Gumbel extreme value distribution, the expression between the parentheses in Eq. (43) is understood in a limiting sense as the exponential distribution,

$$G(y) = G(y; a, 0) = \exp(-y/a)\,. \tag{44}$$

Return Periods

The return period R of a given wind speed, in years, is defined as the inverse of the probability that the specified wind speed will be exceeded in any one year. If λ denotes the mean exceedance rate of the threshold u per year (i.e., the average number of data points above the threshold u per year), then the return period R of the value of X corresponding to the level $x_R = u + y$ is given by the relation

$$R = \frac{1}{\lambda\,\text{Prob}(X > x_R)} = \frac{1}{\lambda\,\text{Prob}(Y > y)} \tag{45}$$

Hence, it follows that

$$\text{Prob}(Y \leq y) = 1 - 1/(\lambda R)\,. \tag{46}$$

Invoking equation (1) for $c \neq 0$ leads to the result

$$x_R = u - a[1 - (\lambda R)^c]/c\,. \tag{47}$$

Similarly, for $c = 0$, it is found that,

$$x_R = u + a \ln(\lambda R)\,, \tag{48}$$

where u is the threshold used in the estimation of c and a.

8 Extreme Value Prediction for Synthetic Data

In this section we illustrate the performance of the ACER method and also the 95% CI estimation. We consider 20 years of synthetic wind speed data, amounting to 2000 data points, which is not much for detailed statistics. However, this case may represent a real situation when nothing but a limited data sample is available. In this case it is crucial to provide extreme value estimates utilizing all data available. As we shall see, the tail extrapolation technique proposed in this paper performs better than asymptotic methods such as POT or Gumbel.

The extreme value statistics will first be analyzed by application to synthetic data for which the exact extreme values can be calculated (Naess and Clausen, 2001). In particular, it is assumed that the underlying (normalized) stochastic process $Z(t)$ is stationary and Gaussian with mean value zero and standard deviation equal to one. It is also assumed that the mean zero up-crossing rate $\nu^+(0)$ is such that the product $\nu^+(0)T = 10^3$ where $T = 1$ year, which seems to be typical for the wind speed process. Using the Poisson assumption, the distribution of the yearly extreme value of $Z(t)$ is then calculated by the formula

$$F^{1\mathrm{yr}}(\eta) = \exp\left\{-\nu^+(\eta)T\right\} = \exp\left\{-10^3 \exp\left(-\frac{\eta^2}{2}\right)\right\}\,, \tag{49}$$

where $T = 1$ year and $\nu^+(\eta)$ is the mean up-crossing rate per year, η is the scaled wind speed. The 100-year return period value $\eta^{100\mathrm{yr}}$ is then calculated from the relation $F^{1\mathrm{yr}}(\eta^{100\mathrm{yr}}) = 1 - 1/100$, which gives $\eta^{100\mathrm{yr}} = 4.80$.

The Monte Carlo simulated data to be used for the synthetic example are generated based on the observation that the peak events extracted from measurements of the wind speed process, are usually separated by 3-4 days. This is done to obtain approximately independent data, cf. (Naess, 1998b). In accordance with this, peak event data are generated from the extreme value distribution

$$F^{3\mathrm{d}}(\eta) = \exp\left\{-q \exp\left(-\frac{\eta^2}{2}\right)\right\}\,, \tag{50}$$

where $q = \nu^+(0)T = 10$, which corresponds to $T = 3.65$ days, and $F^{1\mathrm{yr}}(\eta) = \left(F^{3\mathrm{d}}(\eta)\right)^{100}$.

Since the data points (i.e. $T = 3.65$ days maxima) are independent, $\varepsilon_k(\eta)$ is independent of k. Therefore we put $k = 1$. Since we have 100 data from one year, the data amounts to 2000 data points. For estimation of a 95% confidence interval for each estimated value of the ACER function $\varepsilon_1(\eta)$ for the chosen

range of η-values, the required standard deviation in Eq. (27) was based on 20 estimates of the ACER function using the yearly data. This provided a 95% confidence band on the optimaly fitted curve based on 2000 data. From these data the predicted 100 year return level is obtained from $\hat{\varepsilon}_1(\eta^{100\text{yr}}) = 10^{-4}$.

The POT prediction of the 100 year return level was based on using maximum likelihood estimates (MLE) of parameters for a specific choice of threshold. The 95% confidence interval was obtained from the parametrically bootstrapped PDF of the POT prediction for the given threshold. A sample of 100,000 data sets was used. One of the unfortunate features of the POT method is that the predicted 100 year value may vary significantly with the choice of threshold. So also for the synthetic data. We have followed the standard recommended procedures for identifying a suitable threshold (Coles, 2001).

Similarly, the 100 year return level predicted by the Gumbel method was based on using the moment method for parameter estimation on the sample of 20 yearly extremes. The 95% confidence interval was obtained from the parametrically bootstrapped PDF of the Gumbel prediction. This was based on a sample of size 100,000 data sets of 20 yearly extremes .

In order to get an idea about the performance of the ACER, POT and Gumbel methods, 10 independent 20 yr MC simulations as discussed above were done. Table 1 compares predicted values and confidence intervals. It is seen that the average of the 10 predicted 100 year return levels is slightly better for the ACER method than for both the POT and the Gumbel methods. But more significantly, the range of predicted 100 year return levels by the ACER method is 4.43 - 5.12, while the same for the POT method is 4.05 - 5.89, and for the Gumbel method 4.63 - 5.69. Hence, in this case the ACER method performs consistently better than both these methods. It is also observed that the estimated 95% confidence intervals are generally quite consistent between the three methods.

An example of the ACER plot and results obtained for one set of data is presented in Fig. 1. The predicted 100 year value is 5.12 with a predicted 95% confidence interval (4.57, 5.75). Fig. 2 presents POT predictions based on MLE for different thresholds in terms of the number n of data points above the threshold. The predicted value is 4.23 at $n = 140$, while the 95% confidence interval is (3.95, 4.41). The same data set as in Fig. 1 was used. This was also used for the Gumbel plot shown in Fig. 3. In this case the predicted value is 4.34 with a 95% confidence interval of (4.12, 4.59).

9 Measured wind speed data

In this section we analyze real wind speed data, measured at the weather station at Sula off the coast of Norway. Extreme wind speed prediction is an important issue for design of structures exposed to the weather variations. Significant

Table 1: Return level estimates and 95% CI comparison for A=ACER, P=POT and G=Gumbel

No.	A $\hat{\eta}^{100}$	A CI	P $\hat{\eta}^{100}$	P CI	G $\hat{\eta}^{100}$	G CI
1	4.48	(4.15, 4.76)	4.05	(3.85, 4.18)	4.63	(4.21, 5.09)
2	4.71	(4.25, 5.18)	4.73	(4.16, 5.20)	4.88	(4.36, 5.44)
3	4.82	(4.38, 5.29)	5.89	(4.58, 7.54)	5.69	(4.84, 6.61)
4	5.05	(4.47, 5.65)	4.55	(4.11, 4.85)	4.85	(4.36, 5.38)
5	4.64	(4.27, 5.11)	4.53	(4.09, 4.83)	4.70	(4.26, 5.18)
6	5.00	(4.35, 5.60)	5.24	(4.43, 6.05)	5.39	(4.69, 6.16)
7	4.82	(4.40, 5.40)	4.51	(4.10, 4.78)	4.74	(4.31, 5.21)
8	5.12	(4.57, 5.75)	4.23	(3.95, 4.41)	4.34	(4.12, 4.59)
9	4.43	(4.14, 4.82)	4.60	(4.08, 5.05)	4.74	(4.26, 5.25)
10	4.86	(4.38, 5.35)	4.60	(4.13, 4.93)	4.68	(4.31, 5.09)
Average	4.79		4.70		4.86	

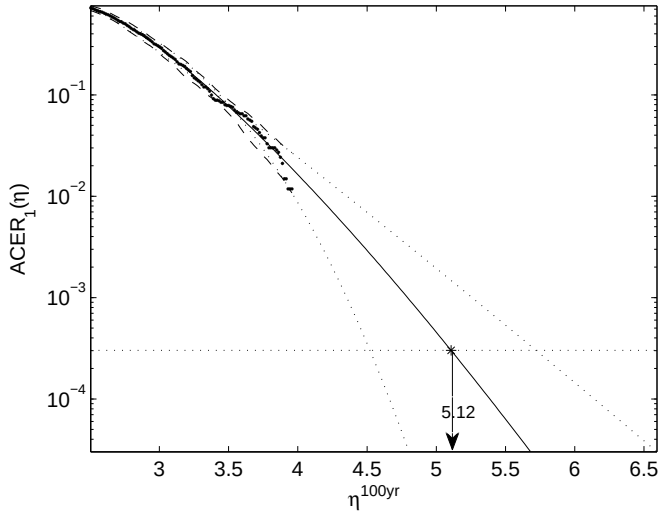


Figure 1: Synthetic data ACER $\hat{\varepsilon}_1$, Monte Carlo simulation (*); optimized curve fit (—); empirical 95% confidence band (- -); optimized confidence band $(\cdots)$. Tail marker $\eta_1 = 2.5$

efforts have been devoted to the problem of predicting extreme wind speeds on the basis of measured data by various authors over several decades, see e.g. (Cook, 1982; Naess, 1998a; Palutikof et al., 1999; Perrin et al., 2006) for

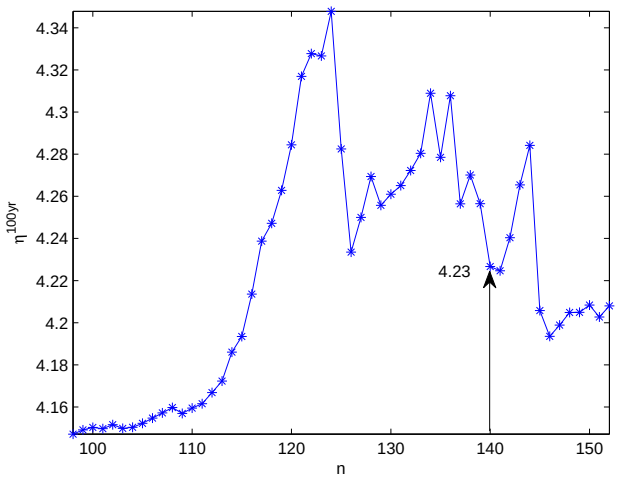


Figure 2: The point estimate $\tilde{\eta}^{100\text{yr}}$ of the 100-year return period value based on 20 years synthetic data as a function of the number n of data points above threshold. The return level estimate = 4.23 at $n = 140$.

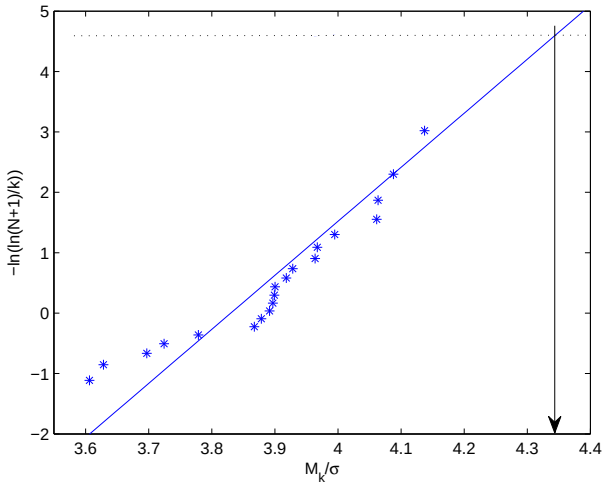


Figure 3: The point estimate $\tilde{\eta}^{100\text{yr}}$ of the 100-year return period value based on 20 years synthetic data. The return level estimate = 4.34.

extensive references to previous work.

Hourly maximum gust wind was recorded during 12 years 1998-2010. The objective is to estimate a 100 year return wind speed. Variation in the wind speed caused by seasonal variations in the wind climate during the year makes the wind speed a non-stationary process on the scale of months. Moreover, due to global climate change, yearly statistics may vary on the scale of years. The

latter is, however, a slow process and for the purpose of long-term prediction we assume here that within a time span of 100 years a quasi-stationary model of the wind speeds applies. However, this may not be entirely true.

Fig. 4 highlights the cascade of ACER estimates $\hat{\varepsilon}_1 \ldots \hat{\varepsilon}_6$ for the case of 12 years of hourly data. Here $\hat{\varepsilon}_6$ is considered to represent the final converged results. By 'converged' we mean that $\hat{\varepsilon}_6 \approx \hat{\varepsilon}_k$ for $k > 6$ in the tail, so that there is no need to consider conditioning of an even higher order than 6. Fig. 4 reveals a rather strong statistical dependence between consecutive data, which is clearly reflected in the effect of conditioning on previous data values. It is also interesting to observe that this effect is to some extent captured already by $\hat{\varepsilon}_2$, that is, by conditioning only on the value of the previous data point. Subsequent conditioning on more than one previous data point does not lead to substantial changes in ACER values, especially for tail values. On the other hand, to bring out fully the dependence structure of these data, it was necessary to carry the conditioning process to (at least) the 6th ACER function, as discussed above.

However, from a practical point of view, the most important information provided by the ACER plot of Fig. 4 is that for the prediction of a 100 year value, one may use the first ACER function. The reason for this is that Fig. 4 shows that all the ACER functions coalesce in the far tail. Hence, we may use any of the ACER functions for the prediction. Then, the obvious choice is to use the first ACER function, which allows us to use all the data in its estimation and thereby increase accuracy.

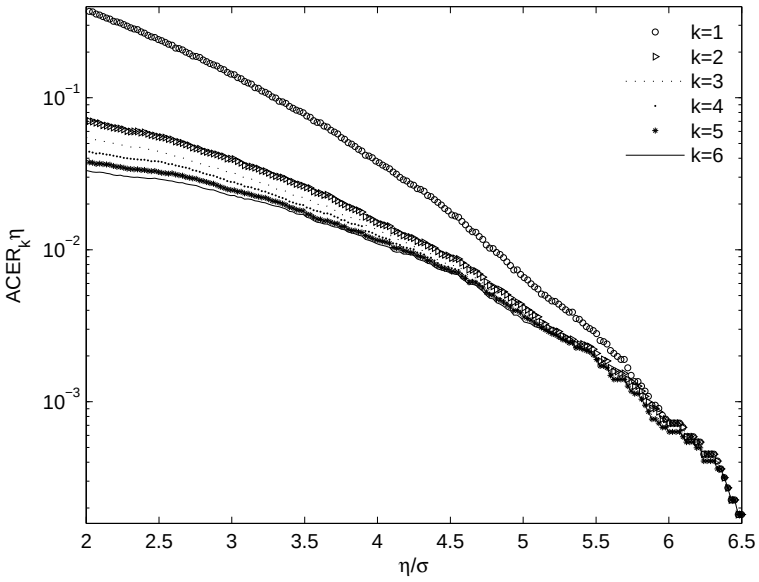


Figure 4: Sula wind speed statistics, 12 years hourly data. Comparison between ACER estimates for different degrees of conditioning. $\sigma = 5.9$ m/s.

In Fig. 5 is shown the results of parametric estimation of the return value and its 95% CI for 12 years of hourly maxima. The predicted 100 year return speed is $\eta^{100\text{yr}} = 46.14$ m/s with 95% confidence interval $(44.44, 48.01)$. $R = 12$ years of data may not be enough to guarantee Eq. (27), since we required $R \geq 20$. Nevertheless, for simplicity, we use it here even with $R = 12$, accepting that it may not be very accurate.

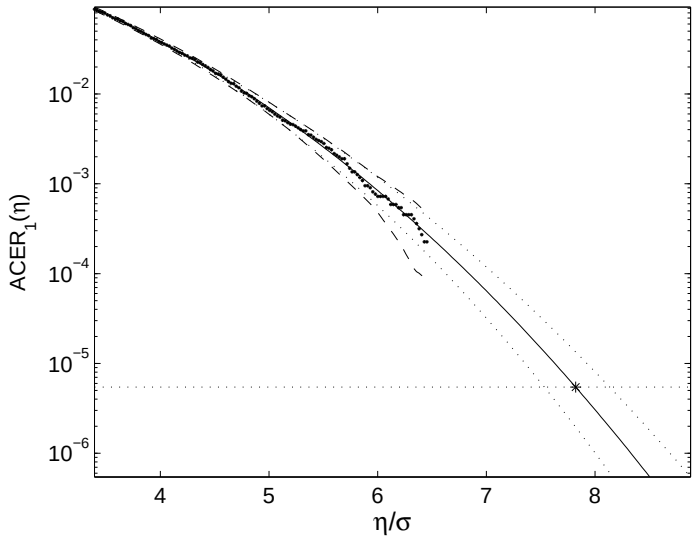


Figure 5: Sula wind speed statistics, 12 years hourly data. $\hat{\varepsilon}_1(\eta)$ (*); Optimized curve fit (—); Empirical 95% confidence band (- -); Optimized confidence band $(\cdots)$. Tail marker $\eta_1 = 20$ m/s. $\sigma = 5.9$ m/s

Fig. 6 presents POT predictions for different threshold numbers based on MLE. The POT prediction is $\eta^{100\text{yr}} = 43.07$ m/s at threshold $n = 80$, while the bootstrapped 95% confidence interval is found to be (40.79, 45.72) m/s based on 100,000 generated samples. It is interesting to observe the unstable characteristics of the predictions over a range of threshold values, while they are quite stable on either side of this range giving predictions that are more in line with the results from the other two methods.

Fig. 7 presents a Gumbel plot based on the 12 yearly extremes extracted from the 12 years of hourly data. The Gumbel prediction $\eta^{100\text{yr}} = 48.14$ m/s, with a parametric bootstrapped 95% confidence interval equal to (40.98, 55.96) m/s.

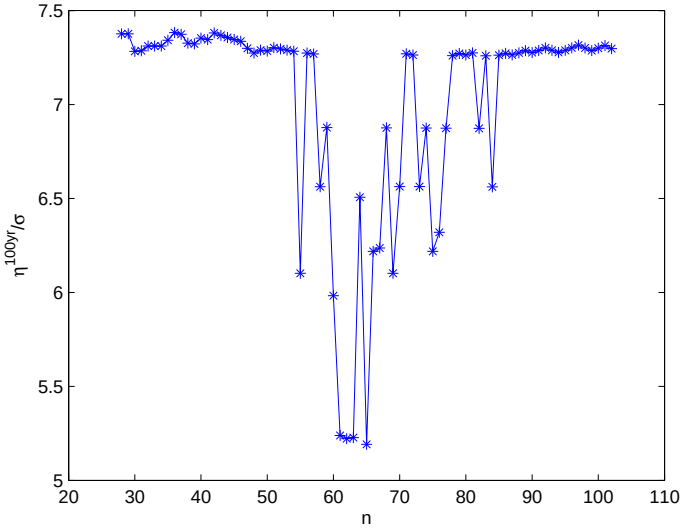


Figure 6: The point estimate $\eta^{100\text{yr}}$ of the 100-year return level based on 12 years hourly data as a function of the number n of data points above threshold. $\sigma = 5.9$ m/s.

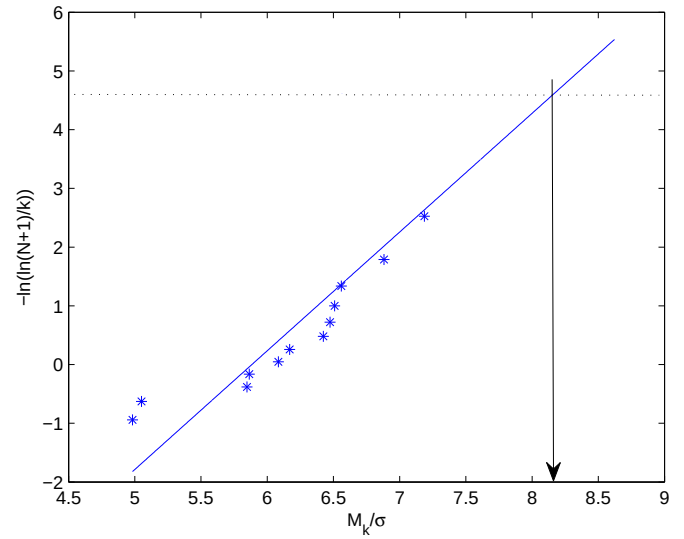


Figure 7: Sula wind speed statistics, 12 years of hourly data. Gumbel plot of yearly extremes. $\sigma = 5.9$ m/s.

10 Extreme value prediction for a narrow band process

In engineering mechanics a classical extreme response prediction problem is the case of a lightly damped mechanical oscillator subjected to random forces.

To illustrate this prediction problem we shall investigate the response process of a linear mechanical oscillator driven by a Gaussian white noise. Let $X(t)$ denote the displacement response; the dynamic model can then be expressed as, $\ddot{X}(t) + 2\zeta\omega_e\dot{X}(t) + \omega_e^2 X(t) = W(t)$, where ζ = relative damping, ω_e = undamped eigenfrequency, and $W(t)$ = a stationary Gaussian white noise (of suitable intensity). By choosing a small value for ζ the response time series will exhibit narrow band characteristics, that is, the spectral density of the response process $X(t)$ will assume significant values only over a narrow range of frequencies. This manifests itself by producing a strong beating of the response time series, which means that the size of the response peaks will change slowly in time, see Fig. 8. A consequence of this is that neighbouring peaks are strongly correlated. Hence the problem with accurate prediction, since the usual assumption of independent peak values is then violated.

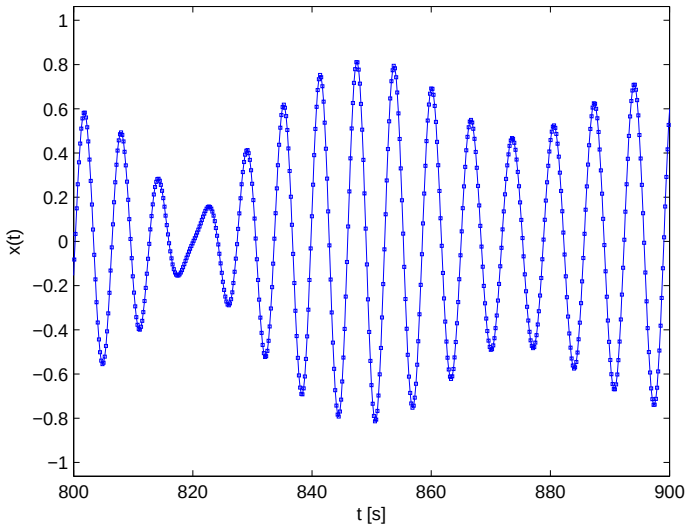


Figure 8: Part of the narrow-band response time series of the linear oscillator.

Many approximations have been proposed to deal with this correlation problem, but no completely satisfactory solution has been presented. In this section we will show that the ACER method solves this problem efficiently and elegantly in a statistical sense. In Fig. 9 are shown some of the ACER functions for the example time series. It may be verified from Fig. 8 that there are approximately 32 sample points between two neighbouring peaks in the time series. To illustrate a point, we have chosen to analyze the time series consisting of all sample points. Usually, in practice, only the time series obtained by extracting the peak values would be used for the ACER analysis. In the present case, the first ACER function is then based on assuming that all the sampled data points are independent, which is obviously completely wrong. The second

ACER function, which is based on counting each exceedance with an immediately preceding non-exceedance, is nothing but an upcrossing rate. Using this ACER function is largely equivalent to assuming independent peak values. It is now interesting to observe that the 25th ACER function can hardly be distinguished from the second ACER function. In fact, the ACER functions after the second do not change appreciably until one starts to approach the 32nd, which corresponds to hitting the previous peak value in the conditioning process. So the important information concerning the dependence structure in the present time series seems to reside in the peak values, which may not be very surprising. It is seen that the ACER functions show a significant change in value as a result of accounting for the correlation effects in the time series. To verify the full dependence structure in the time series it is necessary to continue the conditioning process down to at least the 64th ACER function. In the present case there is virtually no difference between the 32nd and the 64th, which shows that the dependence structure in this particular time series is captured almost completely by conditioning on the previous peak value. It is interesting to contrast the method of dealing with the effect of sampling frequency discussed here with that of (Robinson and Tawn, 2000).

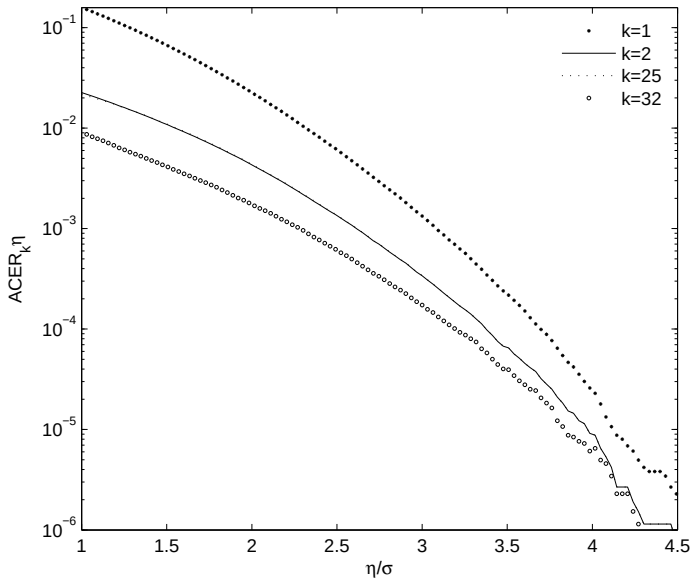


Figure 9: Comparison between ACER estimates for different degrees of conditioning for the narrow-band time series.

11 Extreme tether tension

In this section we shall study data obtained from model tests of an offshore platform for oil production. The Heidrun tension leg platform (TLP) is a large concrete platform installed in 347 m depth in the Norwegian Sea. It is designed with four circular columns forming a square, with a square ring pontoon. Extensive model tests at scale 1:55 were carried out in MARINTEK's 50 m $\times$ 80 m Ocean Basin in Trondheim in 1993. The mass of the TLP in ultimate limit state (ULS) conditions was 257888 tonnes, and the draft was 79.3 m. All data given here are in prototype scale. A sketch depicting the TLP from the side is shown in Fig. 10. The column diameter was 31 m, except in a small section in the wave zone where it was 31.6 m. The centre-to-centre distance between columns was 80.0 m. The pontoon has a rectangular cross-section, with a height 13.0 m and width 16.0 m.

The actual prototype tether group of four tethers at each column was modelled by a single equivalent tether, designed to correspond to the prototype with respect to stiffness, drag and weight properties.

The original test program included a number of different irregular wave test conditions, and a large number of measuring channels, cf. (Naess et al., 2009). In this study, we concentrate on one severe ULS condition. It is specified in terms of the following sea state, which is a unidirectional (long crested) sea: significant wave height $H_s = 15.7$ m, and spectral peak period $T_p = 17.8$ s.

The platform had a 45 degrees heading relative to the waves. The most heavily loaded tether is designated T10, and it is positioned towards the waves. Six different random realizations of duration 3 hours each were run. Thus the resulting statistics correspond to 18 hours duration for the given sea state.

A particular observation from the model tests was the strongly non-Gaussian behaviour of the measured tensions, especially in these high sea states. Thus, resonant high-frequency oscillations occurred, known as "ringing", which are excited by higher-order wave forces on columns in high and steep individual waves (Faltinsen et al., 1995; Stansberg, 1997). This comes in addition to the more commonly known "springing", excited by second-order sum-frequency forces. The extraordinary statistical behaviour was a main reason why these sea states were run with 6 realizations each. A time series sample from the measurements is shown in Fig. 11, which clearly displays the ringing phenomenon caused by a steep wave.

A basic statistical analysis of the 18 hours of time series for tension T10 shows that the dynamical mean of T10 is 97537.5 kN, while the standard deviation equals 6234.22 kN. In Fig. 12 are plotted the ACER functions $\varepsilon_k(\eta)$ for $k = 1, \ldots, 5$. It is seen that there is a significant effect of dependence in the time series, which is reflected in the fact that the $\varepsilon_k(\eta)$, for $k = 2, \ldots, 5$, are noticeably smaller than $\varepsilon_1(\eta)$ over the whole range of response values. There is no tendency for $\varepsilon_1(\eta)$ to merge with $\varepsilon_k(\eta)$ for $k = 2, \ldots, 5$. However, it is seen

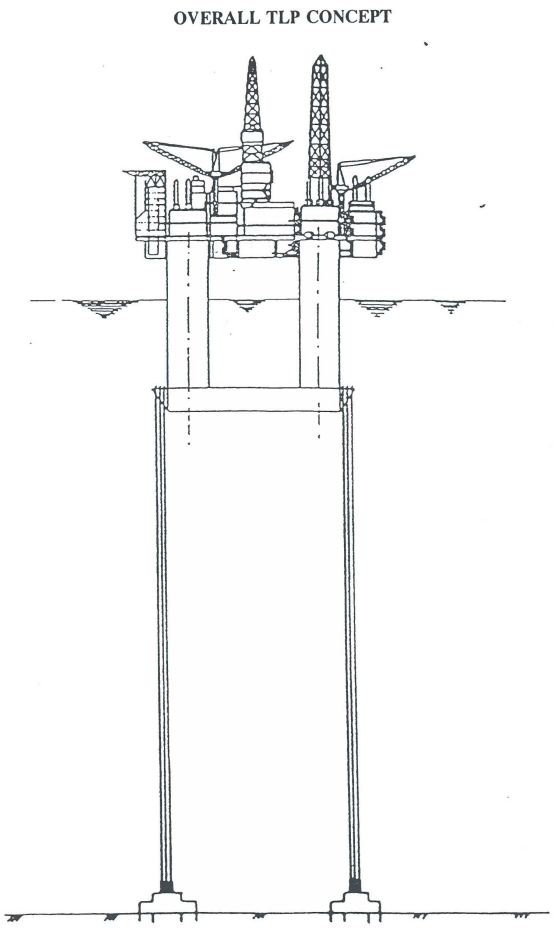


Figure 10: Heidrun TLP as seen from the side.

that already for $k = 2$ a good approximation is obtained, and further, that convergence is certainly achieved for $k = 4$. To emphasize this point, the predicted value of the 90% fractile of the 3 hour extreme value distribution by the ACER method is found to be $\eta_{0.90} = 115965$ kN, with 95% confidence interval (86379, 168745) kN based on $\varepsilon_1(\eta)$, while $\eta_{0.90} = 93723$ kN, with 95% confidence interval (63923, 139448) kN based on $\varepsilon_4(\eta)$. The tail marker is $\eta_1 = 20000$ kN in both cases. It is noticeable that the predicted 90% fractile value by the ACER method based on $\varepsilon_4(\eta)$ is significantly lower (19%) than the corresponding value based on $\varepsilon_1(\eta)$. Hence, the effect of statistical dependence in the response time series on the predicted extremes is of some importance. It is also noted that the predicted statistical uncertainty is approximately the same in both cases.

To highlight the predictions based on the ACER, POT and Gumbel methods, we have estimated the 10 year return period values provided by the three methods. The obtained results can be summarized briefly as follows (all numbers in kN, 95% confidence interval in parenthesis). ACER: 153515, (86122, 204423); POT (at the number n of data points above the threshold = 140):

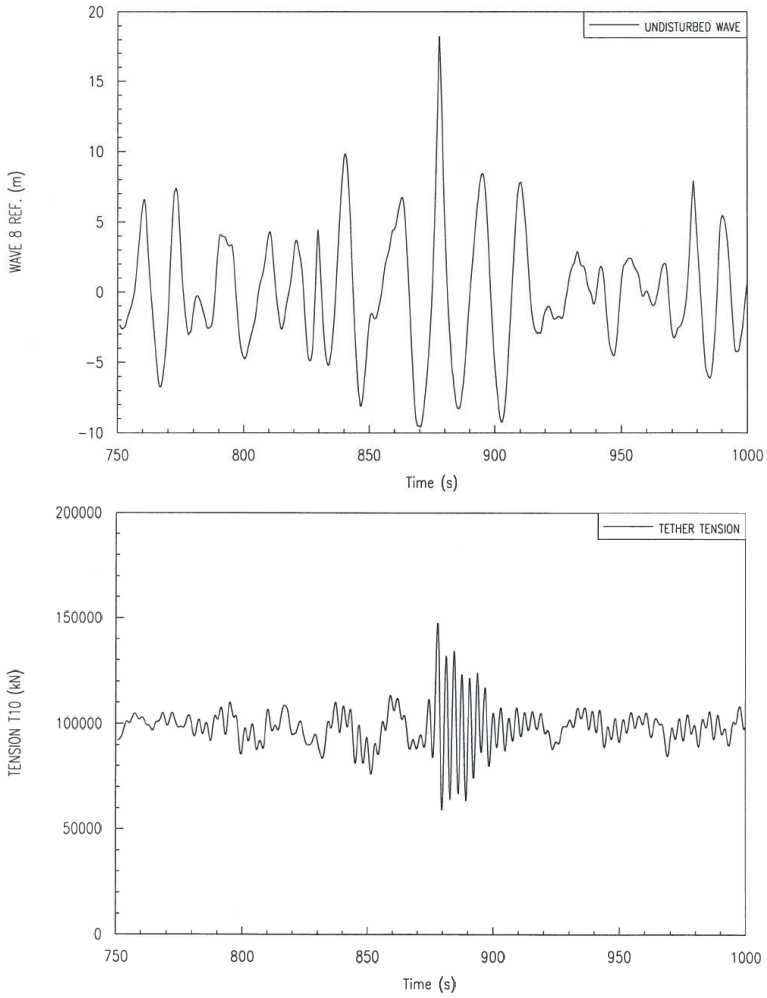


Figure 11: Short time series samples of wave elevation and tether tension T10, with a ringing event caused by a steep wave.

310864, (79414.5, 1.05446e+006) ; Gumbel: 161479, (108723, 219878). Corresponding figures are given in Figs. 13 - 15. Notice the large variability of the POT estimates depending on the choice of threshold.

12 Concluding remarks

This paper studies a new method for extreme value prediction for sampled time series. The method is based on the introduction of a conditional average exceedance rate, which allows dependence in the time series to be properly accounted for. Declustering of the data is therefore avoided, and all the data are used in the analysis. Significantly, the proposed method also aims at capturing to some extent the sub-asymptotic form of the extreme value distribution.

Results for wind speeds, both synthetic and measured, are used to illustrate

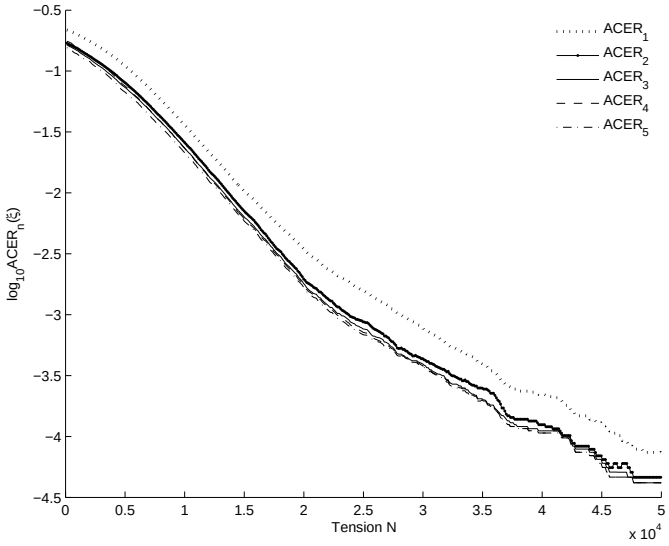


Figure 12: Log plot of empirical ACER $\varepsilon_k(\eta)$, $k = 1, \ldots, 5$.

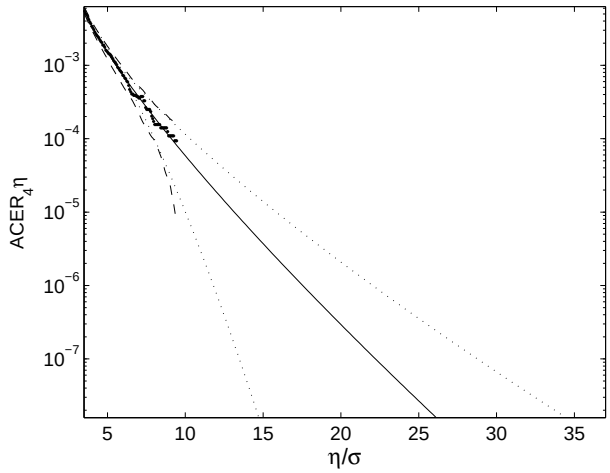


Figure 13: Log plot of empirical ACER $\hat{\varepsilon}_4(\eta)$(—-) with extrapolation by optimally fitted curve (– –). – · –: optimized 95% confidence band; · · ·: reanchored empirical 95% confidence band.

the method. Two prediction problems related to applications in mechanics are also presented. The validation of the method is done by comparison with exact results (when available), or other widely used methods for extreme value statistics, such as the Gumbel and peaks-over-threshold (POT) methods. Comparison of the various predictions indicate that the proposed method may provide

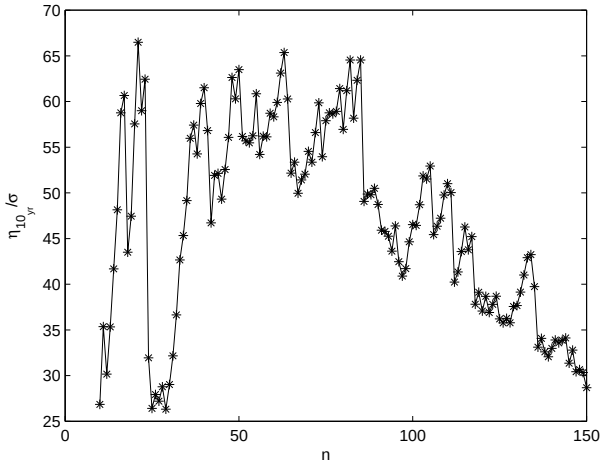


Figure 14: Point estimates of the 10 year value by the POT method as function of number n of data above threshold.

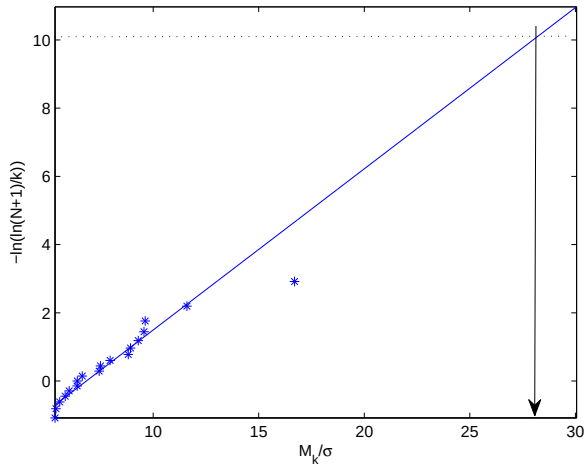


Figure 15: Point estimate of the 10 year value by the Gumbel method.

more accurate results than the Gumbel and POT methods.

Subject to certain restrictions, the proposed method also applies to non-stationary time series, but it cannot directly predict e.g. the effect of climate change in the form of long-term trends in the average exceedance rates extending beyond the data. This must be incorporated into the analysis by explicit modelling techniques.

As a final remark, it may be noted that the ACER method as described in this paper has a natural extension to higher dimensional distributions. The implication is that it is then possible to provide estimates of e.g. the exact

bivariate extreme value distribution for a suitable set of data (Naess, 2011). However, as is easily recognized, the extrapolation problem is not as simply dealt with as for the univariate case studied in this paper.

13 Acknowledgements

This work was supported by the Research Council of Norway (NFR) through the Centre for Ships and Ocean Structures (CeSOS) at the Norwegian University of Science and Technology.

References

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes*. Chichester, UK: John Wiley & Sons, Ltd.

Bury, K. V. (1975). *Statistical Models in Applied Sciences*. New York: John Wiley & Sons, Inc.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.

Coles, S. G. (1994). A temporal study of extreme rainfall. In V. Barnett and K. F. Turkman (Eds.), *Statistics for the Environment 2 - Water Related Issues*, Chapter 4, pp. 61–78. Chichester: John Wiley & Sons.

Cook, N. J. (1982). Towards better estimation of extreme winds. *Journal of Wind Engineering and Industrial Aerodynamics 9*(3), 295–323.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and their Applications*. London: Cambridge University Press.

Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, B 52*(3), 393–442.

Draper, N. R. and H. Smith (1998). *Applied Regression Analysis*. New York, NY: Wiley-Interscience.

Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events*. New York: Springer.

Falk, M., J. Hüsler, and R.-D. Reiss (2004). *Laws of Small Numbers: Extremes and Rare Events* (2. ed.). Basel: Birkhäuser.

Faltinsen, O. M., J. N. Newman, and T. Vinje (1995). Nonlinear wave loads on a slender vertical cylinder. *Journal of Fluid Mechanics 289*, 179–198.

Gill, P., W. Murray, and M. H. Wright (1981). *Practical Optimization*. London: Academic Press.

Gumbel, E. J. (1958). *Statistics of Extremes*. New York, NY: Columbia University Press.

Montgomery, D. C., E. A. Peck, and G. G. Vining (2002). *Introduction to Linear Regression Analysis*. Amsterdam, The Netherlands: Elsevier Science Publishers B. V.

Naess, A. (1984). On the long-term statistics of extremes. *Applied Ocean Research 6*(4), 227–228.

Naess, A. (1985). The joint crossing frequency of stochastic processes and its application to wave theory. *Applied Ocean Research 7*(1), 35–50.

Naess, A. (1990). Approximate first-passage and extremes of narrow-band Gaussian and non-Gaussian random vibrations. *Journal of Sound and Vibration 138*(3), 365–380.

Naess, A. (1998a). Estimation of long return period design values for wind speeds. *Journal of Engineering Mechanics, ASCE 124*(3), 252–259.

Naess, A. (1998b). Statistical extrapolation of extreme value data based on the peaks over threshold method. *Journal of Offshore Mechanics and Arctic Engineering, ASME 120*, 91–96.

Naess, A. (2010). Estimation of extreme values of time series with heavy tails. Preprint Statistics No. 14/2010, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim.

Naess, A. (2011). A note on the bivariate ACER method. Preprint Statistics No. 01/2011, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim.

Naess, A. and P. H. Clausen (2001). Combination of peaks-over-threshold and bootstrapping methods for extreme value prediction. *Structural Safety 23*, 315–330.

Naess, A. and O. Gaidai (2008). Monte Carlo methods for estimating the extreme response of dynamical systems. *Journal of Engineering Mechanics, ASCE 134*(8), 628–636.

Naess, A. and O. Gaidai (2009). Estimation of extreme values from sampled time series. *Structural Safety 31*, 325–334.

Naess, A., O. Gaidai, and S. Haver (2007). Efficient estimation of extreme response of drag dominated offshore structures by Monte Carlo simulation. *Ocean Engineering 34*(16), 2188–2197.

Naess, A., C. T. Stansberg, and O. Batsevych (2009). Prediction of extreme tether tension for a TLP. In *Proceedings 28th International Conference on Offshore Mechanics and Arctic Engineering*, pp. OMAE–2009–80169. New York: ASME.

Numerical Algorithms Group (2010). *NAG Toolbox for Matlab*. Oxford, UK: NAG Ltd.

Palutikof, J. P., B. B. Brabson, D. H. Lister, and S. T. Adcock (1999). A review of methods to calculate extreme wind speeds. *Meteorological Applications 6*, 119–132.

Perrin, O., H. Rootzen, and R. Taesler (2006). A discussion of statistical methods used to estimate extreme wind speeds. *Theoretical and Applied Climatology 85*(3-4), 203–215.

Pickands, J. (1975). Statistical inference using order statistics. *Annals of Statistics 3*, 119–131.

Reiss, R.-D. and M. Thomas (2007). *Statistical Analysis of Extreme Values* (3. ed.). Basel: Birkhäuser.

Robinson, M. E. and J. A. Tawn (2000). Extremal analysis of processes sampled at different frequencies. *Journal of the Royal Statistical Society, Series B 62*(1), 117–136.

Schall, G., M. H. Faber, and R. Rackwitz (1991). The ergodicity assumption for sea states in the reliability estimation of offshore structures. *J Off Mech Arctic Engg, ASME 113*(3), 241–246.

Smith, R. L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability 29*, 37–45.

Stansberg, C. T. (July 1997). Comparing ringing loads from experiments with cylinders of different diameters - an empirical study. In J. H. Vugts (Ed.), *Proceedings 8th International Conference on Behaviour of Offshore Structures (BOSS'97)*. Elsevier.

Vanmarcke, E. H. (1975). On the distribution of the first-passage time for normal stationary random processes. *Journal of Applied Mechanics 42*, 215–220.

Yun, S. (1998). The extremal index of a higher-order stationary Markov chain. *Annals of Applied Probability 8*, 408–437.