Research Article

Estimation of Extreme Values by the Average Conditional Exceedance Rate Method

A. Naess,[1] O. Gaidai,[2] and O. Karpa[3]

[1] *Department of Mathematical Sciences and CeSOS, Norwegian University of Science and Technology, 7491 Trondheim, Norway*
[2] *Norwegian Marine Technology Research Institute, 7491 Trondheim, Norway*
[3] *Centre for Ships and Ocean Structures (CeSOS), Norwegian University of Science and Technology, 7491 Trondheim, Norway*

Correspondence should be addressed to A. Naess; arvidn@math.ntnu.no

This paper details a 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 statistical estimates of the exact extreme value distribution provided by the data in most cases of practical interest. It avoids the problem of having to decluster the data to ensure independence, which is a requisite component in the application of, for example, the standard peaks-over-threshold method. The proposed method also targets the use of subasymptotic 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.

1. Introduction

Extreme value statistics, even in applications, are 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 [1, 2]. Or it is assumed that the exceedances above high thresholds follow a generalized (asymptotic) Pareto distribution with parameters that are estimated from the data [1–4]. 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 since the 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 do 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 [2, 5, 6]. 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. The approach is based on two separate components which are designed to improve on two important aspects of extreme value prediction based on observed data. The first component has the capability to accurately capture and display the effect of statistical dependence in the data, which opens for the opportunity of using all the available data in the analysis. The second component is then constructed so as to make it possible to incorporate to a certain extent also the subasymptotic part of the data into the estimation of extreme values, which is of some importance for accurate estimation. We have used the proposed method on a wide variety of estimation problems, and our experience is that

it represents a very powerful addition to the toolbox of methods for extreme value estimation. Needless to say, what is presented in this paper is by no means considered a closed chapter. It is a novel method, and it is to be expected that several aspects of the proposed approach will see significant improvements.

2. Cascade of Conditioning Approximations

In this section, a sequence of nonparametric distribution functions will be constructed that converges to the exact extreme value distribution for the time series considered. This constitutes the core of the proposed approach.

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 $P(\eta) = \text{Prob}(M_N \le \eta)$ 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 [7, 8] 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

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

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

$$
P(\eta) \approx P_1(\eta) := \prod_{j=1}^{N} \text{Prob}(X_j \le \eta),
\tag{2}
$$

where := means "by definition", the following one-step memory approximation will, to a certain extent, account for the dependence between the X_j's,

$$
\begin{aligned}
&\text{Prob}\{X_j \le \eta \mid X_{j-1} \le \eta, \ldots, X_1 \le \eta\} \\
&\quad \approx \text{Prob}\{X_j \le \eta \mid X_{j-1} \le \eta\},
\end{aligned}
\tag{3}
$$

for $2 \le j \le N$. With this approximation, it is obtained that

$$
\begin{aligned}
P(\eta) &\approx P_2(\eta) \\
&:= \prod_{j=2}^{N} \text{Prob}\{X_j \le \eta \mid X_{j-1} \le \eta\} \text{Prob}(X_1 \le \eta).
\end{aligned}
\tag{4}
$$

By conditioning on one more data point, the one-step memory approximation is extended to

$$
\begin{aligned}
&\text{Prob}\{X_j \le \eta \mid X_{j-1} \le \eta, \ldots, X_1 \le \eta\} \\
&\quad \approx \text{Prob}\{X_j \le \eta \mid X_{j-1} \le \eta, X_{j-2} \le \eta\},
\end{aligned}
\tag{5}
$$

where $3 \le j \le N$, which leads to the approximation

$$
\begin{aligned}
P(\eta) &\approx P_3(\eta) := \prod_{j=3}^{N} \text{Prob}\{X_j \le \eta \mid X_{j-1} \le \eta, X_{j-2} \le \eta\} \\
&\quad \cdot \text{Prob}\{X_2 \le \eta \mid X_1 \le \eta\} \text{Prob}(X_1 \le \eta).
\end{aligned}
\tag{6}
$$

For a general k, $2 \le k \le N$, it is obtained that

$$
\begin{aligned}
P(\eta) &\approx P_k(\eta) \\
&:= \prod_{j=k}^{N} \text{Prob}\{X_j \le \eta \mid X_{j-1} \le \eta, \ldots, X_{j-k+1} \le \eta\} \\
&\quad \cdot \prod_{j=2}^{k-1} \text{Prob}\{X_j \le \eta \mid X_{j-1} \le \eta \ \ldots, X_1 \le \eta\} \\
&\quad \cdot \text{Prob}(X_1 \le \eta),
\end{aligned}
\tag{7}
$$

where $P(\eta) = P_N(\eta)$.

It should be noted that the one-step memory approximation adopted above is not a Markov chain approximation [9–11], nor do the k-step memory approximations lead to kth-order Markov chains [12, 13]. An effort to relinquish the Markov chain assumption to obtain an approximate distribution of clusters of extremes is reported in [14].

It is of interest to have a closer look at the values for $P(\eta)$ obtained by using (7) as compared to (2). Now, (2) can be rewritten in the form

$$
P(\eta) \approx P_1(\eta) = \prod_{j=1}^{N} \left(1 - \alpha_{1j}(\eta)\right),
\tag{8}
$$

where $\alpha_{1j}(\eta) = \text{Prob}\{X_j > \eta\}$, $j = 1, \ldots, N$. Then the approximation based on assuming independent data can be written as

$$P(\eta) \approx F_1(\eta) := \exp\left(-\sum_{j=1}^{N} \alpha_{1j}(\eta)\right), \quad \eta \longrightarrow \infty. \quad (9)$$

Alternatively, (7) can be expressed as,

$$P(\eta) \approx P_k(\eta) = \prod_{j=k}^{N}\left(1 - \alpha_{kj}(\eta)\right)\prod_{j=1}^{k-1}\left(1 - \alpha_{jj}(\eta)\right), \quad (10)$$

where $\alpha_{kj}(\eta) = \text{Prob}\{X_j > \eta \mid X_{j-1} \leq \eta, \ldots, X_{j-k+1} \leq \eta\}$, for $j \geq k \geq 2$, denotes the exceedance probability conditional on $k - 1$ previous nonexceedances. From (10) it is now obtained that

$$P(\eta) \approx F_k(\eta) := \exp\left(-\sum_{j=k}^{N}\alpha_{kj}(\eta) - \sum_{j=1}^{k-1}\alpha_{jj}(\eta)\right),$$

$$\eta \longrightarrow \infty, \quad (11)$$

and $F_k(\eta) \to P(\eta)$ as $k \to N$ with $F_N(\eta) = P(\eta)$ for $\eta \to \infty$.

For the cascade of approximations $F_k(\eta)$ to have practical significance, it is implicitly assumed that there is a cut-off value k_c satisfying $k_c \ll N$ such that effectively $F_{k_c}(\eta) = F_N(\eta)$. It may be noted that for k-dependent stationary data sequences, that is, for data where X_i and X_j are independent whenever $|j - i| > k$, then $P(\eta) = P_{k+1}(\eta)$ exactly, and, under rather mild conditions on the joint distributions of the data, $\lim_{N \to \infty} P_1(\eta) = \lim_{N \to \infty} P(\eta)$ [15]. In fact, it can be shown that $\lim_{N \to \infty} P_1(\eta) = \lim_{N \to \infty} P(\eta)$ is true for weaker conditions than k-dependence [16]. However, for finite values of N the picture is much more complex, and purely asymptotic results should be used with some caution. Cartwright [17] used the notion of k-dependence to investigate the effect on extremes of correlation in sea wave data time series.

Returning to (11), extreme value prediction by the conditioning approach described above reduces to estimation of (combinations) of the $\alpha_{kj}(\eta)$ functions. In accordance with the previous assumption about a cut-off value k_c, for all k-values of interest, $k \ll N$, 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, the following approximation is adopted, which is applicable to both stationary and nonstationary data,

$$F_k(\eta) = \exp\left(-\sum_{j=k}^{N}\alpha_{kj}(\eta)\right), \quad k \geq 1. \quad (12)$$

Going back to 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 $(0, T)$. Equation (9) therefore expresses the approximation that the stream of exceedance events constitute a (nonstationary) Poisson process. This opens for an understanding of (12) 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

The concept of average conditional exceedance rate (ACER) of order k is now introduced as follows:

$$\varepsilon_k(\eta) = \frac{1}{N - k + 1}\sum_{j=k}^{N}\alpha_{kj}(\eta), \quad k = 1, 2, \ldots. \quad (13)$$

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 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 [18, 19].

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 nonexceedances, 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:

$$A_{kj}(\eta) = \mathbf{1}\left\{X_j > \eta, X_{j-1} \leq \eta, \ldots, X_{j-k+1} \leq \eta\right\},$$

$$j = k, \ldots, N, \ k = 2, 3, \ldots,$$

$$\qquad\qquad (14)$$

$$B_{kj}(\eta) = \mathbf{1}\left\{X_{j-1} \leq \eta, \ldots, X_{j-k+1} \leq \eta\right\},$$

$$j = k, \ldots, N, \ k = 2, \ldots,$$

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

$$\alpha_{kj}(\eta) = \frac{\mathbb{E}\left[A_{kj}(\eta)\right]}{\mathbb{E}\left[B_{kj}(\eta)\right]}, \quad j = k, \ldots, N, \ k = 2, \ldots, \quad (15)$$

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

$$\varepsilon_k(\eta) = \lim_{N \to \infty}\frac{\sum_{j=k}^{N} a_{kj}(\eta)}{\sum_{j=k}^{N} b_{kj}(\eta)}, \quad (16)$$

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.

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

$$\tilde{\varepsilon}_k(\eta) = \frac{\sum_{j=k}^{N}\mathbb{E}\left[A_{kj}(\eta)\right]}{N - k + 1}. \quad (17)$$

The advantage of using the modified ACER function $\widetilde{\varepsilon}_k(\eta)$ for $k \geq 2$ is that it is easier to use for nonstationary 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 (17) may be applicable for nonstationary time series, it is recognized that

$$P(\eta) \approx \exp\left(-\sum_{j=k}^{N} \alpha_{kj}(\eta)\right) = \exp\left(-\sum_{j=k}^{N} \frac{\mathbb{E}\left[A_{kj}(\eta)\right]}{\mathbb{E}\left[B_{kj}(\eta)\right]}\right)$$
$$\underset{\eta \to \infty}{\simeq} \exp\left(-\sum_{j=k}^{N} \mathbb{E}\left[A_{kj}(\eta)\right]\right). \tag{18}$$

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,

$$P(\eta) \approx \exp\left(-(N - k + 1)\widehat{\varepsilon}_k(\eta)\right), \tag{19}$$

where

$$\widehat{\varepsilon}_k(\eta) = \frac{1}{N - k + 1}\sum_{j=k}^{N} a_{kj}(\eta). \tag{20}$$

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 $k - 1$ nonexceedances. 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 [7, 20]. For alternative statistical approaches to account for the effect of clustering on the extreme value distribution, the reader may consult [21–26]. In these works, the emphasis is on the notion of an extremal index, which characterizes the clumping or clustering tendency of the data and its effect on the extreme value distribution. In the ACER functions, these effects are automatically accounted for.

Now, let us look at the problem of estimating a confidence interval for $\varepsilon_k(\eta)$, assuming a stationary time series. If R realizations of the requisite length of the time series is available, or, if one long realization can be segmented into R subseries, then the sample standard deviation $\widehat{s}_k(\eta)$ can be estimated by the standard formula

$$\widehat{s}_k(\eta)^2 = \frac{1}{R - 1}\sum_{r=1}^{R}\left(\widehat{\varepsilon}_k^{(r)}(\eta) - \widehat{\varepsilon}_k(\eta)\right)^2, \tag{21}$$

where $\widehat{\varepsilon}_k^{(r)}(\eta)$ denotes the ACER function estimate from realization no. r, and $\widehat{\varepsilon}_k(\eta) = \sum_{r=1}^{R} \widehat{\varepsilon}_k^{(r)}(\eta)/R$.

Assuming that realizations are independent, for a suitable number R, for example, $R \geq 20$, (21) leads to a good approximation of the 95% confidence interval CI $= (C^-(\eta), C^+(\eta))$ for the value $\varepsilon_k(\eta)$, where

$$C^\pm(\eta) = \widehat{\varepsilon}_k(\eta) \pm \frac{1.96\,\widehat{s}_k(\eta)}{\sqrt{R}}. \tag{22}$$

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 $\widehat{E}_k(\eta)$ of $\widetilde{\varepsilon}_k(\eta)$, where

$$\widehat{E}_k(\eta) = \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{N - k + 1} \tag{23}$$

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

$$C^\pm(\eta) = \widehat{\varepsilon}_k(\eta)\left(1 \pm \frac{1.96}{\sqrt{(N - k + 1)\widehat{\varepsilon}_k(\eta)}}\right). \tag{24}$$

4. Estimation of Extremes for the Asymptotic Gumbel Case

The second component of the approach to extreme value estimation presented in this paper was originally derived for a time series with an asymptotic extreme value distribution of the Gumbel type, compared with [27]. We have therefore chosen to highlight this case first, also because the extension of the asymptotic distribution to a parametric class of extreme value distribution tails that are capable of capturing to some extent subasymptotic behaviour is more transparent, and perhaps more obvious, for the Gumbel case. The reason behind the efforts to extend the extreme value distributions to the subasymptotic range is the fact that the ACER functions allow us to use not only asymptotic data, which is clearly an advantage since proving that observed extremes are truly asymptotic is really a nontrivial task.

The implication of the asymptotic distribution being of the Gumbel type on the possible subasymptotic 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,

$$\varepsilon_k(\eta) = q_k(\eta) \exp\left\{-a_k(\eta - b_k)^{c_k}\right\}, \quad \eta \geq \eta_1, \qquad (25)$$

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 $c_k = q_k(\eta) = 1$ corresponds to the asymptotic Gumbel distribution, which is then a special case of the assumed tail behaviour.

From (25) it follows that

$$-\log\left|\log\left(\frac{\varepsilon_k(\eta)}{q_k(\eta)}\right)\right| = -c_k \log(\eta - b_k) - \log(a_k). \qquad (26)$$

Therefore, under the assumptions made, a plot of $-\log|\log(\varepsilon_k(\eta)/q_k(\eta))|$ 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, say q_k, 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,

$$\varepsilon_k(\eta) = q_k \exp\left\{-a_k(\eta - b_k)^{c_k}\right\}, \quad \eta \geq \eta_1, \qquad (27)$$

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, subasymptotic 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 regular 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, compared with [7, 8].

As to the question of finding the parameters a, b, c, q (the subscript k, if it applies, is suppressed), the adopted approach is to determine these parameters by minimizing the following mean square error function, with respect to all four arguments,

$$F(a, b, c, q) = \sum_{j=1}^{J} w_j \left|\log \widehat{\varepsilon}(\eta_j) - \log q + a(\eta_j - b)^c\right|^2, \qquad (28)$$

where $\eta_1 < \cdots < \eta_J$ 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 $\widehat{\varepsilon}(\eta_j)$. The choice of weight factor is to some extent arbitrary. We have previously used $w_j = (\log C^+(\eta_j) - \log C^-(\eta_j))^{-\theta}$ with $\theta = 1$ and 2, combined with a Levenberg-Marquardt least squares optimization method [28]. This has usually worked well provided reasonable and 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 (28) has to stop before that happens. Also, the data should be preconditioned by establishing the tail marker η_1 based on inspection of the empirical ACER functions.

In general, to improve robustness of results, it is recommended to apply a nonlinearly constrained optimization [29]. The set of constraints is written as

$$\log q - a(\eta_i - b)^c \leq 0,$$

$$0 < q < +\infty,$$

$$\min_j X_j < b \leq \eta_1, \qquad (29)$$

$$0 < a < +\infty,$$

$$0 < c < 5.$$

Here, the first nonlinear inequality constraint is evident, since under our assumption we have $\widehat{\varepsilon}_k(\eta_i) = q \exp\{-a(\eta_i - b)^c\}$, and $\widehat{\varepsilon}_k(\eta_i) < 1$ by definition.

A Note of Caution. When the parameter c is equal to 1.0 or close to it, that is, the distribution is close to the Gumbel distribution, 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 (28) 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 $\sigma_{yj}^2 = \text{Var}[y_j]$ (empirical) [30, 31]. Unfortunately, this is not a very practical weight factor for the kind of problem we have here because the summation in (28) 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

$$a^* (b, c) = -\frac{\sum_{j=1}^{N} w_j \left(x_j - \overline{x}\right) \left(y_j - \overline{y}\right)}{\sum_{j=1}^{N} w_j \left(x_j - \overline{x}\right)^2}, \quad (30)$$

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

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 (30) to calculate the corresponding a^* and q^*.

For a simple 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 $\text{CI}_{0.95}$ 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 first estimate of a 95% confidence interval of the predicted extreme value. To obtain a more precise estimate of the confidence interval, a bootstrapping method would be recommended. A comparison of estimated confidence intervals by both these methods will be presented in the section on extreme value prediction for synthetic data. 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. This property is easily recognized by looking at the way the optimized fitting is done. If the tail marker is in the appropriate domain of the ACER function, the optimal fitted curve does not change appreciably by moving the tail marker.

5. Estimation of Extremes for the General Case

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

$$\varepsilon_1 (\eta) \underset{\eta \to \infty}{\simeq} \left[1 + \xi \left(a \left(\eta - b\right)\right)\right]^{-1/\xi}, \quad (31)$$

where $a > 0$, b, ξ are constants. This follows from the explicit form of the so-called generalized extreme value (GEV) distribution Coles [1].

Again, the implication of this assumption on the possible subasymptotic 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. While the extension of the asymptotic Gumbel case to the proposed class of subasymptotic distributions was fairly transparent, this is not equally so for the general case. However, using a similar kind

of approximation, the behaviour of the mean exceedance rate in the subasymptotic part of the tail is assumed to follow a function largely of the form $[1 + \xi(a(\eta - b)^c)]^{-1/\xi}$ ($\eta \geq \eta_1 \geq b$), 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 [32]

$$\varepsilon_k (\eta) = q_k (\eta) \left[1 + \xi_k \left(a_k (\eta - b_k)^{c_k}\right)\right]^{-1/\xi_k}, \quad \eta \geq \eta_1, \quad (32)$$

where the function $q_k(\eta)$ is weakly varying, compared with the function $[1 + \xi_k(a_k(\eta - b_k)^{c_k})]^{-1/\xi_k}$ 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 (25). Another method to account for subasymptotic effects has recently been proposed by Eastoe and Tawn [33], building on ideas developed by Tawn [34], Ledford and Tawn [35] and Heffernan and Tawn [36]. In this approach, the asymptotic form of the marginal distribution of exceedances is kept, but it is modified by a multiplicative factor accounting for the dependence structure of exceedances within a cluster.

An alternative form to (32) would be to assume that

$$\varepsilon_k (\eta) = \left[1 + \xi_k \left(a_k (\eta - b_k)^{c_k} + d_k (\eta)\right)\right]^{-1/\xi_k}, \quad \eta \geq \eta_1, \quad (33)$$

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 (25) is preferable 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 (32), it is 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 will suppress the index k on the ACER functions, which will then be written as

$$\varepsilon (\eta) = q \left[1 + \tilde{a} \left(\eta - b\right)^c\right]^{-\gamma}, \quad \eta \geq \eta_1, \quad (34)$$

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

$$F (\tilde{a}, b, c, q, \gamma) = \sum_{j=1}^{N} w_j \Big| \log \hat{\varepsilon} \left(\eta_j\right) - \log q \\ + \gamma \log \left[1 + \tilde{a}\left(\eta_j - b\right)^c\right] \Big|^2, \quad (35)$$

where w_j is a weight factor as previously defined.

An option for estimating the five parameters $\tilde{a}$, b, c, q, γ 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 (28), 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 (30). To calculate the final optimal set of parameters, the Levenberg-Marquardt method may then be used on the function $\widetilde{F}(\tilde{a}, b, c) = F(\tilde{a}, b, c, q^*(\tilde{a}, b, c), \gamma^*(\tilde{a}, b, c))$ 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, for example, also be found by a sequential quadratic programming (SQP) method [37].

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 will use the predictions given by the two methods that seem to be most favored by practitioners, the Gumbel method and the peaks-over-threshold (POT) method, provided, of course, that the correct asymptotic extreme value distribution is of the Gumbel type.

The Gumbel method is based on recording epochal extreme values and fitting these values to a corresponding Gumbel distribution [38]. 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

$$\text{Prob}\left(M_N \leq \eta\right) = \exp\left\{-\exp\left(-a\left(\eta - b\right)\right)\right\}, \quad (36)$$

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: $b = m_M - 0.57722a^{-1}$ and $a = 1.28255/\sigma_M$ [39]. 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 $\widetilde{N}$, 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 [40, 41]. Note that the assumption that the initial $\widetilde{N}$

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

7.1. The Generalized Pareto Distribution. The POT method for independent data is based on what is called the generalized Pareto (GP) distribution (defined below) in the following manner: it has been shown in [42] 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

$$G\left(y\right) = G\left(y; a, c\right) = \text{Prob}\left(Y \leq y\right) = 1 - \left(1 + c\,\frac{y}{a}\right)_{+}^{-1/c}. \quad (37)$$

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 (37) can be used to represent the conditional cumulative distribution function of the excess $Y = X - u$ of the observed variates X over the threshold u, given that $X > u$ for u is sufficiently large [42]. 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, compared with section below.

For $c = 0$, which corresponds to the Gumbel extreme value distribution, the expression between the parentheses in (37) is understood in a limiting sense as the exponential distribution

$$G\left(y\right) = G\left(y; a, 0\right) = \exp\left(-\frac{y}{a}\right). \quad (38)$$

Since the recorded data in practice are rarely independent, a declustering technique is commonly used to filter the data to achieve approximate independence [1, 2].

7.2. Return Periods. The return period R of a given value x_R of X in terms of a specified length of time τ, for example, a year, is defined as the inverse of the probability that the specified value will be exceeded in any time interval of length τ. If λ denotes the mean exceedance rate of the threshold u per length of time τ (i.e., the average number of data points above the threshold u per τ), then the return period R of the value of X corresponding to the level $x_R = u + y$ is given by the relation

$$R = \frac{1}{\lambda \text{Prob}\left(X > x_R\right)} = \frac{1}{\lambda \text{Prob}\left(Y > y\right)}. \quad (39)$$

Hence, it follows that

$$\text{Prob}\,(Y \le y) = 1 - \frac{1}{(\lambda R)}. \tag{40}$$

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

$$x_R = u - \frac{a\left[1 - (\lambda R)^c\right]}{c}. \tag{41}$$

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

$$x_R = u + a \ln\,(\lambda R), \tag{42}$$

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 will 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 [43]. 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 $v^+(0)$ is such that the product $v^+(0)T = 10^3$, 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

$$F^{1\,\text{yr}}\,(\eta) = \exp\left\{-v^+\,(\eta)\,T\right\} = \exp\left\{-10^3 \exp\left(-\frac{\eta^2}{2}\right)\right\}, \tag{43}$$

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

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, as required by the POT method. In accordance with this, peak event data are generated from the extreme value distribution

$$F^{3\,\text{d}}\,(\eta) = \exp\left\{-q \exp\left(-\frac{\eta^2}{2}\right)\right\}, \tag{44}$$

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

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 (22) was based on 20 estimates of the ACER function using the yearly data. This provided a 95% confidence band on the optimally fitted curve based on 2000 data. From these data, the predicted 100-year return level is obtained from $\widehat{\varepsilon}_1(\eta^{100\,\text{yr}}) = 10^{-4}$. A nonparametric bootstrapping method was also used to estimate a 95% confidence interval based on 1000 resamples of size 2000.

The POT prediction of the 100-year return level was based on using maximum likelihood estimates (MLE) of the parameters in (37) for a specific choice of threshold. The 95% confidence interval was obtained from the parametrically bootstrapped PDF of the POT estimate for the given threshold. A sample of 1000 data sets was used. One of the unfortunate features of the POT method is that the estimated 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 [1].

Note that in spite of the fact that the true asymptotic distribution of exceedances is the exponential distribution in (38), the POT method used here is based on adopting (37). The reason is simply that this is the recommended procedure [1], which is somewhat unfortunate but understandable. The reason being that the GP distribution provides greater flexibility in terms of curve fitting. If the correct asymptotic distribution of exceedances had been used on this example, poor results for the estimated return period values would be obtained. The price to pay for using the GP distribution is that the estimated parameters may easily lead to an asymptotically inconsistent extreme value distribution.

The 100-year return level predicted by the Gumbel method was based on using the method of moments for parameter estimation on the sample of 20 yearly extremes. This choice of estimation method is due to the small sample of extreme values. The 95% confidence interval was obtained from the parametrically bootstrapped PDF of the Gumbel prediction. This was based on a sample of size 10,000 data sets of 20 yearly extremes. The results obtained by the method of moments were compared with the corresponding results obtained by using the maximum likelihood method. While there were individual differences, the overall picture was one of very good agreement.

In order to get an idea about the performance of the ACER, POT, and Gumbel methods, 100 independent 20-year MC simulations as discussed above were done. Table 1 compares predicted values and confidence intervals for a selection of 10 cases together with average values over the 100 simulated cases. It is seen that the average of the 100 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 4.34–5.36, while the same for the POT method is 4.19–5.87 and for the Gumbel method is 4.41–5.71.

TABLE 1: 100-year return level estimates and 95% CI (BCI = CI by bootstrap) for A = ACER, G = Gumbel, and P = POT. Exact value = 4.80.

Sim. No.	A $\widehat{\eta}^{100}$	A CI	A BCI	G $\widehat{\eta}^{100}$	G BCI	P $\widehat{\eta}^{100}$	P BCI
1	5.07	(4.67, 5.21)	(4.69, 5.42)	4.41	(4.14, 4.73)	4.29	(4.13, 4.52)
10	4.65	(4.27, 4.94)	(4.37, 5.03)	4.92	(4.40, 5.58)	4.88	(4.42, 5.40)
20	4.86	(4.49, 5.06)	(4.44, 5.19)	5.04	(4.54, 5.63)	5.04	(4.48, 5.74)
30	4.75	(4.22, 5.01)	(4.33, 5.02)	4.75	(4.27, 5.32)	4.69	(4.24, 5.26)
40	4.54	(4.20, 4.74)	(4.27, 4.88)	4.80	(4.31, 5.39)	4.73	(4.19, 5.31)
50	4.80	(4.35, 5.05)	(4.42, 5.14)	4.91	(4.41, 5.50)	4.79	(4.31, 5.34)
60	4.84	(4.36, 5.20)	(4.48, 5.19)	4.85	(4.36, 5.43)	4.71	(4.32, 5.23)
70	5.02	(4.47, 5.31)	(4.62, 5.36)	4.96	(4.47, 5.53)	4.97	(4.47, 5.71)
80	4.59	(4.33, 4.81)	(4.38, 4.98)	4.76	(4.31, 5.31)	4.68	(4.15, 5.27)
90	4.84	(4.49, 5.11)	(4.60, 5.30)	4.77	(4.34, 5.32)	4.41	(4.23, 4.64)
100	4.62	(4.29, 5.05)	(4.45, 5.09)	4.79	(4.31, 5.41)	4.53	(4.05, 4.88)
Av. 100	4.82	(4.41, 5.09)	(4.48, 5.18)	4.84	(4.37, 5.40)	4.72	(4.27, 5.23)

Hence, in this case the ACER method performs consistently better than both these methods. It is also observed from the estimated 95% confidence intervals that the ACER method, as implemented in this paper, provides slightly higher accuracy than the other two methods. Lastly, it is pointed out that the confidence intervals of the 100-year return level estimated by the ACER method obtained by either the simplified extrapolated confidence band approach or by nonparametric bootstrapping are very similar, except for a slight mean shift. As a final comparison, the 100 bootstrapped confidence intervals obtained for the ACER and Gumbel methods missed the target value three times, while for the POT method this number was 18.

An example of the ACER plot and results obtained for one set of data is presented in Figure 1. The predicted 100-year value is 4.85 with a predicted 95% confidence interval (4.45, 5.09). Figure 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 4.7 at $n = 204$, while the 95% confidence interval is (4.25, 5.28). The same data set as in Figure 1 was used. This was also used for the Gumbel plot shown in Figure 3. In this case the predicted value based on the method of moments (MM) is $\eta_{MM}^{100\,yr} = 4.75$ with a parametric bootstrapped 95% confidence interval of (4.34, 5.27). Prediction based on the Gumbel-Lieblein BLUE method (GL), compared with for example, Cook [44], is $\eta_{GL}^{100\,yr} = 4.73$ with a parametric bootstrapped 95% confidence interval equal to (4.35, 5.14).

9. Measured Wind Speed Data

In this section, we analyze real wind speed data, measured at two weather stations off the coast of Norway: at Nordøyan and at Hekkingen, see Figure 4. Extreme wind speed prediction is an important issue for design of structures exposed to the weather variations. Significant 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, for example, [45–48] for extensive references to previous work.
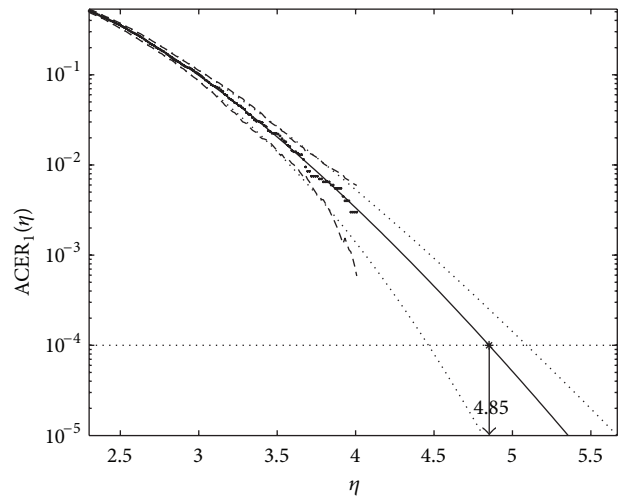

FIGURE 1: Synthetic data ACER $\widehat{\varepsilon}_1$, Monte Carlo simulation ($*$); optimized curve fit (—); empirical 95% confidence band (- -); optimized confidence band ($\cdots$). Tail marker $\eta_1 = 2.3$.

Hourly maximum gust wind was recorded during the 13 years 1999–2012 at Nordøyan and the 14 years 1998–2012 at Hekkingen. The objective is to estimate a 100-year 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. This may not be entirely true, of course.

9.1. Nordøyan. Figure 5 highlights the cascade of ACER estimates $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_{96}$, for the case of 13 years of hourly data recorded at the Nordøyan weather station. Here, $\widehat{\varepsilon}_{96}$ is considered to represent the final converged results. By "converged," we mean that $\widehat{\varepsilon}_{96} \approx \widehat{\varepsilon}_k$ for $k > 96$ in the tail, so that there is no
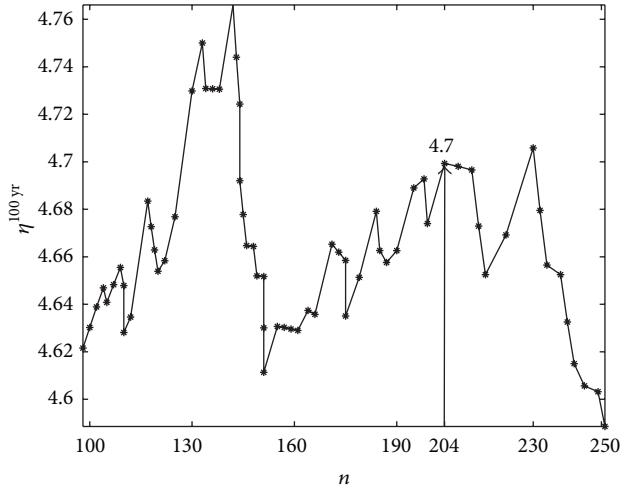
FIGURE 2: The point estimate $\widetilde{\eta}^{100\,yr}$ of the 100-year return period value based on 20 years synthetic data as a function of the number n of data points above threshold. The return level estimate = 4.7 at $n = 204$.
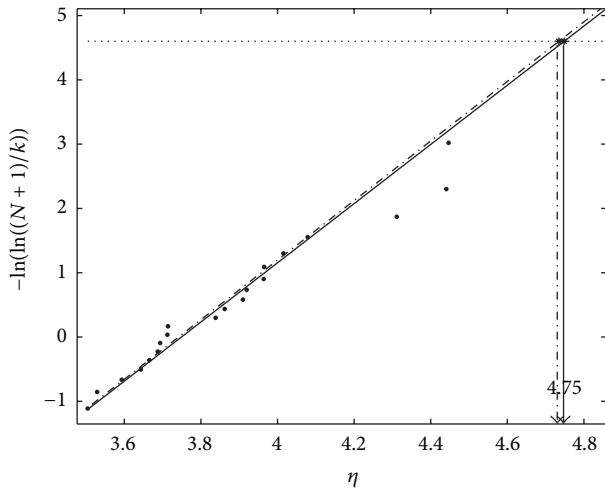

FIGURE 3: The point estimate $\widetilde{\eta}^{100\,yr}$ of the 100-year return period value based on 20 years synthetic data. Lines are fitted by the method of moments—solid line (—) and the Gumbel-Lieblein BLUE method—dash-dotted lite (-·-). The return level estimate by the method of moments is 4.75, by the Gumbel-Lieblein BLUE method is 4.73.

FIGURE 4: Wind speed measurement stations.

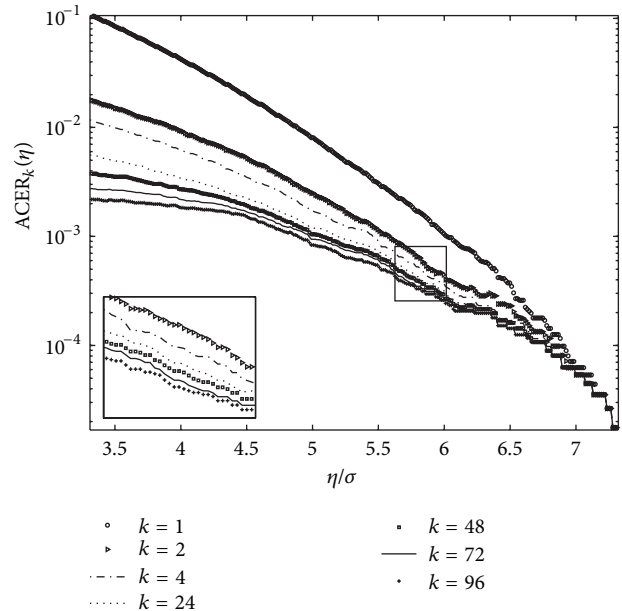| ∘ | $k = 1$ | ▪ | $k = 48$ |
▹	$k = 2$	—	$k = 72$
-·-·	$k = 4$	·	$k = 96$
⋯⋯	$k = 24$		

FIGURE 5: Nordøyan wind speed statistics, 13 years hourly data. Comparison between ACER estimates for different degrees of conditioning. $\sigma = 6.01\,m/s$.

need to consider conditioning of an even higher order than 96. Figure 5 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 $\widehat{\varepsilon}_2$, 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 96th ACER function, as discussed above.

However, from a practical point of view, the most important information provided by the ACER plot of Figure 5 is that for the prediction of a 100-year value, one may use the first ACER function. The reason for this is that Figure 5 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.

In Figure 6 is shown the results of parametric estimation of the return value and its 95% CI for 13 years of hourly
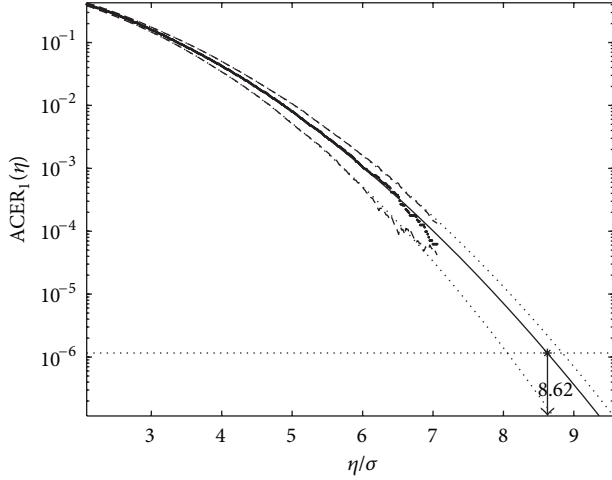
FIGURE 6: Nordøyan wind speed statistics, 13 years hourly data. $\widehat{\varepsilon}_1(\eta)$ (∗); optimized curve fit (—); empirical 95% confidence band (- -); optimized confidence band (⋯). Tail marker $\eta_1 = 12.5$ m/s $= 2.08\sigma$ ($\sigma = 6.01$ m/s).
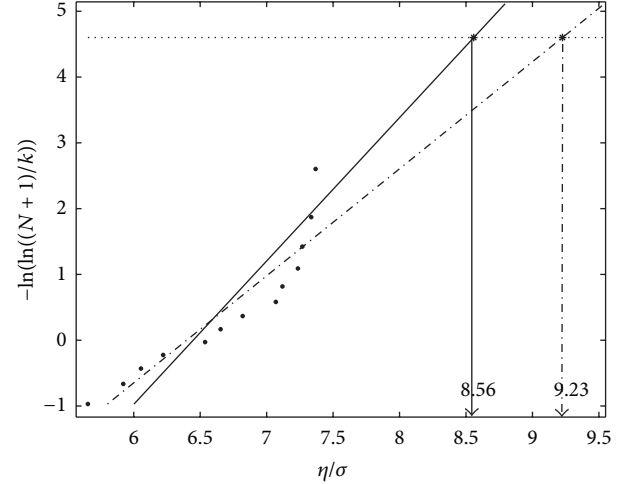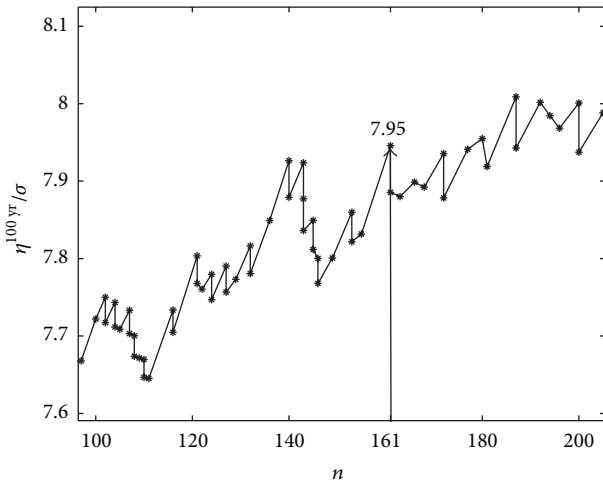

FIGURE 8: Nordøyan wind speed statistics, 13 years of hourly data. Gumbel plot of yearly extremes. Lines are fitted by the method of moments—solid line (—) and the Gumbel-Lieblein BLUE method—dash-dotted lite (- ⋅ -). $\sigma = 6.01$ m/s.

FIGURE 7: The point estimate $\eta^{100\,\mathrm{yr}}$ of the 100-year return level based on 13 years hourly data as a function of the number n of data points above threshold. $\sigma = 6.01$ m/s.

TABLE 2: Predicted 100-year return period levels for Nordøyan Fyr weather station by the ACER method for different degrees of conditioning, annual maxima, and POT methods, respectively.

Method	Spec	$\widehat{\eta}^{100\,\mathrm{yr}}$, m/s	95% CI ($\widehat{\eta}^{100\,\mathrm{yr}}$), m/s
	1	51.85	(48.4, 53.1)
	2	51.48	(46.1, 54.1)
	4	52.56	(46.7, 55.7)
ACER, various k	24	52.90	(47.0, 56.2)
	48	54.62	(47.7, 57.6)
	72	53.81	(46.9, 58.3)
	96	54.97	(47.5, 60.5)
Annual maxima	MM	51.5	(45.2, 59.3)
	GL	55.5	(48.0, 64.9)
POT	—	47.8	(44.8, 52.7)

maxima. The predicted 100-year return speed is $\eta^{100\,\mathrm{yr}} = 51.85$ m/s with 95% confidence interval (48.4, 53.1). $R = 13$ years of data may not be enough to guarantee (22), since we required $R \geq 20$. Nevertheless, for simplicity, we use it here even with $R = 13$, accepting that it may not be very accurate.

Figure 7 presents POT predictions for different threshold numbers based on MLE. The POT prediction is $\eta^{100\,\mathrm{yr}} = 47.8$ m/s at threshold $n = 161$, while the bootstrapped 95% confidence interval is found to be (44.8, 52.7) m/s based on 10,000 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.

Figure 8 presents a Gumbel plot based on the 13 yearly extremes extracted from the 13 years of hourly data. The

Gumbel prediction based on the method of moments (MM) is $\eta_{\mathrm{MM}}^{100\,\mathrm{yr}} = 51.5$ m/s, with a parametric bootstrapped 95% confidence interval equal to (45.2, 59.3) m/s, while prediction based on the Gumbel-Lieblein BLUE method (GL) is $\eta_{\mathrm{GL}}^{100\,\mathrm{yr}} = 55.5$ m/s, with a parametric bootstrapped 95% confidence interval equal to (48.0, 64.9) m/s.

In Table 2 the 100-year return period values for the Nordøyan station are listed together with the predicted 95% confidence intervals for all methods.

9.2. Hekkingen.

Figure 9 shows the cascade of estimated ACER functions $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_{96}$ for the case of 14 years of hourly data. As for Nordøyan, $\widehat{\varepsilon}_{96}$ is used to represent the final converged results. Figure 9 also reveals a rather strong statistical dependence between consecutive data at moderate wind speed levels. This effect is again to some extent captured already by $\widehat{\varepsilon}_2$, so that subsequent conditioning on more than one previous data point does not lead to substantial changes in ACER values, especially for tail values.
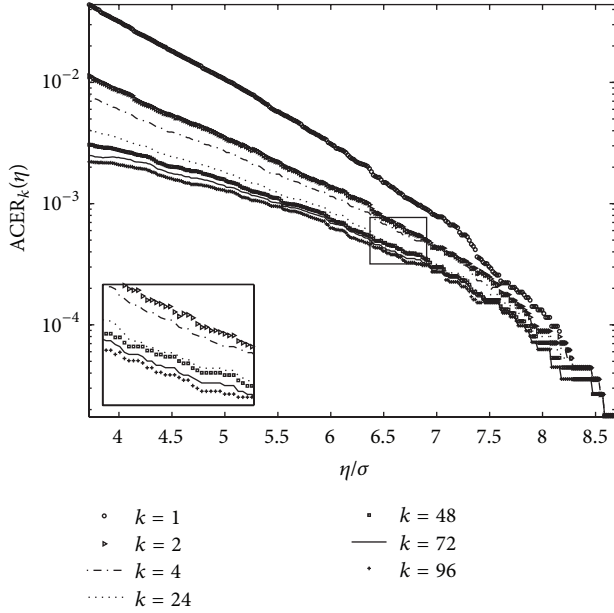
FIGURE 9: Hekkingen wind speed statistics, 14 years hourly data. Comparison between ACER estimates for different degrees of conditioning. $\sigma = 5.72$ m/s.
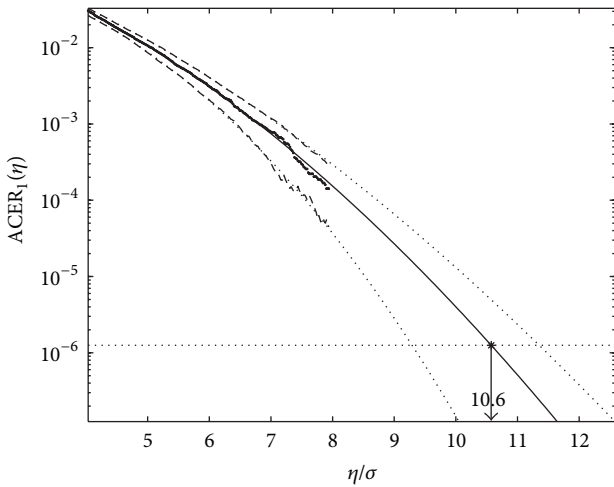

FIGURE 10: Hekkingen wind speed statistics, 14 years hourly data. $\hat{\varepsilon}_1(\eta)$ ($*$); optimized curve fit (—); empirical 95% confidence band (- -); optimized confidence band ($\cdots$). Tail marker $\eta_1 = 23$ m/s = 4.02σ ($\sigma = 5.72$ m/s).

TABLE 3: Predicted 100-year return period levels for Nordøyan Fyr weather station by the ACER method for different degrees of conditioning, annual maxima, and POT methods, respectively.

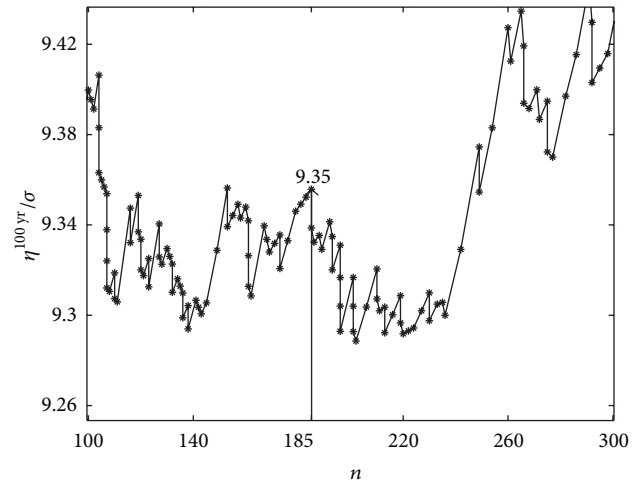| Method	Spec	$\hat{\eta}^{100\,\text{yr}}$, m/s	95% CI $(\hat{\eta}^{100\,\text{yr}})$, m/s
	1	60.47	(53.1, 64.9)
	2	62.23	(53.3, 70.0)
	4	63.03	(53.0, 74.5)
ACER, various k	24	60.63	(51.3, 70.7)
	48	60.44	(51.3, 77.0)
	72	58.06	(51.2, 66.4)
	96	59.19	(52.0, 68.3)
Annual maxima	MM	58.10	(50.8, 67.3)
	GL	60.63	(53.0, 70.1)
POT	—	53.48	(48.9, 57.0)

FIGURE 11: The point estimate $\eta^{100\,\text{yr}}$ of the 100-year return level based on 14 years hourly data as a function of the number n of data points above threshold. $\sigma = 5.72$ m/s.

Also, for the Hekkingen data, the ACER plot of Figure 9 indicates that the ACER functions coalesce in the far tail. Hence, for the practical prediction of a 100-year value, one may use the first ACER function.

In Figure 10 is shown the results of parametric estimation of the return value and its 95% CI for 14 years of hourly maxima. The predicted 100-year return speed is $\eta^{100\,\text{yr}} = 60.47$ m/s with 95% confidence interval (53.1, 64.9). Equation (22) has been used also for this example with $R = 14$.

Figure 11 presents POT predictions for different threshold numbers based on MLE. The POT prediction is $\eta^{100\,\text{yr}} = 53.48$ m/s at threshold $n = 183$, while the bootstrapped

95% confidence interval is found to be (48.9, 57.0) m/s based on 10,000 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.

Figure 12 presents a Gumbel plot based on the 14 yearly extremes extracted from the 14 years of hourly data. The Gumbel prediction based on the method of moments (MM) is $\eta_{\text{MM}}^{100\,\text{yr}} = 58.10$ m/s, with a parametric bootstrapped 95% confidence interval equal to (50.8, 67.3) m/s. Prediction based on the Gumbel-Lieblein BLUE method (GL) is $\eta_{\text{GL}}^{100\,\text{yr}} = 60.63$ m/s, with a parametric bootstrapped 95% confidence interval equal to (53.0, 70.1) m/s.

In Table 3, the 100-year return period values for the Hekkingen station are listed together with the predicted 95% confidence intervals for all methods.
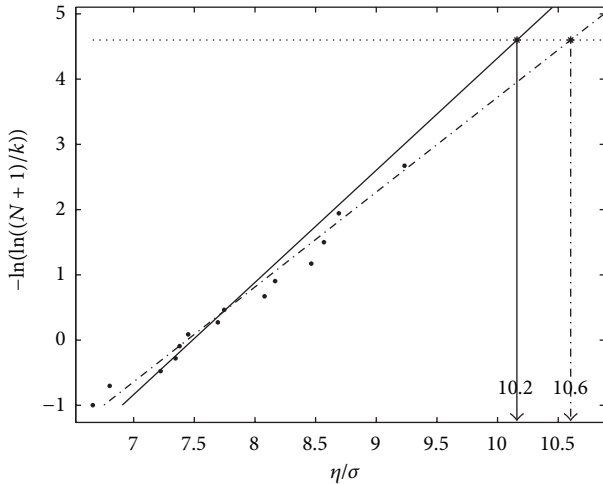
FIGURE 12: Hekkingen wind speed statistics, 14 years of hourly data. Gumbel plot of yearly extremes. Lines are fitted by the method of moments—solid line (—) and the Gumbel-Lieblein BLUE method—dash-dotted lite (- · -). $\sigma = 5.72$ m/s.

FIGURE 13: Part of the narrow-band response time series of the linear oscillator with fully sampled and peak values indicated.

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 will 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 Figure 13. A consequence of this is that neighbouring peaks are strongly correlated, and there is a conspicuous clumping of the peak values. Hence the problem with accurate prediction, since the usual assumption of independent peak values is then violated.

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 Figure 14 are shown some of the ACER functions for the example time series. It may be verified from Figure 13 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
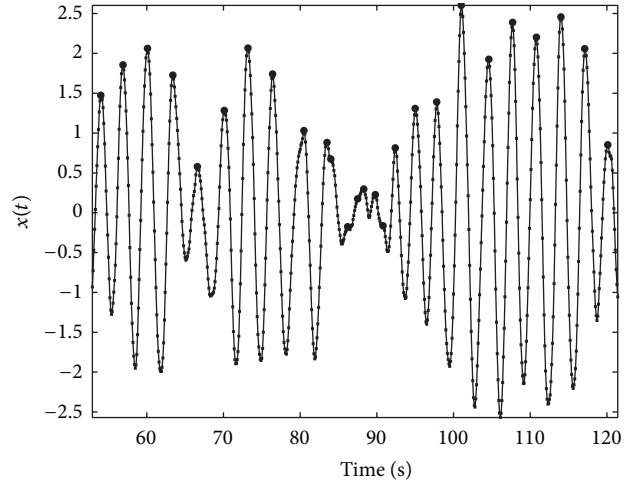
the sampled data points are independent, which is obviously completely wrong. The second ACER function, which is based on counting each exceedance with an immediately preceding nonexceedance, is nothing but an upcrossing rate. Using this ACER function is largely equivalent to assuming independent peak 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 [49].

To illustrate the results obtained by extracting only the peak values from the time series, which would be the approach typically chosen in an engineering analysis, the ACER plots for this case is shown in Figure 15. By comparing results from Figures 14 and 15, it can be verified that they are in very close agreement by recognizing that the second ACER function in Figure 14 corresponds to the first ACER function in Figure 15, and by noting that there is a factor of approximately 32 between corresponding ACER functions in the two figures. This is due to the fact that the time series of peak values contains about 32 times less data than the original time series.
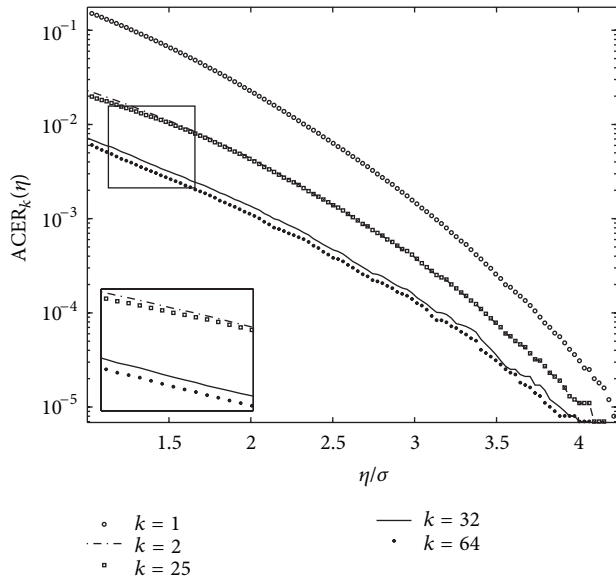
FIGURE 14: Comparison between ACER estimates for different degrees of conditioning for the narrow-band time series.
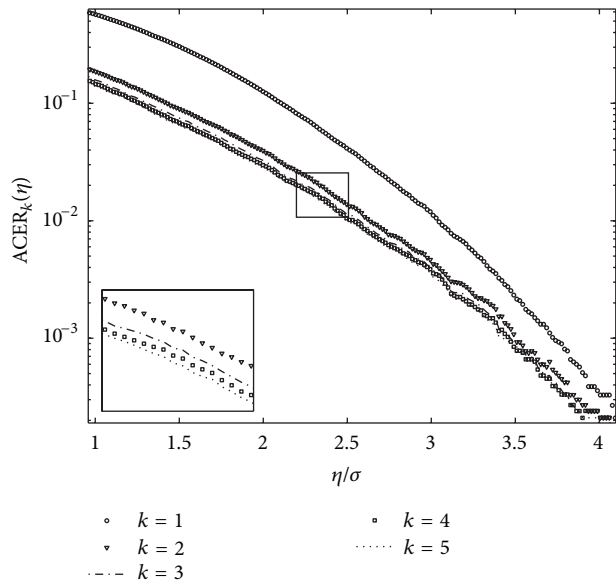

FIGURE 15: Comparison between ACER estimates for different degrees of conditioning based on the time series of the peak values, compared with Figure 13.

11. 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 (ACER), which allows dependence in the time series to be properly and easily 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 subasymptotic form of the extreme value distribution.

Results for wind speeds, both synthetic and measured, are used to illustrate the method. An estimation problem

related to applications in mechanics is 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 the peaks-over-threshold (POT) methods. Comparison of the various estimates indicate that the proposed method provides more accurate results than the Gumbel and POT methods.

Subject to certain restrictions, the proposed method also applies to nonstationary time series, but it cannot directly predict for example, 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 for example, the exact bivariate extreme value distribution for a suitable set of data [50]. However, as is easily recognized, the extrapolation problem is not as simply dealt with as for the univariate case studied in this paper.

Acknowledgment

References

[1] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics, Springer, London, UK, 2001.

[2] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers, *Statistics of Extremes*, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, UK, 2004.

[3] A. C. Davison and R. L. Smith, "Models for exceedances over high thresholds," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 52, no. 3, pp. 393–442, 1990.

[4] R.-D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values*, Birkhäuser, Basel, Switzerland, 3rd edition, 1997.

[5] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events*, vol. 33 of *Applications of Mathematics (New York)*, Springer, Berlin, Germany, 1997.

[6] M. Falk, J. Hüsler, and R.-D. Reiss, *Laws of Small Numbers: Extremes and Rare Events*, Birkhäuser, Basel, Switzerland, 2 nd edition, 2004.

[7] A. Naess and O. Gaidai, "Monte Carlo methods for estimating the extreme response of dynamical systems," *Journal of Engineering Mechanics*, vol. 134, no. 8, pp. 628–636, 2008.

[8] A. Naess, O. Gaidai, and S. Haver, "Efficient estimation of extreme response of drag-dominated offshore structures by Monte Carlo simulation," *Ocean Engineering*, vol. 34, no. 16, pp. 2188–2197, 2007.

[9] R. L. Smith, "The extremal index for a Markov chain," *Journal of Applied Probability*, vol. 29, no. 1, pp. 37–45, 1992.

[10] S. G. Coles, "A temporal study of extreme rainfall," in *Statistics for the Environment 2-Water Related Issues*, V. Barnett and K. F. Turkman, Eds., chapter 4, pp. 61–78, John Wiley & Sons, Chichester, UK, 1994.

[11] R. L. Smith, J. A. Tawn, and S. G. Coles, "Markov chain models for threshold exceedances," *Biometrika*, vol. 84, no. 2, pp. 249–268, 1997.

[12] S. Yun, "The extremal index of a higher-order stationary Markov chain," *The Annals of Applied Probability*, vol. 8, no. 2, pp. 408–437, 1998.

[13] S. Yun, "The distributions of cluster functionals of extreme events in a *d*th-order Markov chain," *Journal of Applied Probability*, vol. 37, no. 1, pp. 29–44, 2000.

[14] J. Segers, "Approximate distributions of clusters of extremes," *Statistics & Probability Letters*, vol. 74, no. 4, pp. 330–336, 2005.

[15] G. S. Watson, "Extreme values in samples from *m*-dependent stationary stochastic processes," *Annals of Mathematical Statistics*, vol. 25, pp. 798–800, 1954.

[16] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics, Springer, New York, NY, USA, 1983.

[17] D. E. Cartwright, "On estimating the mean energy of sea waves from the highest waves in a record," *Proceedings of the Royal Society of London. Series A*, vol. 247, pp. 22–28, 1958.

[18] A. Naess, "On the long-term statistics of extremes," *Applied Ocean Research*, vol. 6, no. 4, pp. 227–228, 1984.

[19] G. Schall, M. H. Faber, and R. Rackwitz, "Ergodicity assumption for sea states in the reliability estimation of offshore structures," *Journal of Offshore Mechanics and Arctic Engineering*, vol. 113, no. 3, pp. 241–246, 1991.

[20] E. H. Vanmarcke, "On the distribution of the first-passage time for normal stationary random processes," *Journal of Applied Mechanics*, vol. 42, no. 1, pp. 215–220, 1975.

[21] M. R. Leadbetter, "Extremes and local dependence in stationary sequences," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 65, no. 2, pp. 291–306, 1983.

[22] T. Hsing, "On the characterization of certain point processes," *Stochastic Processes and Their Applications*, vol. 26, no. 2, pp. 297–316, 1987.

[23] T. Hsing, "Estimating the parameters of rare events," *Stochastic Processes and Their Applications*, vol. 37, no. 1, pp. 117–139, 1991.

[24] M. R. Leadbetter, "On high level exceedance modeling and tail inference," *Journal of Statistical Planning and Inference*, vol. 45, no. 1-2, pp. 247–260, 1995.

[25] C. A. T. Ferro and J. Segers, "Inference for clusters of extreme values," *Journal of the Royal Statistical Society. Series B*, vol. 65, no. 2, pp. 545–556, 2003.

[26] C. Y. Robert, "Inference for the limiting cluster size distribution of extreme values," *The Annals of Statistics*, vol. 37, no. 1, pp. 271–310, 2009.

[27] A. Naess and O. Gaidai, "Estimation of extreme values from sampled time series," *Structural Safety*, vol. 31, no. 4, pp. 325–334, 2009.

[28] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press Inc, London, UK, 1981.

[29] W. Forst and D. Hoffmann, *Optimization—Theory and Practice*, Springer Undergraduate Texts in Mathematics and Technology, Springer, New York, NY, USA, 2010.

[30] N. R. Draper and H. Smith, *Applied Regression Analysis*, Wiley Series in Probability and Statistics: Texts and References Section, John Wiley & Sons, New York, NY, USA, 3rd edition, 1998.

[31] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, Wiley Series in Probability and Statistics: Texts, References, and Pocketbooks Section, Wiley-Interscience, New York, NY, USA, Third edition, 2001.

[32] A. Naess, *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, Norway, 2010.

[33] E. F. Eastoe and J. A. Tawn, "Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds," *Biometrika*, vol. 99, no. 1, pp. 43–55, 2012.

[34] J. A. Tawn, "Discussion of paper by A. C. Davison and R. L. Smith," *Journal of the Royal Statistical Society. Series B*, vol. 52, no. 3, pp. 393–442, 1990.

[35] A. W. Ledford and J. A. Tawn, "Statistics for near independence in multivariate extreme values," *Biometrika*, vol. 83, no. 1, pp. 169–187, 1996.

[36] J. E. Heffernan and J. A. Tawn, "A conditional approach for multivariate extreme values," *Journal of the Royal Statistical Society. Series B*, vol. 66, no. 3, pp. 497–546, 2004.

[37] Numerical Algorithms Group, *NAG Toolbox for Matlab*, NAG, Oxford, UK, 2010.

[38] E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, NY, USA, 1958.

[39] K. V. Bury, *Statistical Models in Applied Science*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 1975.

[40] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, vol. 57 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York, NY, USA, 1993.

[41] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and Their Application*, vol. 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, UK, 1997.

[42] J. Pickands, III, "Statistical inference using extreme order statistics," *The Annals of Statistics*, vol. 3, pp. 119–131, 1975.

[43] A. Naess and P. H. Clausen, "Combination of the peaks-over-threshold and bootstrapping methods for extreme value prediction," *Structural Safety*, vol. 23, no. 4, pp. 315–330, 2001.

[44] N. J. Cook, *The Designer's Guide to Wind Loading of Building Structures*, Butterworths, London, UK, 1985.

[45] N. J. Cook, "Towards better estimation of extreme winds," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 9, no. 3, pp. 295–323, 1982.

[46] A. Naess, "Estimation of long return period design values for wind speeds," *Journal of Engineering Mechanics*, vol. 124, no. 3, pp. 252–259, 1998.

[47] J. P. Palutikof, B. B. Brabson, D. H. Lister, and S. T. Adcock, "A review of methods to calculate extreme wind speeds," *Meteorological Applications*, vol. 6, no. 2, pp. 119–132, 1999.

[48] O. Perrin, H. Rootzén, and R. Taesler, "A discussion of statistical methods used to estimate extreme wind speeds," *Theoretical and Applied Climatology*, vol. 85, no. 3-4, pp. 203–215, 2006.

[49] M. E. Robinson and J. A. Tawn, "Extremal analysis of processes sampled at different frequencies," *Journal of the Royal Statistical Society. Series B*, vol. 62, no. 1, pp. 117–135, 2000.

[50] A. Naess, *A Note on the Bivariate ACER Method*, Preprint Statistics No. 01/2011, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway, 2011.