Prepared for:

Rijkswaterstaat, Rijksinstituut voor Kust en Zee (RIKZ)

Extreme wave statistics

Confidence intervals

Report

February, 2007

WL | delft hydraulics

H4803.30

Prepared for:

Rijkswaterstaat, Rijksinstituut voor Kust en Zee (RIKZ)

Extreme wave statistics

Confidence intervals

Sofia Caires

Report

February, 2007

WL | delft hydraulics

CLIENT:	Rijkswaterstaat, Rijksinstituut voor Kust en Zee
TITLE:	Extreme wave statistics: confidence intervals

ABSTRACT:

Large uncertainties are often found in estimates of extreme value parameters. These arise because estimation is based on a sample that covers a period much shorter than the return periods of interest and depends on the method used to obtain the parameter estimates and associated confidence intervals of the distributions generating the data. The purpose of this work is to assess and compare the performance of several methods of obtaining confidence intervals for return values of the generalized Pareto distribution (GPD) and generalized extreme value (GEV) distribution.

Typically, a method of obtaining a confidence interval for a parameter (e.g. a return value) is based on a method of obtaining parameter estimates of the distribution generating the data. To each method of estimating a parameter one can usually associate several different confidence intervals for that parameter; and although these are often equivalent in an asymptotic or theoretical sense (as the sample size is 'very' large), they can behave quite differently with the sample sizes one has to deal with in practice. In the case of the return values of the GEV distribution, the confidence intervals considered in the present work are all based on parameter estimates obtained by the method of Probability Weighted Moments (PWM). In the case of the return values of the GPD, we consider confidence intervals based on parameter estimates obtained by the method of PWM as well as confidence intervals based on Maximum Likelihood (ML) estimates.

In the case of ML estimators it is concluded that the best confidence intervals are those based on the profile likelihood method and in the case of PWM estimators that the best confidence intervals are those based on the adjusted percentile bootstrap method.

In view of the sample sizes and tail type typically found in the applications targeted in this report, it is recommended that the PWM method be used for the estimation of parameters and return values and the adjusted percentile bootstrap method be used for the construction of confidence intervals.

REFERENCES: RKZ-1697 SAP bestelnummer "4500041303 pos 20"

VER	AUTHOR		DATE	REMARKS	REVIEW		APPROVED BY	
1.0	S. Caires		20/12/2006		H.F.P. van den Boogaard		M. van Gent	
2.0	S. Caires		5/2/2007		H.F.P. van den Boogaard		W. Tilmans	

PROJECT NUMBER:	
KEYWORDS:	Extreme values, confidence intervals, significant wave height, SBW, hydraulic boundary conditions
NUMBER OF PAGES:	32
CONFIDENTIAL:	☐ YES ☒ NO
STATUS:	☐ PRELIMINARY ☐ DRAFT ☒ FINAL

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

In compliance with the Flood Defences Act of The Netherlands ("Wet op de Waterkering, 1996"), the primary coastal structures must be checked every five years (2001, 2006, 2011 etc.) for the required level of protection on the basis of the Hydraulic Boundary Conditions (HBC) and the Safety Assessment Regulation (VTV: *Voorschrift op Toetsen op Veiligheid*). These HBC must be derived anew every five years and established by the Minister of Transport, Public Works and Water Management.

At this moment, there is a degree of uncertainty concerning the quality of the current HBC, in particular those for the Waddenzee. This is because they were obtained from an inconsistent set of measurements and design values (WL, 2002), while for the rest of the Dutch coast (the closed Holland Coast and the Zeeland Delta) the SWAN wave transformation model has been applied (Rijkswaterstaat, 2001).

For 2011 and later the Dutch government plans to define the HBC for the Wadden Sea in the same way as for the rest of the Dutch Coast. In order to produce the best possible hydraulic boundary conditions for that region, and to assess the uncertainty that must be associated with such conditions, the Dutch Directorate for Public Works and Water Management (Rijkswaterstaat) is financing a large study led by WL | Delft Hydraulics. Specifically, Rijkswaterstaat requested WL | Delft Hydraulics to formulate, in the scope of the subproject "Boundary Conditions", which is part of the main project "Strength and Loading of Coastal Structures (SBW: *Sterkte en Belasting Waterkeringen*)", a Plan of Action (WL, 2006a). This plan establishes a strategy to answer the principle question of "How do we arrive at reliable Hydraulic Boundary Conditions for the Wadden Sea for 2011", and lists a sequence of associated activities, one of which is reported here.

1.2 Objectives of this study

One of the initial steps in defining HBC is the determination of offshore statistics (such as return value estimates) on extreme values. These are used by the HYDRA-K program and by the wave model SWAN (Booij et al., 1999).

In accordance with the principles of extreme value theory, statistics on extreme values can be obtained for example by sampling annual maxima data and fitting a Generalized Extreme Value (GEV) distribution to the data or by sampling Peaks Over Threshold (POT) data and fitting the Generalized Pareto Distribution (GPD) to the data. In WL (2006b) we have applied these methods to estimate return values of North Sea significant wave heights.

There are big uncertainties with estimates pertaining to extreme values. These arise because estimation is based on a sample that covers a period much shorter than the return periods of interest and depends on the method used to obtain the parameter estimates and associated confidence intervals of the distributions generating the data. The purpose of this work is to assess and compare the performance of several methods of obtaining confidence intervals for return values of the GPD and GEV distribution. A study on the effects of the uncertainty of the return value estimates on the design values of the dykes is to be reported soon within the same framework of this study.

Typically, a method of obtaining a confidence interval for a parameter (e.g. a return value) is based on a method of obtaining parameter estimates of the distribution generating the data. To each method of estimating a parameter one can usually associate several different competing confidence intervals for that parameter; and although these are often equivalent in an asymptotic or theoretical sense (as the sample size is 'very' large), they can behave quite differently with the sample sizes one has to deal with in practice. In the case of the return values of the GEV distribution, the confidence intervals considered in the present work are all based on parameter estimates obtained by the method of Probability Weighted Moments (PWM). In the case of the return values of the GPD, we consider confidence intervals based on parameter estimates obtained by the method of PWM as well as confidence intervals based on Maximum Likelihood (ML) estimates.

We may say that a confidence interval for a parameter is 'good' if its *coverage rate* is close to a specified level (95%, say). A coverage rate is the percentage of times that a confidence interval really contains the true parameter in (hypothetical) repetitions of the same sampling and estimation process. On the other hand, a good confidence interval is even better if it is *short*, since to know that a parameter is very likely to lie between 5 and 10 (say) is certainly better than to know that the same parameter is very likely to lie between 0 and 20. The 'shortness' of a confidence interval is best quantified by means of its *relative amplitude* (length or width) i.e. its *amplitude* (length or width)—the difference between the upper and lower limits of the interval—divided by the value of the parameter. In this work, our assessment and comparison of confidence intervals will be based mainly on coverage rates and partly on *average* relative amplitudes.

2 Extreme value theory

2.1 Asymptotic distributions

According to extreme value theory, the extreme values in a large sample have an approximate distribution that is independent of the distribution of each variable in the sample. More precisely, let us define $M_n = \max\{X_1,\ldots,X_n\}$, where $X_1,\ldots,X_n$, is a sequence of independent random variables having a common distribution function *F*. If there exist sequences $\{\sigma_n > 0\}$, $\{\mu_n\}$ of constants such that $\mathrm{P}\{(M_n - \mu_n)/\sigma_n \le z\} \to G(z)$ as $n \to \infty$, where *G* is a non-degenerate distribution function, then *G* must be a generalized extreme value (GEV) distribution. This distribution function is given by

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

where *z* take values in three different sets according to the sign of ξ: $z > \mu - \sigma/\xi$ if $\xi > 0$ (the domain of *z* has a lower limit), $z < \mu - \sigma/\xi$ if $\xi < 0$ (the domain of *z* has an upper limit), and $-\infty < z < \infty$ if $\xi = 0$. Here, the parameters μ, σ and ξ are called the location, scale, and shape parameters and satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. For $\xi = 0$ the GEV is the Gumbel distribution, for $\xi > 0$ it is the Fréchet distribution, and for $\xi < 0$ it is the Weibull distribution. For $\xi > 0$ the tail of the GEV is "heavier" (i.e., decreases more slowly) than the tail of the Gumbel distribution, and for $\xi < 0$ it is "lighter" (decreases more quickly and actually reaches 0) than that of the Gumbel distribution. The GEV distribution is said to have a type II tail for $\xi > 0$ and a type III tail for $\xi < 0$. The tail of the Gumbel distribution is called a type I tail. See the book of Coles (2001) for more information.

This result gives rise to the *annual maxima* (AM) method of modelling extremes, in which the GEV distribution is fitted to a sample of block maxima (in this case annual maxima, but the same could be done for e.g. biannual, monthly or even daily maxima).

One of the main applications of extreme value analysis is the estimation of the once per m year (1/m yr) return value. The 1/m yr return value based on the AM method/GEV distribution, z_m, is given by

$$z_m = \begin{cases} \mu - \dfrac{\sigma}{\xi}\left(1 - \left\{-\log\left(1-\dfrac{1}{\mathrm{m}}\right)\right\}^{-\xi}\right), & \text{for } \xi \neq 0 \\ \mu - \sigma\log\left\{-\log\left(1-\dfrac{1}{\mathrm{m}}\right)\right\}, & \text{for } \xi = 0. \end{cases} \tag{2.1}$$

The sample sizes of annual maxima data are usually small, so that model estimates, especially return values, have large uncertainties. This has motivated the development of more sophisticated methods that enable the modelling of more data than just block maxima. These methods are based on two well-known characterizations of extreme value distributions: one based on exceedances of a threshold, and the other based on the behaviour of the *r* largest, for small values of *r*, observations within a block.

We will not consider the *r*-largest approach because it is not often used in practice. Briefly, it consists of collecting the *r*-largest values per year (instead of merely the annual maxima) and fitting the *r*-largest distribution to the data (see, for instance, p. 68 of the book of S. Coles mentioned above). An example of the application of this method to estimate return values of significant wave height is given by Guedes Soares and Scotto (2004).

The approach based on the exceedances of a high threshold, hereafter referred to as the POT (Peaks Over Threshold) method, consists of fitting the generalized Pareto distribution (GPD) to the peaks of clustered excesses over a threshold, the excesses being the observations in a cluster minus the threshold, and calculating return values by taking into account the rate of occurrence of clusters (see Pickands, 1971 and 1975, and Davidson and Smith, 1990). Under very general conditions this procedure ensures that the data can have only three possible, albeit asymptotic, distributions (the three forms of the GPD) and, moreover, that observations belonging to different peak clusters are (approximately) independent. In the POT method, the peak excesses over a high threshold *u* of a time series are assumed to occur in time according to a Poisson process with rate λ_u and to be independently distributed with a GPD, whose distribution function is given by

$$F_u(y) = 1 - \left(1 + \xi \frac{y}{\tilde{\sigma}}\right)^{-1/\xi},$$

where $0 < y < \infty$, $\tilde{\sigma} > 0$ and $-\infty < \xi < \infty$. The two parameters of the GPD are called scale ($\tilde{\sigma}$) and shape (ξ) parameters. For $\xi = 0$ the GPD is the exponential distribution with mean $\tilde{\sigma}$, for $\xi > 0$ it is the Pareto distribution, and for $\xi < 0$ it is a special case of the beta distribution. As for the GEV, the GPD is said to have a type II tail for $\xi > 0$ and a type III tail for $\xi < 0$. The tail of the exponential distribution is a type I tail.

The 1/m yr return value based on a POT/GPD analysis, y_{m}, is given by

$$y_m = \begin{cases} u - \dfrac{\sigma}{\xi}\{(\lambda_{\mathrm{u}}\mathrm{m})^{\xi} - 1\}, & \text{for } \xi \neq 0 \\ u - \sigma \log(\lambda_{\mathrm{u}}\mathrm{m}), & \text{for } \xi = 0. \end{cases} \tag{2.2}$$

2.2 Parameter estimation

There are several methods available for the estimation of the parameters of extreme value distributions. Some of them, for instance the methods of moments and of probability weighted moments (PWM), give explicit expressions for the parameter estimates. The maximum likelihood (ML) method tends to be the preferred estimation method because it is quite general and more flexible than other methods (especially when the number of parameters is increased, as for instance when extending the extreme value analysis to account for non-stationarity), and because it is usually optimal in an asymptotic sense. However, in ordinary extreme value analyses like the ones we are concerned with in this report, the flexibility provided by the ML method is not necessary, and for the range of tails typically found with wave data (not too heavy-tailed distributions) and for small to moderate sample sizes the method of PWM performs better (see Hosking and Wallis, 1987, and Hosking et al., 1985). In this study we shall use the PWM and ML methods to estimate the parameters of the GPD, and the PWM method to estimate the parameters of the GEV distribution. Given the smaller sample sizes typically involved in the AM approach, it is clear that in the context of our report the flexibility of the ML method cannot compensate for its relatively poor performance referred to above.

2.3 Confidence intervals for the GEV and GPD parameters

Typically, approximate confidence intervals for the parameters of a given distribution are obtained, using a normal approximation, as functions of (estimates of) standard errors of parameter estimates, and therefore depend on the particular estimation method used. Such confidence intervals have upper and lower limits of the form *estimate* ± (*constant* × *standard error of estimate*), where the constant determines the degree of confidence; they will be described in more detail in the next section.

In the case of ML estimation, the standard error of the estimate may be obtained either from the expected information matrix or from the observed information matrix. Although asymptotically equivalent, the confidence intervals tend to be more reliable if the observed information is used (Hosking and Wallis, 1987). An alternative, and usually more accurate, method is the *profile likelihood* method (Coles, 2001, p. 57), which is based on the deviance function and, unlike the other two just mentioned, yields *asymmetric* confidence intervals. A confidence interval is said to be asymmetric if the distance between the associated estimate and the lower limit differs from the distance between the estimate and the upper limit; otherwise it is called symmetric.

Generally speaking, realistic confidence intervals for *return values* should be asymmetric, reflecting our intuition that the uncertainty in one direction (the true value of the parameter being *above* its estimate, say) is not the same as the uncertainty in the other direction. This is in contrast with the situation where one wants to find a confidence interval for *the mean* of a symmetric distribution (e.g. the normal).

Approximate, symmetric confidence intervals based on the method of PWM can also be obtained, via a normal approximation, in terms of standard errors of estimates (which are different from those of the ML method).

Besides methods based on the normal approximation and on expressions of standard errors of estimates, a promising class of methods for obtaining confidence intervals is afforded by the *bootstrap* methodology (Efron and Tibshirani, 1993). Some of the available bootstrap-based confidence intervals are also described in detail in the next section.

3 Construction of confidence intervals

In this report we shall present the results of a study about the uncertainty in return value estimates obtained using the ML and the PWM methods in terms of several associated confidence intervals. In our analysis we only consider the confidence intervals of 1/10000 yr return value estimates. The various methods of obtaining confidence intervals are described below and will be assessed and compared in Sections 4 and 5, via simulation, in terms of coverage rate and relative amplitude.

All computations were carried out in Matlab. The GPD and GEV parameter estimates were obtained using the WAFO statistical toolbox (see http://www.maths.lth.se/matstat/wafo/).

3.1 Asymptotic ML- and PWM-based intervals

Confidence intervals for the return value estimates can be obtained using the delta method (which is based on Cramér's theorem; see Ferguson, 1996, p. 45). Specifically, the asymptotic variance of the 1/m yr return value estimates can be estimated by

$$\mathrm{var}(\hat{x}_m) \approx d^T \sum d \ ,$$

where $\hat{x}_m$ equals the z_m of Eq. (2.1) in the case of the GEV distribution and the y_m of Eq. (2.2) in the case of the GPD, with the parameters replaced by their estimates, d is the vector of derivatives of $\hat{z}_m$ with respect to the estimated parameters—ξ and $\tilde{\sigma}$ in the case of POT/GPD estimates and ξ, σ, and μ in the case of AM/GEV estimates—and $\sum$ is the asymptotic covariance matrix of the parameter estimates, both evaluated at the estimates of the parameters.

The asymptotic covariance matrices of the GDP and GEV parameter estimates based on the ML or the PWM are available (see Hosking and Wallis, 1987, for the expressions and derivation of these matrices). In the case of ML estimates we shall use the observed information matrix in place of the expected information matrix since as mentioned earlier the former provides more reliable confidence intervals.

Approximate 95% confidence intervals are obtained by using the fact that the estimates are asymptotically normal centred at the parameter values with the appropriate covariance matrix; their upper and lower limits are calculated as $\hat{x}_m \pm 1.96\sqrt{d^T \sum d}$, where 1.96 is the quantile of probability 0.975 of the standard normal distribution.

We shall refer to the (symmetric) confidence intervals obtained in such way as *asymptotic ML-based intervals* and *asymptotic PWM-based intervals*.

3.2 Intervals based on the profile likelihood

The profile likelihood method is usually more accurate for the computation of confidence intervals based on maximum likelihood estimates (e.g. Coles, 2001), especially when the distribution of estimates is skewed This method is based on the likelihood ratio and is valid under certain regularity conditions (see Coles, 2001, and references therein for more details).

Suppose we have a model depending on a vector of parameters $\theta = (\phi, \varphi)$ with ϕ real, and that we want to find an approximate confidence interval of level 1-α for ϕ. Denote the log-likelihood function (the logarithm of the likelihood) by l and the maximum likelihood estimate by $\hat{\theta} = (\hat{\phi}, \hat{\varphi})$, so that $l(\hat{\theta}) = \max_\theta l(\theta)$ is the likelihood function at $\hat{\theta}$. (At this stage, $\hat{\theta}$ is assumed to have been calculated.)

For each ϕ, $l^*(\phi) = \max_\varphi l(\phi, \varphi)$ is the maximum of the log-likelihood regarded as a function of ϕ alone and φ kept fixed. Asymptotically, $2\left(l(\hat{\theta}) - l^*(\phi)\right)$ has an approximately χ_1^2 distribution, so that $2(l(\hat{\theta}) - l^*(\phi)) < c_{1,\alpha}$ with probability approximately of 1-α, where $c_{1,\alpha}$ is the quantile of probability $1-\alpha$ of the χ_1^2 distribution. (Note: The statistic $2\left(l(\hat{\theta}) - l^*(\phi)\right)$ is twice the likelihood ratio statistic, and the construction of this interval is based on the well-known Wilks theorem; see Coles, 2001). From this follows that an approximate confidence interval of level $1-\alpha$ for ϕ is given by the set

$$\left\{\phi : 2(l(\hat{\theta}) - l^*(\phi)) < c_{1,\alpha}\right\} = \left\{\phi : l^*(\phi) > l(\hat{\theta}) - \frac{c_{1,\alpha}}{2}\right\}.$$

Such confidence intervals need to be determined numerically, for example through small increments of the the value of ϕ and checking whether the inequality $l^*(\phi) > l(\hat{\theta}) - \frac{c_{1,\alpha}}{2}$ is satisfied.

We shall use this method in particular to obtain approximate confidence intervals for return values, which are defined in terms of parameters (e.g.~μ, σ and ξ in the case of the GEV model); the ϕ above is then regarded as the return value and φ as the vector with the remaining (two, in the case of the GEV model) free parameters. The confidence intervals obtained in this way are called *profile likelihood intervals*.

3.3 Bootstrap methods

Although in some cases the delta method can be used to find explicit expressions for the variances of the estimators, it often appears that such a task is extremely complicated or even impossible to carry out. In situations like this, resampling methods like the bootstrap offer a simple and reliable alternative for estimating standard errors of estimators. If the parameter estimates can be assumed approximately normal, such standard errors can be used in essentially the same way as explained in Subsection 3.1 to compute approximate confidence intervals for the parameter. More generally, the bootstrap method allows one to compute *percentile confidence intervals* which also work asymptotically but are generally asymmetric.

We will now briefly explain how the bootstrap can be used to estimate confidence intervals. The reader is referred to Efron and Tibshirani (1993) for an explanation of how and why the method works in each case. For an extensive description of the use of resampling techniques in frequency analysis see Hall et al. (2004).

3.3.1 Standard Bootstrap; percentile bootstrap intervals

In many situations, we have a random sample $x = \{x_i, i = 1,..., n\}$ of observations of some random variable or population X, and we wish to estimate a population parameter θ by an estimator $\hat{\theta} \equiv \hat{\theta}(x_1,..., x_n) \equiv \hat{\theta}(x)$ based on *x*. For instance, θ might be the median of the population (the quantile corresponding to the non-exceedence probability of 0.5) or the population mean (the expectation of X). If $\hat{\theta}$ has a simple expression and the distribution of X has simple mathematical properties, one can determine the variance of $\hat{\theta}$, or at least an approximation to it. In those cases where this is not possible, one can estimate (rather than determine exactly) $\mathrm{var}(\hat{\theta})$ (the variance of $\hat{\theta}$) by the bootstrap method.

The bootstrap method consists of creating bootstrap samples x*, each obtained by randomly sampling *n* times, with replacement, from the original sample *x*. Given B bootstrap samples, which we denote by $x_b^*,\ b = 1,\ldots, B$, we can calculate a set of estimates $\hat{\theta}_b^* = \hat{\theta}(x_b^*)$, each obtained in the same way $\hat{\theta}$ was obtained, but based on x_b^* in place of *x*. Then the bootstrap estimate of the standard error of the estimate $\hat{\theta}$ is given by

$$\hat{s}_B(\hat{\theta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b^* - \hat{\theta}^*)^2}\ \text{, with}\ \hat{\theta}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^*$$

The ideal bootstrap estimate would be $\hat{s}_\infty(\hat{\theta})$, but of course this is not possible to achieve; a limit on B must be stipulated. According to Efron and Tibshirani (1993, pp. 538 50–53), B = 200 bootstrap replications are usually enough for obtaining reasonable estimates of the standard error.

Estimates of the standard error of the estimate are useful for establishing confidence intervals or regions for the unknown parameters. The 95% confidence intervals can be obtained for a parameter θ assuming that the bootstrap distribution of the statistics is normal and calculating the upper and lower limits as $\hat{\theta} \pm 1.96 \cdot \hat{s}_B(\hat{\theta})$.

Instead of these intervals, however, we shall use 95% *percentile bootstrap intervals* (see Van den Boogaard and Diermanse, 2005), which are obtained by taking the quantile of probability 0.025 of the empirical distribution of the sample of bootstrap estimates $\hat{\theta}_b^*,\ b = 1,\ldots, B$, as the lower limit and the quantile of probability 0.975 as the upper limit. Because such confidence intervals are typically asymmetric, they tend to be more realistic than the ones just mentioned, and that is why we are going to use them here; see the motivation given by Hall et al. (2004), and Van den Boogaard and Diermanse (2005). More generally, a percentile bootstrap interval with (nominal) confidence level $1-2\alpha$ has its lower and upper limits given by the α and 1-α percentiles of the empirical distribution of the sample of bootstrap estimates; it may be denoted by $[\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)}]$.

3.3.2 Parametric bootstrap intervals

In the parametric bootstrap one assumes that the data follow a given distribution function $F(\cdot;\theta)$ which depends on the unknown parameter θ. Having estimated θ by $\hat{\theta}$, one then obtains B 'parametric' bootstrap samples, each of size n (like the original sample), by simulating from the estimated model $F(\cdot;\hat{\theta})$ rather than from the original data set $x=\{x_i, i=1,...,n\}$. For each parametric bootstrap sample an estimate $\hat{\theta}_b^*$ of θ is obtained, and as before one can use the sample $\hat{\theta}_1^*,\ldots,\hat{\theta}_B^*$ of bootstrap estimates to compute a percentile confidence interval for θ. We shall refer to the confidence intervals obtained in this way as *parametric* bootstrap intervals (for simplicity, we omit 'percentile').

3.3.3 Some problems with bootstrap confidence intervals

Tajvidi (2003) investigated the performance of several bootstrap methods for constructing confidence intervals for the parameters and quantiles of the GPD and concluded that none of the bootstrap methods gives satisfactory intervals for small sample sizes. In his study he also considered *bias-corrected* and *accelerated intervals* which will not be considered here, but which are discussed in detail in Efron and Tibshrirani (1993, Chapter 14).

Coles and Simiu (2003) state that "it is well known that bootstrap procedures are not consistent for extreme value problems—there is a tendency for the bootstrap sample to generate shorter tails than the true sample distribution". They propose an ad-hoc method to correct/adjust the bootstrap estimates which consists of applying a bias correction to the bootstrap parameter estimates. In their article, maximum likelihood estimates of the GPD are considered and a correction is applied to the bootstrap samples to ensure that the bootstrap means coincide with the ML estimates of the GPD parameters. That is, having estimated the parameters θ of a given distribution and obtained the bootstrap set of estimates $\hat{\theta}_b^*$, adjusted bootstrap estimates $\hat{\theta}_b^{*a}$, are obtained from

$$\hat{\theta}_b^{*a} = \hat{\theta}_b^* - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^* + \hat{\theta} = \hat{\theta}_b^* - \mathrm{mean}(\hat{\theta}^*) + \hat{\theta}. \qquad (3.1)$$

In addition to the percentile bootstrap and parametric bootstrap intervals introduced above, we shall also study in this work their *adjusted versions*, which are obtained in the same way except that the sample of bootstrap estimates is replaced by the sample of *adjusted bootstrap* estimates of Eq. (3.1); these will be referred to as *adjusted percentile bootstrap* and *adjusted parametric bootstrap intervals.*

4 Assessing and comparing confidence intervals for return values: the GPD case

4.1 Introduction

In this study we aim at assessing and comparing the performance of the methods described above for obtaining confidence intervals for return values of the GPD. Because the problem we have in mind is that of estimating offshore return values for the Wadden Sea area, we shall consider only choices of the GPD that are compatible with the characteristics of the offshore data from the Wadden Sea region. These characteristics have been determined by means of significant wave height measurements at Schiermonnikoog Noord (SON).

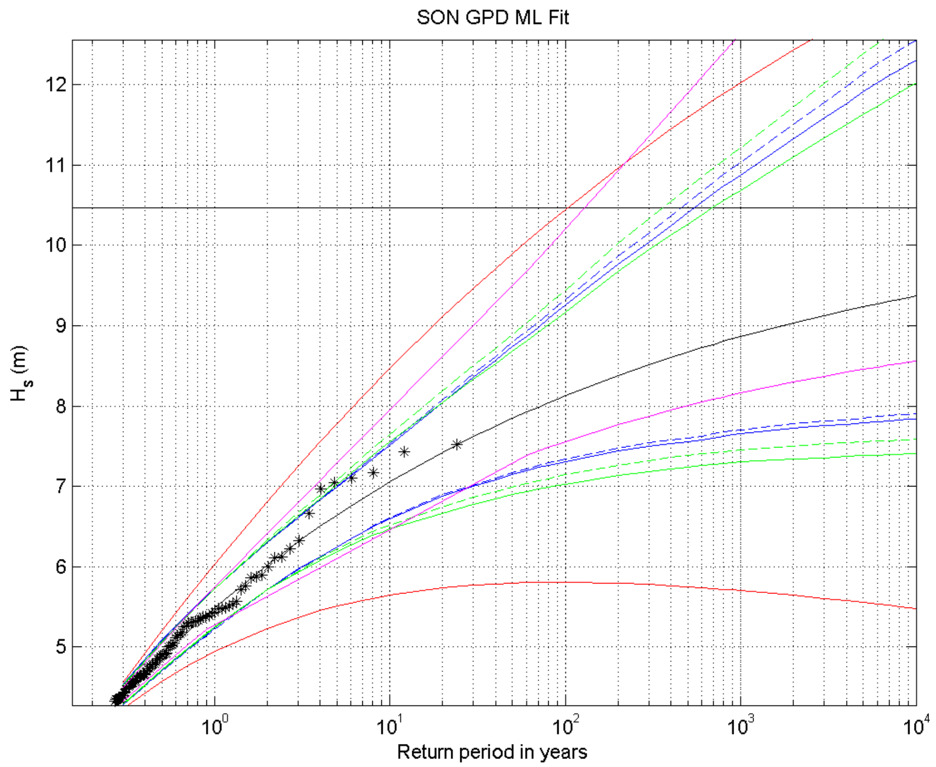

Figure 4.1 – Return value plot of the GPD fitted to the SON data obtained with the ML method (solid black line) with the various associated 95% confidence intervals. The GPD parameter estimates are $\xi = -0.16$ and $\tilde{\sigma} = 1.11$. The magenta lines represent the profile likelihood intervals. The solid blue lines represent the bootstrap intervals. The dashed blue lines represent the adjusted bootstrap intervals. The solid green lines represent the parametric bootstrap intervals. The dashed green lines were obtained using the adjusted parametric bootstrap intervals. The red lines represent the asymptotic (symmetric) ML-based intervals. The data are represented by the asterisks.

Figure 4.1 and Figure 4.2 present the return value estimates and the various associated confidence bands obtained by fitting the GPD with the ML and PWD methods, respectively, to SON POT data consisting of 157 peaks collected above a threshold of 3.69 m; for choice of this threshold and other details see WL (2006b). The 95% confidence intervals computed using the different methods enumerated above are also represented in the figures.

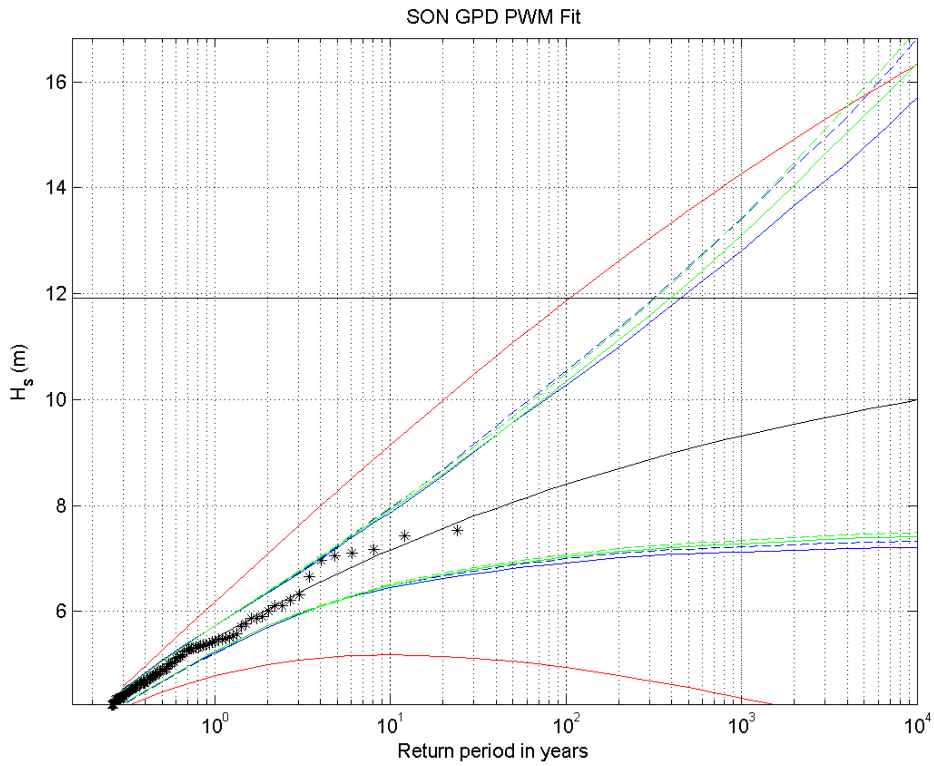

Figure 4.2 – Return value plot of the GPD fitted to the SON data obtained with the PWM method (solid black line) with the various associated 95% confidence intervals. The GPD parameter estimates are $\xi = -0.13$ and $\tilde{\sigma} = 1.08$. The solid blue lines represent the bootstrap intervals. The dashed blue lines represent the adjusted bootstrap intervals. The solid green lines represent the parametric bootstrap intervals. The dashed green lines were obtained using the adjusted parametric bootstrap intervals. The red lines represent the asymptotic PWM-based intervals. The data are represented by the asterisks.

The ML parameter estimates are $\xi = -0.16$ and $\tilde{\sigma} = 1.11$ and the PWM parameter estimates $\xi = -0.13$ and $\tilde{\sigma} = 1.08$. The difference between the estimates provided by the two methods appears to be small, but from visual inspection of the fits (Figure 4.1 and Figure 4.2) it looks like the PWM fit is slightly better.

The confidence bands associated with the ML estimates (Figure 4.1) suggest the following observations:

- The various bootstrap confidence bands are rather close to each other, the adjusted being slightly higher, which is consistent with the statement of Coles and Simu (2003) that "there is a tendency for bootstrap samples to generate shorter tails than the true sample distribution".
- Both the bootstrap and the profile likelihood are skewed (the lower bands are closer to the return value line than the upper bands), but the skewness of the profile likelihood intervals is greater.

- The asymptotic ML-based confidence bands are the only ones that are symmetrically placed around the estimate. Compared to the other confidence bands, the lower band seems to be too low.

Essentially the same comments apply to the confidence bands based on the PWM estimates of Figure 4.2, except that no profile likelihood intervals are available in this case, and the confidence bands of the PWM are wider, which is to be expected given the slightly higher estimate of the shape parameter ξ.

In the remainder of this section we shall assess the adequacy of the various confidence intervals by means of some simulation experiments and in terms of coverage rates. We take the parameter estimates obtained from the PWM GPD fit to the SON data, $\xi = -0.13$, $\tilde{\sigma} = 1.08$, as true, known parameters of our GPD model; the 1/10000 yr return value determined by this choice of parameters corresponds to our point estimate of the return value and is the parameter of interest for which the confidence intervals are to be assessed and compared. We shall restrict ourselves to 95% confidence intervals and consider sample sizes of *n*=50, 75, 100, 150, 200 and 250. In the case of bootstrap-based confidence intervals, we shall also consider different numbers of bootstrap samples, namely B=200, 1000 and 5000.

In each case we carry out 1000 simulations of a sample of size *n* from the GPD with $\xi = -0.13$, $\tilde{\sigma} = 1.08$, estimate its 1/10000 yr return value and compute the relevant 95% confidence interval. The adequacy of each type of confidence interval will be quantified in terms of

- **Coverage rate** (**CR**) **-** the percentage of times that the confidence interval includes the 1/10000 yr return value. The *nominal* coverage rate (i.e., the coverage rate one ideally wants to achieve) will be 95%.
- Coverage rate of the lower limit (**CRLL**) **-** the percentage of times the lower limit of the confidence interval is lower than or equal to the 1/10000 yr return value. The *nominal* CRLL will be 97.5%.
- Coverage rate of the upper limit (**CRUL**) **-** the percentage of times the upper limit of the confidence interval is greater than or equal to the 1/10000 yr return value. The *nominal* CRUL will be 97.5%.
- The **relative amplitude** of the confidence interval (**RA**) – the average ratio of the confidence interval's amplitude divided by the 1/10000 yr return value.

For a detailed description of coverage rates see Van den Boogaard and Hall (2004), who also give results on coverage rates of resampling-based confidence intervals in the case of the Weibull distribution.

4.2 Asymptotic ML- and PWM-based confidence intervals

We begin by analysing the characteristics of the asymptotic ML- and PWM-based 95% confidence intervals. Table 4.1 presents the coverage properties of the asymptotic ML-based intervals and Table 4.2 presents the same information for the asymptotic PWM-based intervals. The results suggest the following remarks:

- The coverage rate of the asymptotic ML-based intervals is below 95% for sample sizes up to 100, and this is due to the lower value of CRUL. The values of CRLL are for sample sizes higher than 50 conservative. For sample sizes above 150 the coverage rates are conservative.
- The coverage rate of the asymptotic PWM-based intervals is conservative, especially for the lower bounds.
- The relative amplitude of the asymptotic PWM-based intervals is higher than that of the asymptotic ML-based intervals.

n	CR (%)	CRLL(%)	CRUL(%)	RA
50	87	96	87	3.11
75	89	99	89	2.22
100	92	100	92	1.88
150	94	100	94	1.51
200	97	100	97	1.35
250	97	100	97	1.16

Table 4.1 – Coverage properties of the asymptotic ML-based 95% confidence intervals in the case of the GPD with parameters $\xi = -0.13$ and $\tilde{\sigma} = 1.08$.

n	CR (%)	CRLL(%)	CRUL(%)	RA
50	97	100	97	5.06
75	98	100	98	3.54
100	97	100	97	2.93
150	98	100	98	2.28
200	99	100	99	1.97
250	99	100	99	1.65

Table 4.2 – Coverage properties of the asymptotic PWM-based 95% confidence intervals in the case of the GPD with parameters $\xi = -0.13$ and $\tilde{\sigma} = 1.08$.

In general terms (symmetric) asymptotic confidence intervals have an unnecessarily high amplitude; they are conservative, especially in terms of the lower limit.

4.3 Profile likelihood intervals

Table 4.3 presents the coverage properties of the confidence intervals based on the profile likelihood coverage rate. Both CRLL and CRUL vary around 95%, and, especially for small sample sizes, the coverage rates are low.

n	CR (%)	CRLL(%)	CRUL(%)	RA
50	87	94	94	13.84
75	89	96	93	4.11
100	89	95	94	2.58
150	93	97	96	1.52
200	92	97	96	1.09
250	91	95	96	0.93

Table 4.3 – Coverage properties of the profile likelihood 95% confidence intervals in the case of the GPD with parameters $\xi = -0.13$ and $\tilde{\sigma} = 1.08$.

4.4 Bootstrap intervals

We shall now analyse the performance of the bootstrap-based confidence intervals. We begin by presenting the results on the intervals based on the ML estimates. Table 4.4 and Table 4.5 present the Coverage properties of the percentile bootstrap and adjusted percentile bootstrap confidence intervals based on the ML method, respectively. Table 4.6 and Table 4.7 give the statistics of the parametric bootstrap and adjusted parametric bootstrap confidence intervals based on the ML method, respectively.

The results presented in the tables suggest the following remarks:

- The coverage rate of the percentile bootstrap intervals is in all cases below 89%, and this is due to the lower value of CRUL. The values of CRLL are in almost all cases of 100%.
- The adjustment of the percentile bootstrap intervals produces an increase in RA and in the coverage rate of up to 6%, but the coverage rate remains in all cases below 91%. The improvements from the adjustment are mainly in the CRUL, with the values of CRLL remaining close to 100%.
- The coverage rate of the parametric bootstrap intervals is in all cases less than that of the percentile bootstrap intervals, but the coverage of the adjusted parametric bootstrap intervals is similar to that of the adjusted standard bootstrap intervals.
- For all the bootstrap-based methods and sample sizes considered, the quality of the coverage rate hardly depends on the size of the bootstrap sample size B. A bootstrap sample size of 1000 seems to be quite adequate.
- The profile likelihood confidence intervals are those with coverage rates, CRLL and CRUL closer to the nominal 97.5%.

We now analyse the coverage rate of the bootstrap confidence intervals based on the PWM estimates. Table 4.8 and Table 4.9 present the results for the percentile bootstrap and adjusted percentile bootstrap confidence intervals, respectively. Table 4.10 and Table 4.11 present the results for the parametric bootstrap and adjusted parametric bootstrap confidence intervals, respectively. The results suggest the following remarks:

- The performance of all the bootstrap intervals based on the PWM estimates is superior to that of the bootstrap intervals based on the ML estimates.
- The coverage rate of the percentile bootstrap intervals based on the PWM method varies between 92 and 95%. The corresponding values of CRLL are 98% and 99%, and those of CRUL are between 93% and 96%. The adjustment of the percentile bootstrap intervals results only in a small improvement of the coverage rate statistics.
- The coverage rate statistics of the parametric and of the adjusted parametric bootstrap intervals are very similar, although the latter have a higher relative amplitude.
- As was the case with the intervals based on the ML estimates, the deviation from the nominal 95% coverage rate hardly depends on the size B of the bootstrap samples.

The adjusted percentile bootstrap confidence intervals are those with coverage rates closer to the nominal 95%.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200	50	81	100	81	4.75
1000		80	100	80	4.60
5000		80	100	80	4.51
200	75	84	100	84	2.21
1000		82	100	82	2.03
5000		81	100	81	1.97
200	100	85	100	85	1.47
1000		84	100	84	1.58
5000		83	100	84	1.39
200	150	85	100	85	1.05
1000		87	100	87	1.06
5000		87	100	87	1.05
200	200	87	100	88	0.85
1000		88	100	89	0.88
5000		87	99	88	0.87
200	250	89	100	89	0.75
1000		89	99	89	0.75
5000		89	99	90	0.74

Table 4.4 – Coverage properties of the percentile bootstrap 95% confidence intervals based on the ML method in the case of the. GPD with parameters $\xi = -0.13$ and $\tilde{\sigma} = 1.08$.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200	50	85	100	85	5.62
1000		86	100	86	5.60
5000		84	100	85	5.51
200	75	88	99	88	2.58
1000		86	100	87	2.40
5000		86	100	86	2.34
200	100	88	100	89	1.67
1000		88	100	88	1.78
5000		87	100	87	1.58
200	150	89	100	89	1.15
1000		90	100	90	1.16
5000		90	99	91	1.14
200	200	90	100	90	0.91
1000		91	100	91	0.95
5000		89	99	90	0.93
200	250	90	99	91	0.80
1000		90	99	91	0.80
5000		91	99	92	0.78

Table 4.5 – Coverage properties of the adjusted percentile bootstrap 95% confidence interval based on the ML method in the case of the GPD with parameters $\xi = -0.13$ and $\tilde{\sigma} = 1.08$.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200	50	76	100	76	3.38
1000		75	100	75	3.81
5000		75	100	75	3.03
200	75	81	100	81	2.06
1000		80	100	80	1.85
5000		78	100	78	1.79
200	100	82	100	82	1.43
1000		82	100	82	1.46
5000		81	100	81	1.32
200	150	84	100	84	1.04
1000		84	100	84	1.05
5000		85	100	85	1.04
200	200	87	100	87	0.86
1000		87	100	87	0.89
5000		87	100	88	0.88
200	250	88	100	88	0.76
1000		88	100	88	0.76
5000		88	100	88	0.75

Table 4.6 – Coverage properties of the parametric bootstrap 95% confidence intervals based on the ML method in the case of the GPD with parameters $\xi = -0.13$ and $\tilde{\sigma} = 1.08$.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200	50	85	100	85	5.14
1000		85	100	85	5.75
5000		85	100	85	4.57
200	75	88	100	88	2.68
1000		86	100	86	2.42
5000		86	100	86	2.33
200	100	90	100	90	1.74
1000		87	100	87	1.77
5000		88	100	88	1.61
200	150	89	100	90	1.19
1000		90	100	90	1.20
5000		91	100	91	1.19
200	200	90	100	90	0.95
1000		92	100	92	0.98
5000		91	100	91	0.97
200	250	91	100	91	0.82
1000		91	100	91	0.82
5000		92	100	92	0.81

Table 4.7 – Coverage properties of the adjusted parametric bootstrap 95% confidence intervals in the case of the GPD with parameters $\xi = -0.13$ and $\tilde{\sigma} = 1.08$.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200		92	99	93	3.26
1000	50	92	99	94	3.16
5000		92	99	93	3.10
200		93	99	94	2.33
1000	75	93	99	94	2.35
5000		93	99	94	2.48
200		92	98	94	1.96
1000	100	94	99	95	1.97
5000		93	98	95	1.97
200		94	98	95	1.46
1000	150	94	98	96	1.53
5000		94	98	96	1.49
200		93	98	96	1.24
1000	200	94	99	95	1.22
5000		93	98	96	1.24
200		94	98	95	1.10
1000	250	95	99	96	1.09
5000		95	98	96	1.09

Table 4.8 – Coverage properties of the percentile bootstrap 95% confidence intervals based on the PWM method in the case of the GPD with parameters $\xi = -0.13$ and $\tilde{\sigma} = 1.08$.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200		94	98	96	4.53
1000	50	95	98	97	4.38
5000		95	99	96	4.30
200		95	98	97	2.90
1000	75	94	98	96	2.91
5000		95	98	97	3.08
200		93	98	96	2.29
1000	100	95	98	97	2.31
5000		95	97	97	2.30
200		94	98	97	1.62
1000	150	94	97	97	1.70
5000		95	98	97	1.66
200		95	97	97	1.34
1000	200	95	98	96	1.31
5000		94	97	97	1.34
200		95	98	97	1.17
1000	250	95	98	97	1.16
5000		95	98	97	1.16

Table 4.9 – Coverage properties of the adjusted percentile bootstrap 95% confidence intervals based on the PWM method in the case of the GPD with parameters $\xi = -0.13$ and $\tilde{\sigma} = 1.08$.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200	50	94	100	94	4.88
1000		95	100	95	4.56
5000		94	100	94	4.41
200	75	95	100	95	2.96
1000		94	100	94	3.00
5000		95	100	95	3.17
200	100	94	99	94	2.38
1000		95	100	95	2.39
5000		95	100	96	2.38
200	150	95	100	95	1.64
1000		95	99	96	1.72
5000		95	100	95	1.66
200	200	95	99	95	1.36
1000		95	100	95	1.32
5000		95	99	95	1.35
200	250	95	99	95	1.17
1000		95	100	96	1.16
5000		95	99	96	1.16

Table 4.10 – Coverage properties of the parametric bootstrap 95% confidence intervals based on the PWM method in the case of the GPD with $\xi = -0.13$ and $\tilde{\sigma} = 1.08$

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200	50	94	100	95	6.26
1000		95	100	95	5.71
5000		94	100	94	5.48
200	75	95	100	96	3.34
1000		96	100	96	3.43
5000		95	100	96	3.64
200	100	94	99	95	2.61
1000		95	100	96	2.62
5000		96	100	96	2.61
200	150	96	100	96	1.73
1000		96	99	96	1.82
5000		95	99	96	1.76
200	200	95	99	96	1.41
1000		94	99	96	1.37
5000		95	99	96	1.41
200	250	94	99	95	1.21
1000		95	100	96	1.20
5000		95	99	96	1.20

Table 4.11 – Coverage properties of the adjusted parametric bootstrap 95% confidence intervals based on the PWM method in the case of the GPD with $\xi = -0.13$ and $\tilde{\sigma} = 1.08$

5 Assessing and comparing confidence intervals for return values: the GEV case

5.1 Introduction

In Section 4 we have compared the performance of the methods described in Section 3 for obtaining confidence intervals for return values of the GPD. In this section we shall determine the characteristics of the considered different methods for determining the PWM –based confidence intervals for return values of the GEV distribution. We shall again only consider the choices of the GEV that are compatible with the characteristics of the offshore data from the Wadden Sea region, determined by means of measurements at the SON location.

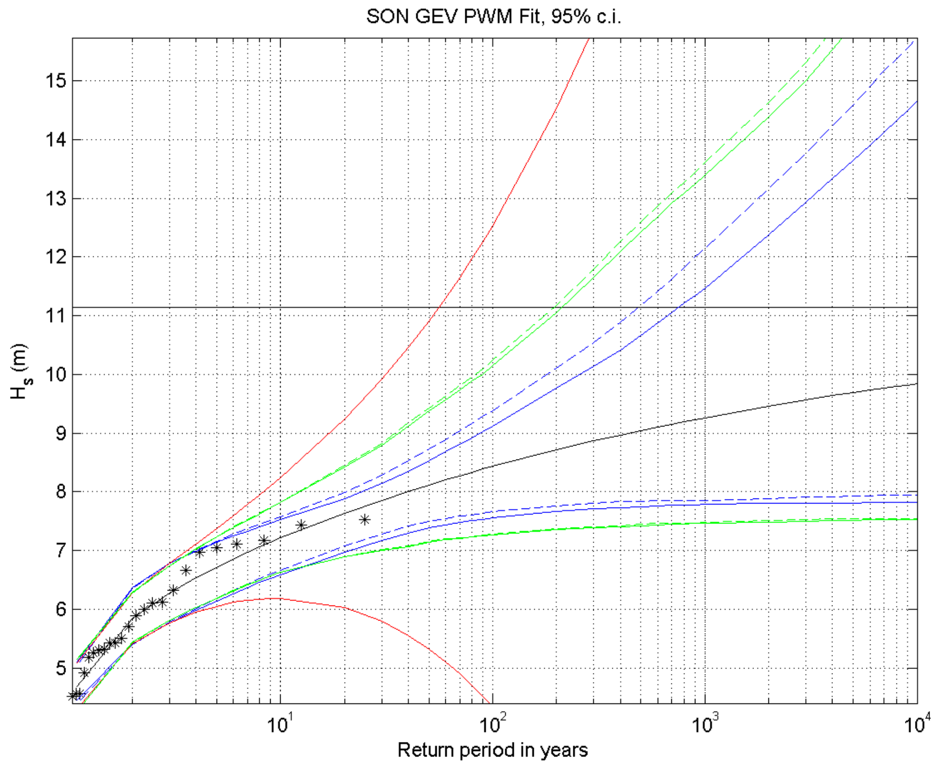

Figure 5.1 – Return value plot of the GEV distribution fitted to the SON data obtained with the PWM method (solid black line) with the various associated 95% confidence intervals. The GEV parameter estimates are $\xi = -0.16$, $\sigma = 0.89$ and $\mu = 5.53$. The solid blue lines represent the bootstrap intervals. The dashed blue lines represent the adjusted bootstrap intervals. The solid green lines represent the parametric bootstrap intervals. The dashed green lines were obtained using the adjusted parametric bootstrap intervals. The red lines represent the asymptotic PWM-based intervals. The data are represented by the asterisks.

Figure 5.1 present the return value estimates and the various associated 95% confidence bands obtained by fitting the GEV with the PWD method to SON AM data.

The PWM parameter estimates are $\xi = -0.16$, $\sigma = 0.89$ and $\mu = 5.53$. Analysing the ML associated confidence bands we see that:

- The different (adjusted) parametric bootstrap confidence bands are wider than the (adjusted) standard bootstrap confidence bands.
- The adjusted confidence bands are wider than the corresponding parametric or standard bootstrap confidence bands.
- The asymptotic PWM-based bands look rather conservative.

Again, we shall assess the adequacy of the various confidence intervals through simulation experiments as describer for the GPD case in Section 4 and again carrying out 1000 simulations in each case. We take the parameter estimates obtained from the PWM GEV fit to the SON AM data, $\xi = -0.16$, $\sigma = 0.89$ and $\mu = 5.53$, as true, known parameters of our GEV model; the 1/10000 yr return value determined by this choice of parameters corresponds to our point estimate of the return value and is the parameter of interest for which the confidence intervals are to be assessed and compared. We shall restrict ourselves to 95% confidence intervals and consider sample sizes of *n*=10, 25, 50, and 100. In the case of bootstrap-based confidence intervals, we shall again consider various choices of bootstrap sample sizes, namely B=200, 1000 and 5000.

5.2 Asymptotic PWM-based intervals

We start by analysing the characteristics of the asymptotic PWM-based 95% confidence intervals which are presented in Table 5.1. The results show that for sample sizes of 10 the values of CR and CRLL are quite poor. In all cases the values of CRUL are 100%, while for sample sizes above 25 all coverage rates are conservative.

n	CR (%)	CRLL(%)	CRUL(%)	RA
10	84	84	100	9.18
25	95	95	100	8.18
50	99	99	100	5.52
100	100	100	100	3.47

Table 5.1 – Coverage properties of the asymptotic PWM-based 95% confidence intervals in the case of the GEV distribution with parameters $\xi = -0.16$, $\sigma = 0.89$ and $\mu = 5.53$.

5.3 Bootstrap intervals

We shall now analyse the coverage rate of the bootstrap confidence intervals. Table 5.2 and Table 5.3 present the results for the percentile bootstrap and adjusted percentile bootstrap confidence intervals, respectively. Table 5.4 and Table 5.5 provide the results for the parametric bootstrap and adjusted parametric bootstrap confidence intervals, respectively. The results suggest the following remarks:

- In all cases the values of CRLL are greater than or equal to 98%. The highest values are found in the parametric bootstrap intervals.

- The standard bootstrap CR values vary between 93 and 99%. The most conservative values are found in the smaller sample sizes. The adjustment of the standard bootstrap intervals produces an increase of the RA and a slight improvement of the coverage rates.
- The parametric bootstrap CR values vary between 94 and 95%. The adjustment of the parametric bootstrap intervals produces an increase of the RA and only a very slight improvement of the coverage rates.
- The deviation from the nominal 95% coverage rate hardly depends on the size B of the bootstrap samples.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200	10	98	100	98	3.48
1000		99	100	99	3.56
5000		98	100	98	3.57
200	25	94	99	95	1.04
1000		94	100	94	1.03
5000		94	99	95	1.01
200	50	95	99	96	0.60
1000		94	99	95	0.62
5000		94	98	95	0.62
200	100	94	99	95	0.40
1000		93	98	95	0.40
5000		94	99	95	0.40

Table 5.2 – Coverage properties of the percentile bootstrap 95% confidence intervals based on the PWM method in the case of the GEV distribution with parameters $\xi = -0.16$, $\sigma = 0.89$ and $\mu = 5.53$.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200	10	98	100	98	8.33
1000		98	100	98	8.44
5000		97	100	97	8.36
200	25	94	98	96	1.29
1000		94	99	95	1.27
5000		94	98	96	1.25
200	50	95	99	96	0.65
1000		94	98	96	0.68
5000		94	98	96	0.68
200	100	94	98	96	0.41
1000		93	98	95	0.41
5000		94	98	95	0.41

Table 5.3 – Coverage properties of the adjusted percentile bootstrap 95% confidence intervals based on the PWM method in the case of the GEV distribution with parameters $\xi = -0.16$, $\sigma = 0.89$ and $\mu = 5.53$.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200		94	100	94	6.98
1000	10	95	100	95	7.22
5000		95	100	95	7.31
200		95	100	95	1.58
1000	25	94	100	94	1.52
5000		94	100	94	1.51
200		94	100	94	0.71
1000	50	95	100	95	0.76
5000		95	100	95	0.74
200		95	100	95	0.43
1000	100	94	100	94	0.43
5000		94	99	95	0.43

Table 5.4 – Coverage properties of the parametric bootstrap 95% confidence intervals based on the PWM method in the case of the GEV distribution with parameters $\xi = -0.16$, $\sigma = 0.89$ and $\mu = 5.53$.

b	n	CR (%)	CRLL(%)	CRUL(%)	RA
200		95	100	95	13.58
1000	10	95	100	95	14.27
5000		95	100	95	14.85
200		94	100	94	1.82
1000	25	94	100	94	1.73
5000		94	100	94	1.72
200		95	100	95	0.74
1000	50	95	100	95	0.79
5000		95	100	95	0.78
200		95	99	95	0.44
1000	100	94	100	94	0.44
5000		94	99	95	0.44

Table 5.5 – Coverage properties of the adjusted parametric bootstrap 95% confidence intervals based on the PWM method in the case of the GEV distribution with parameters $\xi = -0.16$, $\sigma = 0.89$ and $\mu = 5.53$.

6 Final remarks

6.1 Summary

In this report several methods of constructing confidence intervals for the return values of an extreme value analysis were considered. The properties and quality of these confidence intervals were assessed and compared in terms of coverage rates and, partly, of relative amplitudes by means of a simulation study.

The construction of a confidence interval for a return value of an extreme event involves a number of choices. The main steps involved are as follows:

1. The selection of a sample of observed extreme values. Given a time series of some physical quantity (here significant wave height) the Peaks over Threshold (POT) and Annual Maxima (AM) are the most commonly applied data selection methods, and were here considered as well.
2. The selection of an extreme value distribution as a statistical model for the data produced in the first step. According to theory the generalized Pareto distribution (GPD) should be selected when the sample is collected by means of the POT method, while in the case of the AM method a generalized extreme value (GEV) distribution should be chosen.
3. The selection of a method to estimate the unknown parameters of the distribution chosen in Step 2. On the basis of the estimated parameters (and thus the estimated distribution) the extreme values corresponding to one or more prescribed return period(s) can be estimated.
4. The selection of a method to quantify the uncertainty in the estimates of the distributions's parameters and in the associated return values. The uncertainty in the parameters and return values is usually quantified by a confidence interval of some significance level (here 95%).

In this study, the estimation methods mentioned in Step 3 were the maximum likelihood (ML) and the probability weighted moments (PWM) methods. In the case of the return values of the GEV distribution, the confidence intervals studied were all based on parameter estimates obtained by the method of PWM. In the case of the return values of the GPD, both confidence intervals based on parameter estimates obtained by the method of PWM and confidence intervals based on ML estimates were considered.

The methods of quantifying the uncertainty mentioned in Step 4 were based on asymptotic ML- and PWM-based, profile likelihood, percentile bootstrap, adjusted percentile bootstrap, parametric bootstrap and adjusted parametric bootstrap confidence intervals.

Only 95% confidence intervals of 1/10000 yr return values were considered. Because the type of application contemplated in this report is that of estimating offshore return values for the Wadden Sea area, only choices of the GPD and GEV distribution that are representative of the characteristics of the offshore data from the Wadden Sea region (as determined by means of measurements of significant wave height at Schiermonnikoog Noord, SON) were considered.

6.2 Conclusions

The main conclusions of the present study are as follows:

- In the case of the GPD, for return values based on parameter estimates obtained by the method of ML the profile likelihood method produces confidence intervals with coverage rates closer to the nominal 95%.
- When the PWM-method is applied to estimate the parameters, the adjusted percentile bootstrap method turns out to produce the best 95% confidence intervals from of the point of view of coverage rates. This holds with both the GPD and the GEV distribution. Moreover, such intervals have relatively low amplitudes or widths.
- For all the bootstrap-based methods and sample sizes considered, the quality of the coverage rate does not depend much on the bootstrap sample size B. A bootstrap sample size of 1000 seems to be quite adequate for most practical purposes.

6.3 Recommendations

In view of the sample sizes and tail type typically found in the applications targeted in this report, it is recommended that the computation of estimates and confidence intervals be made with the PWM method and the adjusted percentile bootstrap method, respectively.

A study on the effects of the uncertainty of the return value estimates on the design values of dykes will be reported soon within the same framework of this study.

References

Booij, N., R.C. Ris. and L.H. Holthuijsen, 1999: A third-generation wave model for coastal regions, Part 1: Model description and validation, J. Geophys. Res., **104 (C4)**, 7649– 7666.

Coles, S., 2001: An Introduction to the Statistical Modelling of Extreme Values. Springer Texts in Statistics, Springer-Verlag: London.

Coles, S., and E. Simiu, 2003: Estimating uncertainty in the extreme value analysis of data generated by a hurricane simulation model. J. Engrg. Mech., **129 (11)**, 1288-1294.

Davison, A. C., and R. L. Smith, 1990: Models for exceedances over high thresholds (with discussion). *J. Roy. Stat. Soc*., **52B**, 393–442.

Efron, B. and Tibshirani, R. J., 1993: *An Introduction to the Bootstrap*. Monographs on Statistics & Applied Probability 57**,** Chapman and Hall/CRC, 436p.

Ferguson, T. S., 1996: *A Course in Large Sample Theory*, Chapman and Hall, New York.

Guedes Soares, C. and M. G. Scotto, 2004: Application of the r largest-order statistics for long-term predictions of significant wave height, *Coastal Eng*., **21 (5-6)**, pp. 387-394.

Hall, M. J., Van den Boogaard, H. F .P., Fernando, R. C. and A .E. Mynett, 2004: The construction of confidence intervals for frequency analysis using resampling techniques. *Hydrol. Earth Sys. Sci*., **8(2)**, 235-246

Hosking, J. R. M., and J. R. Wallis, 1987: Parameter and quantile estimation for the Generalized Pareto Distribution. Technometrics, **29**, 339–349.

Hosking, J. R. M., J. R. Wallis, and E.F. Wood, 1985: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**, 251-261.

Pickands, J., 1971: The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability*, **8**, 745-756.

Pickands, J., 1975: Statistical inference using extreme order statistics. *Annals of Statistics*, **3**, 119-131.

Rijkswaterstaat, 2001: Hydraulische Randvoorwaarden 2001 voor het toetsen van primaire waterkeringen. (Hydraulic Boundary Conditions Book). Den Haag.

Tajvidi, N., 2003: Confidence Intervals and Accuracy Estimation for heavy-tailed Generalized Pareto Distributions. *Extremes*, **6**, 111-123.

Van den Boogaard, H. F. P. and F. L. M. Diermanse, 2005: Confidence intervals for extreme value analysis. *Proc. International Symposium on Stochastic Hydraulics,* Nijmegen, The Netherlands.

Van den Boogaard, H. F .P. and M .J. Hall, 2004: The construction of confidence intervals for frequency analysis using resampling techniques: a supplementary note. *Hydrol. Earth Sys. Sci.*, **8(6)**, 1174-1178.

WL, 2002: Kwaliteit randvoorwaardenboek en kwaliteit SWAN. WL|Delft Hydraulics Report H4061, Januari 2002.

WL, 2006a: SBW Plan of Action on the boundary conditions for the Waddenzee. WL | Delft Hydraulics Report H4750, April 2006.

WL, 2006b: Extreme wave statistics: Methodology and applications to North Sea wave data. WL | Delft Hydraulics Report H4803.30, October 2006.