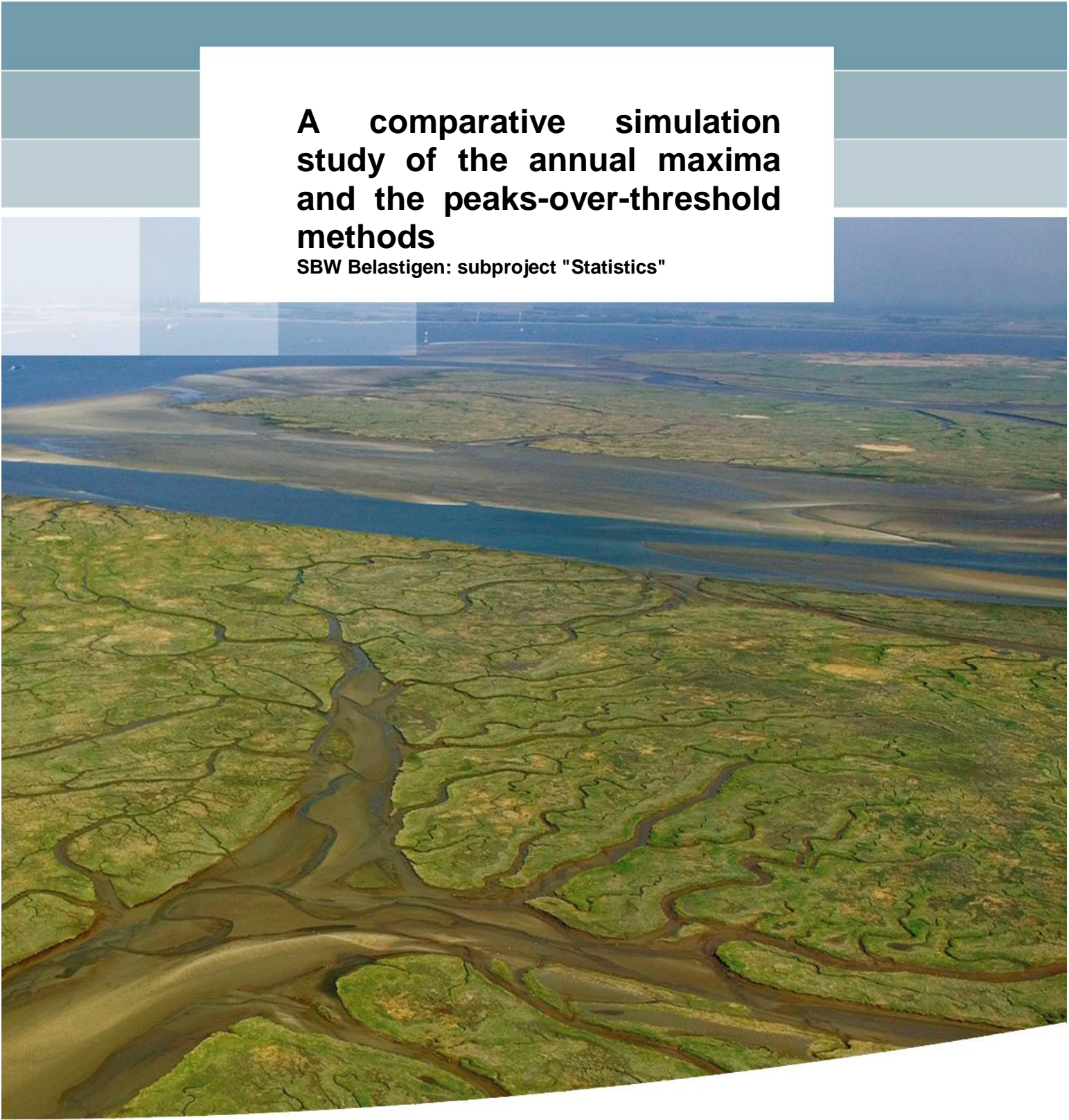


**A comparative simulation  
study of the annual maxima  
and the peaks-over-threshold  
methods**

SBW Belastigen: subproject "Statistics"



# **A comparative simulation study of the annual maxima and the peaks- over-threshold methods**

**SBW Belastigen: subproject "Statistics"**

Sofia Caires

**Title**

A comparative simulation study of the annual maxima and the peaks-over-threshold methods

**Client**

Rijkswaterstaat - Centre for Water Management

**Project**

1200264-002

**Pages**

58

**Keywords**

Extreme value analysis, sample size, non-stationarity, serial dependence, mean square error of estimates, environmental data

**Abstract** (see the Executive summary for a non-technical description)

In order to assess their relative merits in the context of the determination of Hydraulic Boundary Conditions (HBC), the Annual Maxima / Generalized Extreme Value distribution (AM/GEV) and Peaks over Threshold / Generalized Pareto Distribution (POT/GPD) approaches are compared in terms of their accuracy, as measured by mean square errors, in estimating exceedance probabilities on the basis of time series with various lengths and with characteristics, that mimic those of real time series, such as non-stationarity and serial dependence.

Two types of simulation studies were carried out. Both studies took into account the characteristics of the data currently available on the so-called basic variables. The first study focused on the finite-sample properties of the estimators of the GEV and GPD models based on independent and identically distributed data, both the Maximum Likelihood (ML) and Probability Weighted Moments (PWM) estimation methods having been considered. The second study focused on the finite-sample properties of the AM/GEV and the POT/GPD approaches applied to non-stationary and dependent data and using the method of PWM.

The conclusions of the first study were that with POT samples having an average of two or more observations per year, the GPD estimates are more accurate than the corresponding GEV estimates, and that with more than 200 years of data the accuracies of the two approaches are similar and rather good. Furthermore, it was concluded that with less than 50 years of data the method of PWM should, on the basis of its error characteristics and robustness, be preferred to the ML method, and that with longer data sets the two estimation methods have comparable accuracies.

In the second study, based on non-stationary and serially dependent data, it was concluded that the POT/GPD estimates of the shape parameter are more accurate than those of the AM/GEV approach with time series less than 100 years long. With 100-year long time series the performance of the two approaches is comparable, the accuracy of the AM/GEV approach being slightly greater with lighter tails and that of the POT/GPD approach being slightly greater with heavier tails. In terms of return value estimates (namely of the 4,000-yr and 10,000 return values), the POT/GPD approach is significantly more accurate. Only for time series of 200 years or more do the two approaches yield comparably small mean square errors. Still, even with 200-year long time series the relative root-mean square errors of the POT/GPD approach are about 2/3 of those of the AM/GEV approach whenever the underlying tail index exceeds -0.1. A noteworthy aspect of this second study is that the choice of the threshold in the POT/GPD approach has been chosen automatically; if visual inspection or a more sophisticated and theoretically grounded method is used to choose the threshold, the POT/GPD approach is expected to perform even better.

Based on the results of this study, we recommend that, irrespective of the variable of interest, the POT/GPD approach be used for the extreme value analyses of the data required for the computation of HBC. Furthermore, we recommend that the parameters of the GPD be estimated using the method of PWM.

**References**

Projectplan SBW Belastingen 2009, Deltares (A. van der Westhuysen)

Version	Date	Author	Initials	Review	Initials	Approval	Initials
1	Nov. 2009	Sofia Caires	SC	Douwe Dillingh Prof. Laurens de Haan		Marcel van Gent	
2	Dec. 2009	Sofia Calres	SC	Douwe Dillingh		Marcel van Gent	

State

final

**Title**

Uncertainties in the annual maxima and the peak-over-threshold approaches

## Executive summary

### General

According to the Dutch Water Defence Act (“Wet op de Waterkering, 1996”) the strength of the Dutch primary water defences must be checked every five years (1996, 2001, 2006, 2011, etc.) for the required level of protection from loads with return periods varying from 250 to 10,000 years, depending on the area protected by the water defence. These loads are determined on the basis of Hydraulic Boundary Conditions (HBC) and must also be derived anew and approved by the Dutch Ministry of Transport, Public Works and Water Management (VenW) every five years.

With the aim of filling knowledge gaps in the determination of the strengths and loads of the water defences, Rijkswaterstaat - Centre for Water Management in The Netherlands is funding the long-term research project “Strengths and loads of water defences” (in Dutch: “Sterkte en Belastingen Waterkeringen”; in short: SBW).

The SBW program presently comprises nine projects, of which seven are related to the strengths and two to the loads of water defences: SBW-Waddenzee and SBW-Belastingen.

The SBW-Belastingen project is divided into a number of sub-projects. One of these sub-projects is the Statistical Models sub-project. The HBC used to assess the primary water defences have return periods that are several orders of magnitude larger than the periods of available environmental data used in the determination of the loads. One of the main challenges of the statistical modelling used in the determination of HBC is thus to decide how the very low probability of extreme (not yet observed) events should be determined from the limited information available, which varies from a few decades to a century of data. The main goal of the Statistical Models sub-project is therefore to evaluate and possibly improve the methods and approaches used in the modelling of the extremes of the relevant *basic variables*. These basic variables are the wind velocities, still water levels (at sea), lake water levels, wave parameters at sea, and river discharges. The present study is part of the program for 2009 of the sub-project Statistical Models of SBW-Belastingen.

### Statement of the problem

Two approaches are usually considered for the estimation of extreme values:

- the Annual Maxima / Generalized Extreme Value distribution, in short the AM/GEV approach, and
- the Peaks over Threshold / Generalized Pareto Distribution, in short the POT/GPD approach.

In general, and especially in view of the typically short (covering a few decades) time series available on the basic variables, the POT/GPD approach is preferable because it makes use of more data, thanks to the possibility of choosing a *threshold* that is exceeded on average by more than one peak per year, than the AM/GEV approach, in which only the yearly maxima are considered. The POT/GPD approach is therefore expected to yield more accurate estimates of parameters and exceedance probabilities as it is based on more observations (namely storm peaks).

## **Title**

Uncertainties in the annual maxima and the peak-over-threshold approaches

Although the choice of the threshold in the POT/GPD approach represents an opportunity to increase the number of observations used to estimate the parameters and hence to reduce the errors of the estimates, it is often seen by some practitioners as a problem and source of subjectivity. Consequently, because of its comparative straightforwardness the AM/GEV approach often gets the better of the POT/GEV approach and the latter is not even considered.

## **Aim of the study**

In order to exhibit the relative merits of the AM/GEV and POT/GPD approaches in the context of the determination of HBC, in this study we compare them in terms of their accuracy in estimating exceedance probabilities on the basis of time series with various lengths and, most importantly, with characteristics that mimic those of real time series. The aim of this study is thus to provide a global assessment and comparison of the two approaches on the basis of small to large time series of simulated data possessing realistic features such as non-stationarity and serial dependence. It is hoped that the results of this study will help practitioners involved in the determination of HBC to choose between the two methods, for instance on the basis of the length of the time series and of the expected underlying shape parameter of the data available to them.

## **Approach**

In order to compare the two approaches two types of simulation studies are carried out. Both simulation studies take into account the characteristics of the time series of the basic variables currently available. The first study focuses on the finite-sample properties of the Maximum Likelihood (ML) and Probability Weighted Moments (PWM) estimators of the GEV and the GPD models, based on independent and identically distributed observations. The second study focuses on the finite-sample properties of the AM/GEV and the POT/GPD approaches applied to non-stationary and serially dependent observations.

## **Conclusions**

The conclusions of the first study are that with POT samples having an average of two or more observations per year, the GPD estimates are more accurate than the corresponding GEV estimates, and that with more than 200 years of data the accuracies of the two approaches are similar and rather good. Furthermore, it is concluded that with less than 50 years of data the method of PWM should, on the basis of its error characteristics and robustness, be preferred to the ML method, and that with longer data sets the two estimation methods have comparable accuracies.

In the second study, based on non-stationary and serially dependent data, it is concluded that the POT/GPD estimates of the shape parameter are more accurate than those of the AM/GEV approach with time series less than 100 years long. With 100-year long time series the performance of the two approaches is comparable, the accuracy of the AM/GEV approach being slightly greater with lighter tails (yielding less extreme data with an upper bound) and that of the POT/GPD approach being slightly greater with heavier tails (yielding more extreme data). In terms of return value estimates (namely of the 4,000-yr and 10,000 return values), the POT/GPD approach is significantly more accurate. Only for time series of 200 years or more do the two approaches yield comparably small mean square errors. Still, even with 200-year long time series the relative root-mean square errors of the POT/GPD approach are about 2/3 of those of the AM/GEV approach whenever the underlying tail is medium light to heavy (more precisely, when the tail index exceeds -0.1). A noteworthy

**Title**

Uncertainties in the annual maxima and the peak-over-threshold approaches

aspect of this second study is that the choice of the threshold in the POT/GPD approach is chosen in an automatic way; if visual inspection or a more sophisticated and theoretically grounded method is used to choose the threshold, the POT/GPD approach is expected to perform even better.

**Recommendations**

Based on the results of this study, we recommend that, irrespective of the variable of interest, the POT/GPD approach be used for the extreme value analyses of the data required for the computation of HBC. Furthermore, we recommend that the parameters of the GPD be estimated using the method of PWM.

## Contents

<b>List of Tables</b>	<b>i</b>
<b>List of Figures</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Framework	1
1.2 Background and motivation	1
1.3 Objective	2
1.4 Approach	2
<b>2 Extreme value analysis</b>	<b>3</b>
2.1 Introduction	3
2.2 Extreme value theory	3
2.2.1 Block maxima	3
2.2.2 Peaks-over-threshold method	4
2.2.3 Domains of attraction	7
2.3 Estimation	7
2.4 Serial dependence	8
<b>3 Setup of the study</b>	<b>9</b>
3.1 Characteristics of the real data	9
3.2 Previous simulation studies	10
3.3 The present simulation study	11
<b>4 Finite-sample properties of the ML and PWM estimators of the GEV and GPD models</b>	<b>15</b>
4.1 Introduction	15
4.2 Analysis of the results	16
4.3 Effect of the scale and location parameters on the relative errors in the case of the GEV	22
<b>5 Finite-sample properties of the two methods with non-stationary and dependent data</b>	<b>27</b>
5.1 Introduction	27
5.2 Analysis of the results based on time series with marginal GEV distributions	27
5.2.1 Effect of the number of simulations on the error statistics	31
5.2.2 Comparison between the performance of the estimators with i.i.d. and non-i.i.d. data	32
5.3 Analysis of the results based on time series with marginal Gamma or Beta distributions	32
<b>6 Conclusions</b>	<b>36</b>
<b>References</b>	<b>37</b>

## Appendices

<b>A Further details on the simulation of the time series</b>	<b>A-1</b>
A.1 GEV distribution	A-1
A.2 Gamma distribution	A-1
A.3 Beta distribution of the first kind	A-2
A.4 Beta distribution function of the second kind with a scale parameter	A-3
<b>B Report review</b>	<b>B-1</b>
B.1 Introduction	B-1
B.2 Review report	B-1
B.3 Reply to the review	B-5



## List of Tables

Table 3.1	Non-zero coefficients of the AR(19) model obtained by Guedes Soares et al. (1996) from three-hourly significant wave height data from the North Atlantic. 12
Table 3.2	Average monthly means (M) and standard deviations (S) of the SON data from 1970 to 2002. .... 13
Table 4.1	Relative RMSE and bias of the ML and PWM estimators of the 4,000-yr return values. *Values in excess of 1,000. .... 18
Table 4.2	Relative RMSE and bias of the ML and PWM estimators of the 10,000-yr return values. *Values in excess of 1,000. .... 19
Table 4.3	RMSE and bias of the ML and PWM estimators of $\xi$ . .... 20
Table 4.4	Relative RMSE and bias of the ML and PWM estimators of $\sigma$ and $\tilde{\sigma}$ ..... 21
Table 4.5	Failure rate (%) of the ML method. .... 21
Table 4.6	Parameters of the GEV distribution of the yearly maxima of three-hourly i.i.d. GEV distributed variables with scale parameter 1, location parameter 0, and $\xi$ as given in the first row of the table. .... 22
Table 4.7	Relative RMSE and bias of the ML and PWM estimators of the 4,000-yr return value of the GEV. *Values in excess of 1,000. .... 23
Table 4.8	Relative RMSE and bias of the ML and PWM estimators of the 10,000-yr return value of the GEV. *Values in excess of 1,000. .... 24
Table 4.9	RMSE and bias of the ML and PWM estimators of $\xi$ in the GEV model. .... 25
Table 4.10	Relative RMSE and bias of the ML and PWM estimators of $\sigma$ in the GEV model. .... 26
Table 4.11	Failure rate (%) of the ML estimation in the GEV model. .... 26
Table 5.1	Estimates of the parameters of the GEV model fitted to 100,000 year maxima. 28
Table 5.2	RMSE and bias of the PWM parameter estimates in the AM/GEV and POT/GPD approaches, computed from 100 simulations ..... 30
Table 5.3	RMSE and bias of the PWM parameter estimators of the AM/GEV and POT/GPD approaches computed from 100 and 1,000 simulations. Results only for the type I tail. .... 31
Table 5.4	Error statistics of the PWM estimators of the GEV distribution with i.i.d. data when the parameters correspond to those of the non-i.i.d. data given in Table 5.1. Based on 100,000 simulations. .... 32
Table 5.5	Estimates of the parameters of the GEV distribution fitted to 100,000 yearly maxima of time series with Gamma ( $\xi=0$ ) Beta ( $\xi=-0.1$ ) and Beta of the second kind ( $\xi=0.1$ ) marginal distributions. .... 33

Table 5.6	RMSE and bias of the PWM estimates of the 4,000- and 10,000-yr return value and of $\xi$ in the AM/GEV and POT/GPD approaches applied to non-stationary and dependent time series with Gamma, Beta and Beta of the second kind marginal distributions. Computed from 100 simulations.....	34
Table 5.7	RMSE and bias of the ML estimates of the 4,000- and 10,000-yr return value and of $\xi$ in the AM/GEV and POT/GPD approaches applied to non-stationary and dependent time series with Gamma, Beta and Beta of the second kind marginal distributions. Computed from 100 simulations. * No estimate available due to a high failure rate.....	35

## List of Figures

Figure 5.1	Illustration of the variation of the estimates of the shape parameter, of $\sigma^*$ and of the 10,000 return value with the threshold in the POT/GPD approach. The vertical line indicates the threshold chosen. The example shown is from a simulation with $n_y=200$ and $\xi=0$ . .....	29
Figure 5.2	Illustration of the variation of the tail index, $\sigma^*$ and 10,000 return value estimates with the threshold in the POT/GPD approach. The vertical line indicates the automatically chosen threshold. The example shown is from a simulation with $n_y=100$ and $\xi=-0.1$ . .....	33

# 1 Introduction

## 1.1 Framework

According to the Dutch Water Defence Act (“Wet op de Waterkering, 1996”) the strength of the Dutch primary water defences must be assessed every five years (1996, 2001, 2006, 2011, etc.) for the required level of protection, which, depending on the area protected by the water defence, may vary from 250 to 10,000 year loads. These loads are determined on the basis of Hydraulic Boundary Conditions (HBC) and must also be derived anew and approved by the Dutch Ministry of Transport, Public Works and Water Management (VenW) every five years. The project WTI (Wettelijk Toets Instrumentarium, or, in English, Legal Assessment Instruments), which is based on the HBC and on the Safety Assessment Regulation (VTV: Voorschrift op Toetsen op Veiligheid), plays the crucial role in the assessment of the primary water defences.

With the aim of filling knowledge gaps in the determination of the strengths and loads of the water defences, Rijkswaterstaat - Centre for Water Management (to be called RWS-CWM in the remainder of this report; in Dutch: “Waterdienst”) in The Netherlands is funding the long-term research project “Strengths and loads of water defences” (in Dutch: “Sterkte en Belastingen Waterkeringen”; in short: SBW). The SBW products are considered in each assessment round by the WTI project.

The SBW program presently comprises nine projects, of which seven are related to the strengths and two to the loads of water defences: SBW-Waddenzee and SBW-Belastingen<sup>1</sup>. They aim at determining the quality of models and methods and at improving them where needed, in order that from 2011 onwards more accurate HBC be determined.

The SBW-Belastingen project is divided into a number of sub-projects. One of these sub-projects is the Statistical Models sub-project. The HBC used to assess the primary water defences have return periods that are several orders of magnitude larger than the periods of available environmental data used in the determination of the loads. One of the main challenges of the statistical modelling used in the determination of HBC is thus to decide how the very low probability of extreme (not yet observed) events should be determined from the limited information available, which varies from a few decades to a century of data. The main goal of the Statistical Models sub-project is therefore to evaluate and possibly improve the methods and approaches used in the modelling of the extremes of the relevant basic variables. These basic variables are the wind velocities, still water levels (at sea), lake water levels, wave parameters at sea, and river discharges. The present study is part of the program for 2009 of the sub-project Statistical Models of SBW-Belastingen (De Waal et al., 2009).

## 1.2 Background and motivation

Two approaches are usually considered for the estimation of extreme values:

- 1 the annual maxima/Generalized Extreme Value distribution (AM/GEV) approach, and
- 2 the Peaks over Threshold/Generalized Pareto Distribution (POT/GPD) approach.

---

1. *Dutch for loads.*

On the grounds of theoretical arguments, the POT/GPD method is expected to generally perform better than the AM/GEV method. While the latter typically uses a single observation per year (the annual maximum) the former makes use of several observations per year, namely of *storm peaks* which exceed a certain *threshold* determined by the user, and this more intensive use of the data usually proves to be advantageous for the analysis of the comparatively short (comprising a couple of decades) time series of measurements available. This is so because the more observations are used to estimate the parameters of a probability model, the more accurate the estimates of the parameters, and hence the estimates of quantities determined by the parameters (such as exceedance probabilities), tend to be. As explained by de Haan and Zhou (2009), by “choosing the threshold in an intelligent way, either by visual inspection or by using a theoretically justified threshold selection procedure, it is possible to improve the tail estimation substantially. The AM/GEV approach has no room for this improvement since the year is usually the only realistic time period one can consider”. However, although the choice of the threshold should be seen as an opportunity, it is often seen by some practitioners as a problem and a source of subjectivity, and consequently the comparative straightforwardness of the AM/GEV approach often gets the better of the POT/GPD approach in applications.

### 1.3 Objective

In order to exhibit the relative merits of the AM/GEV and POT/GPD approaches in the context of the determination of HBC it is therefore of interest to compare them in terms of their accuracy in estimating exceedance probabilities on the basis of time series with various lengths and, most importantly, with characteristics that mimic those of real time series. The objective of this study is to provide a global assessment and comparison of the two approaches on the basis of small to large time series of simulated data possessing realistic features such as non-stationarity and serial dependence. It is hoped that the results of this study will help practitioners involved in the determination of HBC to choose between the two methods, for instance on the basis of the length of the time series available to them.

### 1.4 Approach

In order to compare the two approaches two types of simulation studies were carried out. Both simulation studies took into account the characteristics of the basic variable data currently available. The first studied focused on the finite-sample properties of the GEV and the GPD estimators. The second study on the finite-sample properties of the AM/GEV and the POT/GPD approaches.

## 2 Extreme value analysis

### 2.1 Introduction

This section introduces briefly the principles of extreme value theory and describes the methods used in extreme value analysis (EVA) as well as other statistical methods.

### 2.2 Extreme value theory

#### 2.2.1 Block maxima

In order to explain the basic ideas of extreme value theory, let us write  $M_n = \max\{X_1, \dots, X_n\}$ , where  $X_1, X_2, \dots$  is a sequence of independent and identically distributed (i.i.d) random variables with distribution function  $F$ . In its simplest form, the *extremal types theorem* states the following: If there exist sequences of constants  $\{\sigma_n > 0\}$  and  $\{\mu_n\}$  such that  $P\{\sigma_n M_n + \mu_n \leq z\} \rightarrow G(z)$  as  $n \rightarrow \infty$ , where  $G$  is a *non-degenerate distribution function*<sup>2</sup>, then  $G$  must be a Generalized Extreme Value (GEV) distribution, which is given by

$$G(z) = \begin{cases} \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{for } \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{z - \mu}{\sigma}\right)\right]\right\} & \text{for } \xi = 0, \end{cases} \quad (2.1)$$

where  $z$  takes values in three different sets according to the sign of the shape parameter  $\xi$ :  $z > \mu - \sigma/\xi$  if  $\xi > 0$  (the domain of  $z$  has a lower bound),  $z < \mu - \sigma/\xi$  if  $\xi < 0$  (the domain of  $z$  has an upper bound), and  $-\infty < z < \infty$  if  $\xi = 0$ .

In other words, if the distribution function of (a normalization of) the maximum value in a random sample of size  $n$  converges to a distribution function as  $n$  tends to infinity, then that distribution function must be a GEV distribution. Moreover, this and other results of extreme value theory *hold true even under general dependence and non-stationary conditions* (see Chapter 5 of Coles, 2001), i.e., even when the variables  $X_1, X_2, \dots$  are not independent or/nor identically distributed.

In Eq. (2.1) the parameters  $\mu$ ,  $\sigma$  and  $\xi$  are called the location, scale and shape parameters, and they 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 *(reverse) 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 is said to have a *type II tail* when the shape parameter is positive ( $\xi > 0$ ) and a *type III tail* when the shape parameter is

2. A distribution function is said to be degenerate if it allocates probability 1 to a single point.

negative ( $\xi < 0$ , the domain of  $z$  has an upper bound)<sup>3</sup>. The tail of the Gumbel distribution ( $\xi = 0$ ) is called a *type I or exponential tail*.

Since the distribution function  $F$  of the variables  $X_1, X_2, \dots$  in the extremal types theorem is associated with a unique (i.e. independent of the particular sequences  $\{\sigma_n > 0\}$  and  $\{\mu_n\}$ ) value of  $\xi$ , the parameter  $\xi$  is called the *tail index* of  $F$ .

The extremal types theorem gives rise to the *annual maxima* (AM) method of modelling extremes, in which the GEV distribution is fitted to a sample of block maxima (e.g. to annual maxima, though biannual, monthly or even daily maxima can of course be used as well).

One of the main applications of extreme value theory is the estimation of the *once per m year* (*m-yr*) *return value*, the value which is exceeded on average once every  $m$  years. The  $m$ -yr return value (for  $m > 1$ ) based on the AM method/GEV distribution,  $z_m$ , is given by<sup>4</sup>

$$z_m = \begin{cases} \mu - \frac{\sigma}{\xi} \left( 1 - \left\{ -\ln \left( 1 - \frac{1}{m} \right) \right\}^{-\xi} \right), & \text{for } \xi \neq 0 \\ \mu - \sigma \ln \left\{ -\ln \left( 1 - \frac{1}{m} \right) \right\}, & \text{for } \xi = 0. \end{cases} \quad (2.2)$$

Since the time series available typically cover only a few decades, the sample sizes of annual maxima data are usually small. Consequently, the estimates of the parameters of the GEV distribution, and hence the estimates of the return values, have large variances (uncertainties). This has motivated the development of a more sophisticated method based on exceedances of a threshold described next.

### 2.2.2 Peaks-over-threshold method

The approach based on the exceedances of a high threshold, hereafter referred to as the *Peaks-over-Threshold* (POT) 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 given below) 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  $\lambda_u$  and to be independently distributed with a GPD, whose distribution function is given by

- 
3. *Note that some articles (e.g. Hosking and Wallis, 1987) use another convention for the sign of the shape parameter: a negative shape parameter in those references corresponds to a type II distribution.*
  4. *In this report the natural logarithm of  $x$  is written as  $\ln(x)$ .*

$$F_u(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\tilde{\sigma}}\right)^{-1/\xi}, & \text{for } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), & \text{for } \xi = 0, \end{cases} \quad (2.3)$$

where,  $y > 0$ ,  $\tilde{\sigma} > 0$  and  $(1 + \xi(y/\tilde{\sigma})) > 0$ . The two parameters of the GPD are called the scale ( $\tilde{\sigma}$ ) and shape ( $\xi$ ) parameters. When  $\xi = 0$  the GPD is said to have a type I tail and amounts to the exponential distribution with mean  $\tilde{\sigma}$ ; when  $\xi > 0$  it has a type II tail and it is the Pareto distribution; and when  $\xi < 0$  it has a type III tail and it is a special case of the beta distribution. If  $\xi < 0$ , just as with the GEV distribution, the support of the GPD has an upper bound,  $-\tilde{\sigma}/\xi$ , which is called the *upper end-point* of the GPD and is to be thought of as the upper-limit of the excesses, the upper limit of the variable of interest being then  $u - \tilde{\sigma}/\xi$ .

The  $m$ -yr return value based on a POT/GPD analysis,  $z_m$ , is given by

$$z_m = \begin{cases} u + \frac{\tilde{\sigma}}{\xi} \{ (\lambda_u m)^\xi - 1 \}, & \text{for } \xi \neq 0 \\ u + \tilde{\sigma} \ln(\lambda_u m), & \text{for } \xi = 0. \end{cases} \quad (2.4)$$

Note that this expression is obtained from Eq. (2.3) by solving  $(1 - F_u(y)) = \frac{1}{\lambda_u m}$  for  $y$  and then adding the threshold  $u$  to the result.

Just as block maxima have the GEV as their approximate distribution, the threshold excesses have a corresponding approximate distribution in the GPD. Moreover, the parameters of the GPD corresponding to the threshold excesses of a given time series are uniquely determined by those of the associated GEV distribution of block maxima. In particular, the shape parameter is the same, and the scale parameters of the two distributions are related by

$$\tilde{\sigma} = \sigma + \xi(u - \mu). \quad (2.5)$$

The choice of the threshold in the POT method (which is analogous to the choice of block size in the block maxima approach) represents a trade-off between bias and variance: too low a threshold is likely to violate the asymptotic basis of the model, leading to bias; too high a threshold will generate fewer excesses with which to estimate the model, leading to high variance.

An important property of the POT/GPD approach is the *threshold stability property*. Suppose that the GPD is a correct (or approximately correct) model for describing the excesses over a threshold  $u_0$ , i.e. that

$$P(X - u_0 > x \mid X > u_0) \equiv \frac{P(X > x + u_0)}{P(X > u_0)} = (1 + \xi x / \tilde{\sigma}_0)^{-1/\xi}, \quad (2.6)$$

for some scale parameter  $\tilde{\sigma}_0$ , where the left-hand side is to be read as the ‘‘probability that an excess  $X - u_0$  above  $u_0$  exceeds  $x$ ’’. Then it follows that if  $u_1 > u_0$



$$\begin{aligned}
P(X - u_1 > x | X > u_1) &\equiv \frac{P(X > x + u_1)}{P(X > u_1)} = \frac{P(X - u_0 > x + u_1 - u_0)}{P(X - u_0 > u_1 - u_0)} \\
&= \frac{(1 + \xi(x + u_1 - u_0) / \tilde{\sigma}_0)^{-1/\xi}}{(1 + \xi(u_1 - u_0) / \tilde{\sigma}_0)^{-1/\xi}} \\
&= \left( \frac{\tilde{\sigma}_0 + \xi(x + u_1 - u_0)}{\tilde{\sigma}_0 + \xi(u_1 - u_0)} \right)^{-1/\xi} \\
&= \left( 1 + \frac{\xi}{\tilde{\sigma}_0 + \xi(u_1 - u_0)} x \right)^{-1/\xi}
\end{aligned} \tag{2.7}$$

which means that the excesses over a higher threshold  $u_1$  also follow a GPD with the same shape parameter and with a scale parameter  $\tilde{\sigma}_1 = \tilde{\sigma}_0 + \xi(u_1 - u_0)$ . Since this last relation between the two scale parameters can be written as  $\tilde{\sigma}_1 - \xi u_1 = \tilde{\sigma}_0 - \xi u_0$ , we conclude that if the GPD is a correct model for the excesses above a given threshold then the GPD is also a correct model for the excesses above any threshold  $u$  above that threshold and that the scale parameter  $\tilde{\sigma}_u$  corresponding to  $u$  is such that

$$\sigma^* = \tilde{\sigma}_u - \xi u \tag{2.8}$$

does not depend on  $u$  (it is constant in  $u$ ).

This stability property of the GPD can be used to find the minimum threshold at which a GPD model applies to the data, since in the range of threshold values for which the GPD provides a good approximation to the excesses the *estimates* of  $\tilde{\sigma}_u - \xi u$  should remain approximately constant.

In this report we have used the threshold stability property to choose the threshold on the basis of which the sample of peak excesses is selected and used to estimate the GPD. More precisely, we have looked for threshold values around which the estimate of the shape parameter shows the least variation. We have automatized such a choice of the threshold using the following procedure:

- 1 POT samples with at least 10 and at most 300 peaks are collected by systematically decreasing the threshold, and for each of these samples GPD fits are obtained. Note that if there is a POT sample with, say, 20 peaks, then it does not follow that there is also a POT sample with 19 peaks, since different peaks may have the same value and even a small increase of the threshold can eliminate more than one of the peaks collected at a lower threshold.
- 2 For each sample size  $n$ , a set of parameter estimates based on sample sizes ranging from  $n-l$  to  $n+l$  peaks, where  $l$  is some fixed value (see below), are obtained, and the standard deviation ( $v$ ) and the linear slope ( $b$ ) of such a set of estimates is computed. In the case of the shape parameter, for example, this procedure yields one standard deviation for each value of  $n$ , and each standard deviation quantifies the variability of the parameter estimates around a 'window' of  $2l+1$  sample sizes ( $2l+1=(n+l)-(n-l)+1$ ).
- 3 The threshold, or sample size  $n$ , that is then used for making inferences is the one yielding the smallest ( $v+b$ ) value computed in bullet 2.

In previous analyses of wave and wind speed data (Caires, 2007 and 2009) several tests were carried out to determine the best choice of the window size on which the standard deviations of the second bullet are computed.  $l=10$  turned out to be a good choice: using about 21 ( $2l+1$  for  $l=10$ ) estimates, the automatically determined threshold coincides often with the one that we would have chosen by visual inspection of plots. In most cases the results are insensitive to changes in  $l$  for  $l$  between 10 and 15. With larger values of  $l$  the threshold chosen is often too low.

### 2.2.3 Domains of attraction

It is worth emphasizing that although the AM and POT methods are based on fitting limit distributions (the GEV and the GPD) to certain samples (samples of block maxima and samples of peak excesses), the distribution functions governing the behaviour of those samples can be quite arbitrary; in particular, it is not necessary at all that those distributions be the GEV or GPD. What is necessary for the inferences resulting from the AM and POT methods to be valid is that the distributions governing the data belong to the *domain of attraction* of extremes. A distribution function  $F$  is said to belong to the domain of attraction of the extreme value distributions if there is a positive function  $f$  such that

$$\frac{1 - F(u + xf(u))}{1 - F(u)} \rightarrow (1 + \xi x)^{-1/\xi} \quad (2.9)$$

as  $u \uparrow x_F$ , where  $x_F$  is the *right endpoint* of  $F$ , i.e. the number  $x_F \leq \infty$  such that  $F(x_F) = 1$  and  $F(x_F - \varepsilon) < 1$  for all  $\varepsilon > 0$  (see Section 1.2 of de Haan and Ferreira, 2006).

The function on the right in this limiting relation can be recognized as 1 minus a GPD with scale parameter equal to 1, and the function on the left should be interpreted as the probability that the excess over the threshold  $u$  of an observation governed by  $F$  is greater than  $xf(u)$  conditionally on it being greater than  $u$ .

In sections 3.3, 5.3 and in Appendix A we will meet three distributions that are neither the GEV nor the GPD but which belong to the domain of attraction of the extreme value distributions, and hence to which the results of extreme value theory apply. In order to verify that such distributions are indeed in the domain of attraction of the extreme value distributions it suffices to exhibit a function  $f$  with which the above limiting relation (as  $u \uparrow x_F$ ) holds; see Appendix A.

## 2.3 Estimation

There are several numerical methods available for the estimation of the parameters of extreme value distributions. Some of them, for instance the method of moments and the method of probability weighted moments (PWM), give explicit expressions for the parameter estimates, but others, such as the maximum likelihood (ML) method, require numerically solving non-linear systems of equations in order to determine the parameter estimates. The ML method tends to be the preferred estimation method because it is quite general and more flexible than other methods, for instance when extending the extreme value approach to account for non-stationarity. In the case of the GEV and GPD distributions, the ML estimators

are consistent<sup>5</sup> for  $\xi > -0.5$  and the PWM estimators for  $\xi < 0.5$ . On the basis of the work of Hosking et al. (1985) and Hosking and Wallis (1987), for the range of tails typical of the basic variables, whose distributions are not too heavy-tailed, and for small to moderate sample sizes, the method of PWM is expected to perform better than the ML method in the estimation of the GEV and the GPD parameters. The performance of both ML and PWM methods will be investigated in Chapter 4 under a variety of scenarios that are relevant to the determination of HBC and which have not been considered by the above mentioned authors.

## 2.4 Serial dependence

The data on the basic variables—wave, water level, wind speed, etc.—which have to be analysed in order to determine HBC are typically available every hour to every 6 hours, depending on whether they consist of measurements or of model results. These data have to be regarded as *statistically dependent*, in the sense that the value of, say, the significant wave height at a given time is to a certain extent predictable from its value 6 hours earlier. However, in order to carry out an extreme value analysis it is required that the data used to fit the distribution be approximately independent. Consequently, the samples used in the AM and POT methods have to be extracted from the original time series in such a way as to make this assumption realistic. In the case of the AM method this is achieved by ensuring that if the annual maximum of a given year belongs to a storm that straddles that year and the next, then the annual maximum of the following year is not to be picked within that storm. In the case of the POT method this is done by a process of *declustering* in which only the peak (highest) observations in clusters of successive exceedances of a specified threshold are retained and, of these, only those which in some sense are sufficiently apart (so that they belong to more or less ‘independent storms’) are included in the POT sample. Based on storm duration studies, it has been concluded that the hypothesis of independence of the POT data is tenable by treating cluster maxima at a distance of less than 2-4 days apart as belonging to the same cluster (storm); see e.g. Dillingh et al., 1993.

---

5. Converge in probability to the true value of the parameter being estimated.

### 3 Setup of the study

The objective of this study is to provide advice on the choice between the POT/GPD and AM/GEV approaches to estimate univariate distributions of extreme values of basic variables on the basis of Monte Carlo simulation. For most of the Dutch coast, the return periods required for the determination of HBC are either of 4,000 or 10,000 year. It is therefore natural to assess the performance of the two methods regarding the estimation of the 4,000-yr and 10,000-yr return values across a range of scenarios specifying the tail of the data and the amount of available data.

#### 3.1 Characteristics of the real data

A number of points need to be made about the set of quality-checked measurements of the basic variables available for the determination of the HBC, as these are relevant to the choice of scenarios chosen in our simulations:

- About 30 years of wave data are available; these include measurements of significant wave height and wave period.
- Depending on the measuring station, there are time series with 5 to 50 years of wind data available.
- Water levels have been observed since the end of the 19<sup>th</sup> century and, depending on the tidal gauge location, time series with 70 to 120 years of still water level (SWL) are available.
- There are about 100 years of discharge levels of the Rhine and Meuse rivers.
- Finally, there are 30 years of water level measurements of the IJssel lake.

The wave, SWL and wind speed data have already been the subject of AM/GEV and POT/GPD analyses, and these analyses suggest realistic specifications of the length of the time series to be simulated and of the shape parameter  $\xi$  of the GEV and GPD distributions:

- Dillingh et al. (1993) analysed the SWL data up to 1985. The length of the time series considered varied from 50 to 100 years. The estimates of  $\xi$  of the GPD varied between -0.17 and 0.01 and those of the GEV distribution between -0.24 and 0.04.
- Caires (2007) analysed the North Sea significant wave height data from 1979 to 2002, obtaining estimates of  $\xi$  between -0.32 and -0.07 with the GPD and between -0.38 and 0.18 with the GEV distribution.
- Caires (2009) analysed the wind velocity data from 1970 to 2008, obtaining estimates of  $\xi$  between -0.05 and 0.02 with the GPD and between -0.24 and 0.17 with the GEV distribution.

Unfortunately, no POT/GPD nor AM/GEV analyses have been carried out with the water level data from lake IJssel and with the Rhine and Meuse river discharges. It seems, however, reasonable to assume that the shape parameters of these data are within the ranges found in the wave, wind and SWL data.

In summary, it can be concluded that the length of the available time series varies from 5 to 120 years and that the shape parameters underlying the data are likely to vary from -0.3 to 0.1.

### 3.2 Previous simulation studies

The finite-sample properties of the ML and PWM estimators of the GEV and GPD in the case of independent and identically distributed observations following exactly these distributions have been studied via simulation by Hosking et al. (1985) and Hosking and Wallis (1987), respectively. These authors considered five values of the shape parameter (-0.4, -0.2, 0, 0.2 and 0.4). In the case of the GEV, the performance of the PWM estimators is found to be generally better than that of the ML estimators for sample sizes of up to 50; for larger sample sizes the performance of the methods was found to be comparable. In the case of the GPD, and for sample sizes of up to 100, the performance of the PWM estimators is superior to that of the ML estimators for shape parameters 0, 0.2 and 0.4 in terms of the root mean square error, and superior in the whole range of shape parameters considered in terms of bias. For sample sizes bigger than 100 the behaviour of the two methods is similar. When the sample sizes are less than 50 the ML estimators of both distributions sometimes do not exist especially when the shape parameter negative.

There are a number of simulation studies (Cunnane, 1973, Yevjevich and Taesombut, 1978, and Tavares da Silva, 1983) on the performance of the ML estimators of the Gumbel and Exponential distributions (the case  $\xi = 0$  in the GEV and GPD) under the assumption that  $\lambda$ , the expected number of clusters per year (recall the beginning of Subsection 2.2.1), is bigger than 1. These studies say that for  $\lambda > c$ , with  $c$  varying between 1.65 and 2 depending on the study, the variance (uncertainty) in the estimates of the exponential distribution is lower than that of the Gumbel distribution. On the other hand, all studies conclude that the opposite is true when  $\lambda = 1$  and the AM and POT are based on time series of the same length.

A different type of study is that of van den Brink et al. (2005). These authors used 'ensemble simulations' with a total length of 7540 years from a certain climate model to compute storm surges at a North Sea location, from which 65 records of 116 years were constructed and analyzed using the AM/GEV and POT/GPD methods, the parameters of the distributions being estimated by the ML method. The methods were compared in terms of the uncertainties in the estimates of the 10,000-year return value, the 'true' 10,000-year return value being estimated from the entire 7540-year dataset. The conclusions of this study were that in the POT/GPD approach a threshold could not be determined objectively from the 116-year long time series, that if a threshold was to be chosen then such a threshold should be associated with a value of  $\lambda$  not larger than 4, and that given the difficulties and uncertainties associated with the choice of the threshold the AM/GEV approach is preferable.

Finally, de Valk (1993) simulated wave data with a type I tail and sample sizes of 10, 25 and 100 years and studied the performance of the POT/GPD method based on three fixed thresholds of 4, 5 and 6 metres. His conclusion was that unless the sample size is 100 years the return value estimates based on the fitted GPD underestimate the true return value. He also considered Bayesian estimates of the parameters of the GPD and concluded that for each of the sample sizes considered one can choose a threshold for which the bias and standard deviation of the estimates are acceptable.

### 3.3 The present simulation study

The available simulation studies assessing the performances of the POT/GPD and/or AM/GEV methods, and in particular those mentioned in the preceding subsection, do not cover all the scenarios that are relevant for the determination of HBC. Thus, they do not cover the whole range of shape parameters that are likely to underlie the basic variables, do not cover the whole range of data available on those variables, do not consider non-stationarity and serial dependence in the simulated time series, or do not make use of the threshold stability property in the POT/GPD analyses as a means of choosing the threshold. In this study we carry out two simulation studies with the purpose of quantifying the performance of the POT/GPD and the AM/GEV approaches with time series of dependent, non-stationary data whose length can vary from 5 to 120 years and whose underlying shape parameters vary from -0.3 to 0.1.

The first study concerns the finite-sample properties of the ML and PWM estimators of the GEV and GPD. It is based on simulated series of independent and identically distributed measurements with durations varying from 10 to 200 years, and it is similar to the studies of Hosking et al. (1985) and Hosking and Wallis (1987), except that more shape parameter values in the range of -0.3 to 0.1 are considered.

The second study concerns the finite-sample properties of the AM/GEV and POT/GPD approaches with time series of dependent measurements subject to seasonal variations and with a fixed tail index (the underlying shape parameter of the associated extreme value distribution; recall Subsection 2.2), the duration of the simulated time series also ranging from 10 to 200 years.

In both studies, the two methods are assessed in terms of their accuracy in the estimation of the underlying shape parameter and of two high percentiles, namely the 4,000- and 10,000-yr return values.

The serial dependence incorporated into the time series of the second study is based on normal (or Gaussian) autoregressive (AR) models. Guedes Soares et al. (1996) used autoregressive moving average (ARMA) processes to model time series of significant wave height measurements offshore Portugal and found that AR models (which constitute a special case of the ARMA models) could be used to satisfactorily describe 'de-seasonalized' time series (i.e. the series with the seasonal and average components removed) of three-hourly significant wave height measurements. One of the models fitted by them was used here as a means of specifying a dependence structure for the simulated series of three-hourly measurements. This is an AR(19) model, which is to say that the value of an observation at a given time can be expressed as a linear combination of observations at *lags* 1 to 19, i.e. as a subset of the previous 19 observations, plus a random error; the non-zero coefficients of this linear combination are given in Table 3.1.

Lag	Coefficient
1	1.045
2	-0.259
3	0.043
7	0.157
10	-0.080
16	0.044
18	-0.074
19	0.072

Table 3.1 Non-zero coefficients of the AR(19) model obtained by Guedes Soares et al. (1996) from three-hourly significant wave height data from the North Atlantic.

AR time series such as these are stationary and have a normal marginal distribution with mean 0 and a certain variance. In order to create a time series possessing realistic features, the time series of the AR(19) have to undergo a couple of transformations. In the first place, the simulated AR(19) series of three hourly measurements, denoted by  $X_t$ , are *standardized*, i.e. divided by their standard deviation, which in the case of the AR model of Guedes Soares et al. (1996) equals 1.48. This transformation yields a stationary time series, given by  $Y_t = X_t/1.48$ , possessing a *standard normal* marginal distribution, i.e. a normal distribution with mean 0 and variance 1. Next, to this standardized series one applies the standard normal distribution function  $\Phi$ , obtaining a *uniform time series*  $U_t = \Phi(Y_t)$  possessing a *standard uniform* marginal distribution (i.e. a uniform distribution on the interval  $[0,1]$ ).

Now if  $U$  is a standard uniform random variable one may obtain a random variable  $V$  with an arbitrary continuous and strictly increasing distribution function  $F$  by means of the transformation  $V = F^{-1}(U)$ , where as usual  $F^{-1}$  denotes the inverse function of  $F$ . Consequently, if one wants to transform the uniform time series in such a way as to obtain a time series that follows a particular distribution function  $F_t$  at time  $t$ , one simply has to set  $V_t = F_t^{-1}(U_t)$ , where  $F_t^{-1}$  is the inverse of  $F_t$ . Thus the third and final transformation consists of applying the inverse of a certain distribution function  $F_t$  to  $U_t$ , the uniform time series at time  $t$ . This  $F_t$ , however, is not arbitrary: it is chosen so as to have a specified tail index, and a certain mean and a certain standard deviation determined by the month in which the measurement is taken. In this way, the time series of  $V_t$  measurements that result from the three transformations possess a dependence structure, a unique tail index, and marginal means and variances that follow a seasonal pattern.

In order to specify realistic monthly means and standard deviations we have based ourselves on the three-hourly time series of significant wave height measurements from 1979 to 2002 from the Schiermonnikoog Noord (SON) buoy located in the North Sea. Using these 24 years of data, we have computed the monthly sample averages and standard deviations for each year and then averaged the 24 sets of monthly averages and standard deviations to obtain a single set of monthly averages and monthly standard deviations, given in Table 3.2. It is this set of monthly averages and monthly standard deviations that is used to specify the  $F_t$  just mentioned for a given  $t$ .

	M	S
Jan	1.39	0.80
Feb	1.34	0.78
Mar	1.22	0.71
Apr	1.02	0.66
May	0.90	0.52
Jun	0.98	0.58
Jul	0.95	0.55
Aug	0.96	0.59
Sep	1.17	0.71
Oct	1.30	0.76
Nov	1.38	0.77
Dec	1.44	0.80

Table 3.2 Average monthly means ( $M$ ) and standard deviations ( $S$ ) of the SON data from 1970 to 2002.

Note that if  $F_t$  is a distribution function with the desired tail index and two *free parameters* (i.e. two parameters that are unrestricted by the tail index, which is fixed beforehand), and possessing mean and variance, then one can use the method of moments to determine values of the two free parameters that yield a mean and a variance equal to the targeted monthly mean and monthly variance. Of course, there are many families of distributions satisfying these requirements, so in spite of having fixed the dependence structure, the tail index, and the monthly means and variances, we are still left with an infinite number of degrees of freedom to fully specify our time series model. In order to complete the specification of the time series model we have considered two families of distributions as candidates for the  $F_t$ s.

In the first place we have considered the GEV distribution with the tail index fixed at -0.3, -0.2, -0.1, 0 or 0.1 and the location and scale parameters  $\mu_t$  and  $\sigma_t$  determined by the method of moments so as to force  $F_t$  to have the mean and variance appropriate to the month in which the measurement at time  $t$  falls.

Secondly, we have considered a trio of models comprising the Gamma distribution, whose tail index is 0 (i.e. its tail is of type I or exponential), the Beta distribution function, whose tail index is negative (it has a type III tail), and the Beta distribution function of the second kind with a scale parameter, whose tail index is positive (it has a type II tail). Each of these models has two free parameters which can be determined so as to yield time series with the required tail index and the required monthly means and standard deviations. In contrast with our simulations based on the GEV distribution, however, our simulations with this second set of distributions only cover tail indices of -0.1, 0 and 0.1, due to the high costs of the computations involving the Beta distributions.

Appendix A provides the definitions of this trio of distributions and a summary of their properties which are relevant to the carrying out of the simulations. Parallel results on the GEV distribution are also included in Appendix A.

The purpose of considering the Gamma and the two types of Beta distribution in our simulation study is to introduce an extra dimension in the estimation of extreme values based on the limiting distributions provided by probability theory. More precisely, when applying the methods of extreme value theory to the simulated time series with GEV marginal distributions the quality of the approximation provided by the limiting GEV distribution is affected by the



dependence and non-stationarity of the data, but not by the choice of the marginal distributions, which are “already in the limit“. In contrast, with the simulated time series based on the Gamma and on the two types of Beta distributions one may expect the approximation provided by the limiting distributions to be worse, as none of those distributions is “already in the limit“.

Given the complex form of the dependent and non-stationary simulated time series, their high percentiles cannot be computed analytically. Moreover, since those percentiles concern very small probabilities of exceedance, huge numbers of simulations would be required to estimate them non-parametrically with any decent degree of accuracy. Therefore, we have estimated the true percentiles of each time series model by simulating 100,000 years of data and computing AM/GEV estimates from them; it is these AM/GEV estimates that are taken to be the true parameter values and subsequently used to assess the performance of our two approaches.

Finally, notice that although we have used real data to create time series with realistic *features*, the time series *themselves* are not realistic, in the sense that they do not mimic a particular type of data from a particular geographical region. Ideally, it would be more satisfactory to create and analyze different time series mimicking the behaviour of all the basic variables in various geographical regions, but that would represent a huge computational effort well beyond the scope of the present work; furthermore, our approach, by incorporating the main features of environmental time series and plausible values of the tail index, should yield rather general conclusions.

## 4 Finite-sample properties of the ML and PWM estimators of the GEV and GPD models

### 4.1 Introduction

In this chapter we present the results of the simulations illustrating the finite sample properties of the ML and PWM estimators of the GEV and GPD models. The simulations were carried out for five values of the shape parameter, namely  $\xi = -0.3, -0.2, -0.1, 0$  and  $0.1$ , and five sample sizes representing five different durations in years, namely  $n_y = 10, 20, 50, 100$  and  $200$  years.

In each case we carry out  $k = 100,000$  simulations of a sample of size  $\lambda n_y$  from the GPD with  $\tilde{\sigma} = 1$  and of a sample of size  $n_y$  from the GEV with  $\sigma = 1$  and  $\mu = 0$ . A single value of  $\lambda$  is considered in the simulations, namely  $\lambda = 5$ . On the basis of analyses of real time series of the basic variables, it is expected that at least five storm peaks can be picked per year in a POT/GPD analysis, so in particular it does not make much sense to compare the two distributions using equal sample sizes (i.e. taking  $\lambda = 1$ ). Since different numbers of years of data are considered in the simulations, results for other values of  $\lambda$  can also be deduced from the results for  $\lambda = 5$ .

In each simulation, the parameters of the distributions and the associated 4,000- and 10,000-yr return values are estimated using the ML and PWM methods.

The differences between the ML and PWM estimates and the true values of the parameters were quantified in terms of the *bias*, defined by  $k^{-1} \sum \hat{\theta}_i - \theta$ , and the root-mean-square error, defined by  $RMSE = \sqrt{k^{-1} \sum (\hat{\theta}_i - \theta)^2}$ , where  $\theta$  is the true parameter value and the  $\hat{\theta}_i$ s are the estimates of  $\theta$ . Unless  $\theta$  under consideration is the shape parameter, the results are presented in terms of percentual relative bias and RMSE, which are the bias and RMSE normalized (divided) by the true parameter values and multiplied by 100.

In all simulations the scale parameter is set to 1, which involves no loss of generality since both the ML and the PWM methods are invariant under scale transformations of the data. The relative RMSE and bias of the estimates of the percentiles, however, are not invariant under scale transformations, but this is not a problem because our main interest lies in comparing the values of the error statistics in the two methods and in how those values vary with the sample size and not in their magnitudes.

Since in our notation of Eqs. (2.2), (2.4) and (2.5) the parameters of the GEV and GPD models used in the simulations satisfy  $\tilde{\sigma} = 1$ ,  $\sigma = 1$  and  $\mu = 0$ , the  $m$ -year return value of the GEV corresponds approximately (for large  $m$ ) to the  $\lambda m$ -year return value of the GPD model when  $\lambda = 1$ .

The results of the simulations are presented and analysed in the next section. In Section 4.3 we illustrate the effect of choosing other scale and location parameters on the magnitude of the bias and RMSE statistics of the GEV distribution.

## 4.2 Analysis of the results

Table 4.1 to Table 4.4 present the error statistics of the ML and PWM estimators of the scale and shape parameters and of the 4,000- and 10,000-yr return values of the GEV and GPD models.

In the case of the GEV the sample size is  $n_y$  and in the case of the GPD the sample size is  $\lambda n_y = 5n_y$ . Since  $\lambda \neq 1$ , the 4,000- and 10,000-yr return values of the GEV and GPD models differ; cf. Table 4.1, Table 4.2, and Eq. (2.4).

The ML estimates cannot always be obtained because the likelihood function has a maximum value with probability less than 1 (though the probability approaches 1 as the sample size increases). Table 4.5 gives the ‘failure rate’ of the ML estimation of the GPD and GEV. It can be seen that the failure rates increase with decreasing shape parameter (lighter tails) and decreasing sample size. Given that the sample sizes of the GEV are smaller, there are more failures in the estimation of the GEV than in that of the GPD.

Looking at the performance of the PWM and ML methods per parameter, we can draw the following conclusions:

- *Return values:* In the case of the GEV, for  $n_y$  below 50 years PWM generally behaves better than ML, especially for  $\xi > -0.2$ ; for larger sample sizes the behaviour of the PWM and ML estimators is similar. In the case of the GPD, for  $n_y$  below 50 years ML generally behaves better for  $\xi \leq -0.1$ , while PWM behaves better for  $\xi > -0.1$ ; for larger sample sizes the behaviour of the two methods is similar.
- *Shape and scale parameter estimates:* In the case of the GEV, PWM is generally better in terms of bias and RMSE. In the case of the GPD, the differences in the error statistics of the two methods are small for sample sizes of more than 20 years; for smaller sample sizes PWM does slightly better in terms of RMSE for  $\xi \geq -0.1$  and always better in terms of bias.

Comparing the errors of the GEV and GPD estimators, the relative RMSEs and the absolute values of the relative bias of the GPD estimates are lower, as expected. However, for sample sizes of 200 the differences are quite small.

In order to have an idea of how the GPD error statistics vary with  $\lambda$  we analyse the results further. When  $\lambda = 1$ , if we compare the performance of the GPD estimates with  $n_y$  equal to 10 and 20 (which correspond to a number of peaks of 50 and 100, respectively) with those of the GEV estimates with  $n_y$  equal to 50 and 100, we have the following conclusions:

- In the case of the GEV, the RMSEs of the estimate of  $\xi$  range from 0.11 to 0.12 for  $n_y = 50$  (from 0.07 to 0.08 for  $n_y = 100$ ); in the case of the GPD, the RMSEs of the estimate of  $\xi$  range from 0.17 to 0.20 for  $n_y = 10$  (from 0.12 to 0.14 for  $n_y = 20$ ).

- In the case of the GEV, the relative RMSEs of the estimate of  $\sigma$  range from 11% to 14% for  $n_y = 50$  (from 7% to 9% for  $n_y = 100$ ); in the case of the GPD, the RMSEs of the estimate of  $\tilde{\sigma}$  range from 22% to 25% for  $n_y = 10$  (from 14% to 16% for  $n_y = 20$ ).

Thus, whenever the sample size (number of peaks) of the POT method is equal to that of the AM method (the annual maxima) the uncertainty of the GEV estimates is lower.

In the situation where  $\lambda = 2$ , the comparison of the GPD estimates for  $n_y = 20$  (a sample size of 100) with the GEV estimates for  $n_y = 50$  yields the following conclusions:

- In the case of the GEV, the RMSEs of the estimate of  $\xi$  range from 0.11 to 0.12, and in the case of the GPD the RMSEs of the estimate of  $\xi$  range from 0.12 to 0.14.
- The relative RMSEs of the GEV estimate of  $\sigma$  range from 11% to 14%, and those of the GPD estimate of  $\tilde{\sigma}$  from 14% to 16%

The errors of the two methods are thus comparable, as had been noted by Tavares and da Silva (1983) when  $\xi=0$ ; when  $\lambda > 2$ , however, the uncertainty in the GEV estimates is greater than that of the GPD estimates.

In conclusion, we can say that for  $\lambda > 2$  the uncertainties of the GPD estimates are lower than the corresponding GEV estimates, and that for data sets of more than 200 years the uncertainties of the GEV estimates are low.

Furthermore, with data sets of less than 50 years, the PWM method is preferable to the ML method, both because of its better error characteristics and robustness against 'non existence' problems. With the longer datasets considered here, the two methods provide comparably accurate estimates. Asymptotically, however, the ML method provides better estimates than the PWM method for  $\xi < 0.05$ ; this was pointed out to us by the Reviewer (see Section B.2).

GEV						GPD							
4,000-yr rv		3.06	4.05	5.64	8.29	12.92	4,000-yr rv		3.16	4.31	6.29	9.90	16.92
$\xi$		-0.30	-0.20	-0.10	0.00	0.10	$\xi$		-0.30	-0.20	-0.10	0.00	16.92
$n_y$		Relative bias (%)					$n_y$		Relative bias (%)				
10	PWM	34.47	40.68	48.81	55.53	62.92	10	PWM	10.68	13.12	15.89	18.98	23.10
10	ML	—*	—*	—*	—*	—*	10	ML	-6.15	-4.41	0.22	10.00	26.20
20	PWM	11.77	14.47	17.86	22.62	28.27	20	PWM	5.15	6.43	7.78	9.47	11.75
20	ML	371.56	94.32	—*	—*	—*	20	ML	-4.12	-3.47	-1.56	2.37	9.42
50	PWM	3.82	4.53	5.97	7.72	10.50	50	PWM	2.04	2.44	3.11	3.83	4.71
50	ML	-5.31	-1.68	2.83	8.86	17.90	50	ML	-2.17	-2.01	-1.14	0.44	2.96
100	PWM	1.70	2.12	2.82	3.96	5.37	100	PWM	0.98	1.30	1.56	1.92	2.33
100	ML	-3.88	-1.76	0.37	3.19	6.75	100	ML	-1.28	-1.14	-0.72	0.12	1.33
200	PWM	0.70	0.92	1.26	1.81	2.79	200	PWM	0.47	0.65	0.73	0.91	1.17
200	ML	-2.64	-1.33	-0.18	1.15	3.05	200	ML	-0.76	-0.64	-0.45	0.01	0.65
$n_y$		Relative RMSE (%)					$n_y$		Relative RMSE (%)				
10	PWM	113.69	135.10	167.08	202.31	254.95	10	PWM	48.09	58.92	73.36	91.38	118.17
10	ML	—*	—*	—*	—*	—*	10	ML	25.95	39.99	67.83	123.38	183.67
20	PWM	48.56	60.27	75.68	99.60	135.25	20	PWM	29.11	36.36	44.85	55.69	70.25
20	ML	—*	—*	—*	—*	—*	20	ML	15.39	22.90	34.50	52.63	78.86
50	PWM	23.80	29.64	37.41	48.56	65.70	50	PWM	16.77	20.85	25.85	31.74	39.26
50	ML	24.35	29.88	45.79	61.11	97.86	50	ML	8.88	13.31	19.53	27.97	39.08
100	PWM	15.72	19.40	24.67	31.83	42.06	100	PWM	11.45	14.27	17.63	21.64	26.41
100	ML	15.86	17.96	24.15	32.97	43.85	100	ML	6.01	9.07	13.35	18.97	25.69
200	PWM	10.64	13.27	16.71	21.46	27.99	200	PWM	7.94	9.92	12.24	15.04	18.42
200	ML	10.70	11.97	15.73	21.08	27.43	200	ML	4.08	6.29	9.25	13.18	17.77

Table 4.1 Relative RMSE and bias of the ML and PWM estimators of the 4,000-yr return values. \*Values in excess of 1,000.

GEV							GPD						
10,000-yr rv		3.12	4.21	6.02	9.21	15.12	10,000-yr rv	3.20	4.43	6.61	10.82	19.50	
	$\xi$	-0.30	-0.20	-0.10	0.00	0.10		$\xi$	-0.30	-0.20	-0.10	0.00	0.10
$n_y$		Relative bias (%)					$n_y$		Relative bias (%)				
10	PWM	45.32	55.68	69.69	82.26	97.19	10	PWM	12.88	16.51	20.89	25.96	32.58
10	ML	—*	—*	—*	—*	—*	10	ML	-5.86	-3.48	2.91	16.72	39.82
20	PWM	14.53	18.62	24.00	31.73	41.60	20	PWM	6.08	7.94	10.05	12.69	16.22
20	ML	—*	323.96	—*	—*	—*	20	ML	-4.09	-3.23	-0.74	4.49	13.96
50	PWM	4.58	5.70	7.80	10.52	14.89	50	PWM	2.37	2.99	3.96	5.06	6.40
50	ML	-4.99	-0.84	4.76	12.68	25.04	50	ML	-2.20	-1.98	-0.91	1.10	4.39
100	PWM	2.04	2.65	3.66	5.30	7.47	100	PWM	1.14	1.58	1.97	2.52	3.15
100	ML	-3.83	-1.52	1.01	4.57	9.17	100	ML	-1.31	-1.14	-0.63	0.42	1.99
200	PWM	0.85	1.16	1.65	2.44	3.80	200	PWM	0.54	0.79	0.93	1.20	1.58
200	ML	-2.67	-1.27	0.07	1.72	4.10	200	ML	-0.79	-0.65	-0.42	0.16	0.97
$n_y$		Relative RMSE (%)					$n_y$		Relative RMSE (%)				
10	PWM	150.35	183.72	234.82	292.03	381.82	10	PWM	54.71	68.98	88.64	113.73	152.66
10	ML	—*	—*	—*	—*	—*	10	ML	28.69	46.82	86.76	178.68	265.63
20	PWM	56.27	72.25	93.26	128.17	183.30	20	PWM	31.79	40.81	51.72	65.81	85.08
20	ML	—*	—*	—*	—*	—*	20	ML	16.23	24.92	39.16	62.75	98.29
50	PWM	25.98	33.17	42.94	57.37	81.00	50	PWM	17.88	22.77	28.89	36.19	45.49
50	ML	25.88	33.56	60.57	76.48	139.00	50	ML	9.32	14.29	21.51	31.65	45.33
100	PWM	16.90	21.29	27.69	36.59	49.68	100	PWM	12.11	15.43	19.48	24.37	30.12
100	ML	16.58	19.53	27.05	38.10	51.97	100	ML	6.30	9.71	14.63	21.23	29.26
200	PWM	11.36	14.44	18.55	24.29	32.31	200	PWM	8.37	10.68	13.45	16.82	20.84
200	ML	11.17	12.91	17.39	23.82	31.56	200	ML	4.28	6.73	10.11	14.68	20.08

Table 4.2 Relative RMSE and bias of the ML and PWM estimators of the 10,000-yr return values. \*Values in excess of 1,000.

GEV							GPD								
		$\xi$	-0.30	-0.20	-0.10	0.00	0.10			$\xi$	-0.30	-0.20	-0.10	0.00	0.10
$n_y$		Bias					$n_y$		Relative bias (%)						
10	PWM	-0.02	-0.03	-0.04	-0.05	-0.07	10	PWM	-0.02	-0.02	-0.03	-0.03	-0.04		
10	ML	-0.13	-0.09	-0.06	-0.04	-0.01	10	ML	-0.08	-0.07	-0.07	-0.06	-0.06		
20	PWM	-0.01	-0.01	-0.02	-0.02	-0.03	20	PWM	-0.01	-0.01	-0.01	-0.01	-0.02		
20	ML	-0.08	-0.06	-0.04	-0.02	-0.01	20	ML	-0.04	-0.03	-0.03	-0.03	-0.03		
50	PWM	0.00	0.00	-0.01	-0.01	-0.01	50	PWM	0.00	0.00	-0.01	-0.01	-0.01		
50	ML	-0.04	-0.03	-0.02	-0.01	0.00	50	ML	-0.02	-0.02	-0.01	-0.01	-0.01		
100	PWM	0.00	0.00	0.00	0.00	-0.01	100	PWM	0.00	0.00	0.00	0.00	0.00		
100	ML	-0.03	-0.01	-0.01	0.00	0.00	100	ML	-0.01	-0.01	-0.01	-0.01	-0.01		
200	PWM	0.00	0.00	0.00	0.00	0.00	200	PWM	0.00	0.00	0.00	0.00	0.00		
200	ML	-0.02	-0.01	0.00	0.00	0.00	200	ML	-0.01	0.00	0.00	0.00	0.00		
$n_y$		RMSE					$n_y$		RMSE						
10	PWM	0.30	0.29	0.29	0.29	0.30	10	PWM	0.20	0.18	0.18	0.17	0.17		
10	ML	0.50	0.50	0.51	0.51	0.53	10	ML	0.17	0.17	0.18	0.18	0.19		
20	PWM	0.18	0.18	0.18	0.19	0.19	20	PWM	0.14	0.13	0.12	0.12	0.12		
20	ML	0.27	0.25	0.24	0.24	0.25	20	ML	0.10	0.11	0.11	0.12	0.12		
50	PWM	0.11	0.11	0.11	0.11	0.12	50	PWM	0.08	0.08	0.08	0.07	0.07		
50	ML	0.15	0.13	0.12	0.12	0.12	50	ML	0.06	0.06	0.06	0.07	0.07		
100	PWM	0.07	0.07	0.07	0.08	0.08	100	PWM	0.06	0.06	0.05	0.05	0.05		
100	ML	0.10	0.08	0.07	0.08	0.08	100	ML	0.04	0.04	0.04	0.05	0.05		
200	PWM	0.05	0.05	0.05	0.05	0.06	200	PWM	0.04	0.04	0.04	0.04	0.04		
200	ML	0.07	0.05	0.05	0.05	0.06	200	ML	0.03	0.03	0.03	0.03	0.04		

Table 4.3 RMSE and bias of the ML and PWM estimators of  $\xi$ .

GEV						GPD							
$\sigma$	1	1	1	1	1	$\tilde{\sigma}$	1	1	1	1	1		
$\xi$	-0.30	-0.20	-0.10	0.00	0.10	$\xi$	-0.30	-0.20	-0.10	0.00	0.10		
$n_y$	Relative bias (%)					$n_y$	Relative bias (%)						
10	PWM	-2.84	-2.47	-2.10	-1.58	-1.07	10	PWM	2.43	2.48	2.78	2.91	3.50
10	ML	-17.08	-14.26	-12.78	-11.89	-11.61	10	ML	7.49	7.03	6.84	6.37	6.27
20	PWM	-1.14	-0.94	-0.83	-0.65	-0.39	20	PWM	1.13	1.20	1.39	1.46	1.67
20	ML	-7.44	-5.73	-5.23	-5.15	-5.23	20	ML	3.69	3.42	3.27	3.06	2.90
50	PWM	-0.35	-0.29	-0.27	-0.20	-0.19	50	PWM	0.46	0.48	0.52	0.57	0.68
50	ML	-2.84	-1.98	-1.96	-2.06	-2.18	50	ML	1.60	1.43	1.29	1.17	1.12
100	PWM	-0.17	-0.12	-0.07	-0.13	-0.15	100	PWM	0.24	0.22	0.24	0.29	0.31
100	ML	-1.16	-0.86	-0.92	-1.05	-1.13	100	ML	0.88	0.74	0.64	0.59	0.53
200	PWM	-0.03	-0.04	-0.03	-0.05	-0.03	200	PWM	0.14	0.10	0.15	0.16	0.17
200	ML	-0.37	-0.39	-0.47	-0.52	-0.50	200	ML	0.50	0.39	0.36	0.31	0.27
$n_y$	Relative RMSE (%)					$n_y$	Relative RMSE (%)						
10	PWM	27.24	27.60	28.47	29.82	31.93	10	PWM	22.46	22.45	22.52	22.48	22.87
10	ML	36.24	35.18	34.79	34.58	35.04	10	ML	22.43	22.95	23.55	24.00	24.68
20	PWM	18.39	18.61	19.18	20.14	21.42	20	PWM	15.55	15.54	15.62	15.55	15.77
20	ML	23.61	21.36	20.42	20.38	20.96	20	ML	14.27	14.63	15.13	15.55	16.13
50	PWM	11.33	11.43	11.76	12.36	13.19	50	PWM	9.74	9.71	9.71	9.70	9.86
50	ML	13.97	11.87	11.54	11.90	12.43	50	ML	8.31	8.65	8.98	9.29	9.75
100	PWM	7.97	8.04	8.26	8.68	9.28	100	PWM	6.83	6.85	6.85	6.85	6.89
100	ML	9.32	7.97	7.95	8.24	8.66	100	ML	5.69	5.93	6.21	6.47	6.72
200	PWM	5.60	5.67	5.83	6.13	6.52	200	PWM	4.84	4.82	4.81	4.83	4.90
200	ML	6.31	5.56	5.58	5.79	6.05	200	ML	3.94	4.11	4.31	4.51	4.75

Table 4.4 Relative RMSE and bias of the ML and PWM estimators of  $\sigma$  and  $\tilde{\sigma}$ .

$n_y$	GEV					GPD				
10	4.63	3.51	2.72	2.14	1.63	0.22	0.05	0.02	0.01	0.00
20	1.64	0.78	0.39	0.17	0.10	0.00	0.00	0.00	0.00	0.00
50	0.98	0.41	0.10	0.02	0.01	0.00	0.00	0.00	0.00	0.00
100	1.24	0.44	0.08	0.01	0.00	0.00	0.00	0.00	0.00	0.00
200	1.38	0.38	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.5 Failure rate (%) of the ML method.



### 4.3 Effect of the scale and location parameters on the relative errors in the case of the GEV

To illustrate how the magnitudes of the errors in the estimation of percentiles vary with the choice of the scale and location parameters, we consider the case of yearly maxima of independent and identically (i.i.d.) three-hourly variables with a GEV distribution with scale parameter 1 and location parameter 0 (the parameters used in the previous section). Since the variables are i.i.d., the distribution of the yearly maxima is given by the product of the distribution of each three-hourly value in a year:

$$P(X_1 \leq x, \dots, X_n \leq x) = F(x)^n. \quad (4.1)$$

where  $n$  is the number of observations in a year. From (4.1) and (2.1), one can conclude that the yearly maxima of i.i.d. GEV variables is GEV distributed with the following parameters:

$$\xi_n = \xi, \sigma_n = \begin{cases} \sigma_n = \sigma, & \text{for } \xi = 0 \\ \sigma_n = \frac{\sigma}{n^{-\xi}}, & \text{for } \xi \neq 0 \end{cases} \text{ and } \mu_n = \begin{cases} \mu_n = \mu + \sigma \ln(n), & \text{for } \xi = 0 \\ \mu_n = \mu + \frac{\sigma}{\xi n^{-\xi}} - \frac{\sigma}{\xi}, & \text{for } \xi \neq 0 \end{cases}. \quad (4.2)$$

In the case of three-hourly time series,  $n = 2920$  and the parameters of the model governing the yearly maxima are as given in Table 4.6.

$\xi$	-0.30	-0.20	-0.10	0.00	0.10
$\sigma$	0.09	0.20	0.45	1.00	2.22
$\mu$	3.02	3.99	5.50	7.98	12.21

Table 4.6 Parameters of the GEV distribution of the yearly maxima of three-hourly i.i.d. GEV distributed variables with scale parameter 1, location parameter 0, and  $\xi$  as given in the first row of the table.

Without loss of generality, and for the sake of simplicity, only the GEV distribution is considered in this section. As in the simulation study presented in the previous section, for each GEV model given in Table 4.6 we carry out  $k = 100,000$  simulations and compute the error statistics of the GEV estimates. The results on the 4,000-yr and 10,000-yr return values, on the scale parameter and on the shape parameter are presented in Table 4.7 to Table 4.10.

Comparing the GEV error statistics given in Table 4.1 to Table 4.4 with those in Table 4.7 to Table 4.10, one can conclude that the errors in the shape and scale parameter estimates are comparable. In both cases the RMSE of the estimate of  $\xi$  decreases from about 0.5 to about 0.05 with increasing sample size, and the relative RMSE of the estimate of  $\sigma$  decreases from about 35% to about 6% with increasing sample size. Furthermore, in terms of variations with increasing sample sizes the error statistics allow the same conclusions that were drawn in the previous section.

On the other hand, the choice of different scale and location parameters influences the magnitude of the errors in the estimates of the percentiles. Thus, whereas in the previous section the relative RMSE of the 10,000-yr return value estimate based on a sample size of 50 years ranged from 26% to 140% depending on  $\xi$ , here it ranges from 2% to 97%; cf. Table 4.2 and Table 4.8.

As shown in Table 4.11, this combination of model parameters, especially the low values of the scale parameter, also yields more failures of the ML method; cf. Table 4.5 and Table 4.11.

4,000-yr rv		3.31	4.81	8.04	16.27	40.9
$\xi$		-0.30	-0.20	-0.10	0.00	0.10
$n_y$		Relative bias (%)				
10	PWM	2.91	6.94	15.42	28.30	44.14
10	ML	—*	—*	—*	—*	—*
20	PWM	0.99	2.47	5.64	11.53	19.83
20	ML	39.28	16.79	73.71	—*	—*
50	PWM	0.32	0.77	1.88	3.93	7.37
50	ML	-0.21	-0.08	0.89	4.52	12.55
100	PWM	0.14	0.36	0.89	2.02	3.77
100	ML	-0.44	-0.25	0.11	1.63	4.76
200	PWM	0.06	0.16	0.40	0.92	1.96
200	ML	-0.66	-0.29	-0.06	0.59	2.14
$n_y$		Relative RMSE (%)				
10	PWM	9.59	23.06	52.77	103.11	178.85
10	ML	—*	—*	—*	—*	—*
20	PWM	4.10	10.29	23.90	50.76	94.88
20	ML	—*	—*	—*	—*	—*
50	PWM	2.01	5.06	11.82	24.75	46.09
50	ML	2.01	4.94	14.46	31.15	67.04
100	PWM	1.33	3.31	7.79	16.22	29.50
100	ML	1.60	3.04	7.63	16.80	31.54
200	PWM	0.90	2.27	5.28	10.94	19.64
200	ML	1.48	2.14	4.97	10.74	19.24

Table 4.7 Relative RMSE and bias of the ML and PWM estimators of the 4,000-yr return value of the GEV.  
\*Values in excess of 1,000.

10,000-yr rv		3.31	4.84	8.21	17.19	45.79
$\xi$		-0.30	-0.20	-0.10	0.00	0.10
$n_y$		Relative bias (%)				
10	PWM	3.90	9.81	23.01	44.07	71.28
10	ML	—*	—*	—*	—*	—*
20	PWM	1.25	3.28	7.92	17.00	30.51
20	ML	201.43	58.60	317.14	—*	—*
50	PWM	0.39	1.01	2.58	5.64	10.92
50	ML	-0.15	0.08	1.57	6.79	18.34
100	PWM	0.18	0.47	1.21	2.84	5.48
100	ML	-0.42	-0.21	0.33	2.45	6.76
200	PWM	0.07	0.20	0.54	1.31	2.79
200	ML	-0.65	-0.29	0.02	0.92	3.00
$n_y$		Relative RMSE (%)				
10	PWM	12.93	32.38	77.54	156.47	280.01
10	ML	—*	—*	—*	—*	—*
20	PWM	4.84	12.73	30.79	68.68	134.42
20	ML	—*	—*	—*	—*	—*
50	PWM	2.23	5.85	14.18	30.74	59.40
50	ML	2.20	5.78	20.00	40.98	96.89
100	PWM	1.45	3.75	9.14	19.61	36.43
100	ML	1.67	3.42	8.93	20.41	40.09
200	PWM	0.98	2.55	6.13	13.01	23.69
200	ML	1.52	2.37	5.74	12.76	23.14

Table 4.8

Relative RMSE and bias of the ML and PWM estimators of the 10,000-yr return value of the GEV. \*Values in excess of 1,000.

		$\xi$	-0.30	-0.20	-0.10	0.00	0.10	
$n_y$		<b>Bias</b>						
<b>10</b>	<b>PWM</b>	-0.02	-0.03	-0.04	-0.05	-0.07		
<b>10</b>	<b>ML</b>	0.05	0.02	-0.05	-0.02	0.00		
<b>20</b>	<b>PWM</b>	-0.01	-0.01	-0.02	-0.02	-0.03		
<b>20</b>	<b>ML</b>	-0.01	-0.04	-0.04	-0.02	-0.01		
<b>50</b>	<b>PWM</b>	0.00	0.00	-0.01	-0.01	-0.01		
<b>50</b>	<b>ML</b>	0.00	-0.02	-0.02	-0.01	0.00		
<b>100</b>	<b>PWM</b>	0.00	0.00	0.00	0.00	-0.01		
<b>100</b>	<b>ML</b>	0.00	-0.01	-0.01	0.00	0.00		
<b>200</b>	<b>PWM</b>	0.00	0.00	0.00	0.00	0.00		
<b>200</b>	<b>ML</b>	0.00	-0.01	0.00	0.00	0.00		
$n_y$		<b>RMSE</b>						
<b>10</b>	<b>PWM</b>	0.30	0.29	0.29	0.29	0.30		
<b>10</b>	<b>ML</b>	0.37	0.39	0.49	0.50	0.51		
<b>20</b>	<b>PWM</b>	0.18	0.18	0.18	0.19	0.19		
<b>20</b>	<b>ML</b>	0.21	0.22	0.25	0.24	0.25		
<b>50</b>	<b>PWM</b>	0.11	0.11	0.11	0.11	0.12		
<b>50</b>	<b>ML</b>	0.10	0.11	0.12	0.12	0.12		
<b>100</b>	<b>PWM</b>	0.07	0.07	0.07	0.08	0.08		
<b>100</b>	<b>ML</b>	0.06	0.07	0.07	0.08	0.08		
<b>200</b>	<b>PWM</b>	0.05	0.05	0.05	0.05	0.06		
<b>200</b>	<b>ML</b>	0.04	0.05	0.05	0.05	0.06		

Table 4.9 RMSE and bias of the ML and PWM estimators of  $\xi$  in the GEV model.

$\sigma$		0.09	0.2	0.45	1	2.22
$\xi$		-0.30	-0.20	-0.10	0.00	0.10
$n_y$		Relative bias (%)				
10	PWM	-2.84	-2.47	-2.10	-1.58	-1.07
10	ML	-12.87	-11.57	-9.98	-11.06	-11.05
20	PWM	-1.14	-0.94	-0.83	-0.65	-0.39
20	ML	-5.11	-4.38	-4.96	-5.16	-5.20
50	PWM	-0.35	-0.29	-0.27	-0.20	-0.19
50	ML	-2.05	-1.67	-1.93	-2.06	-2.17
100	PWM	-0.17	-0.12	-0.07	-0.13	-0.15
100	ML	-0.71	-0.77	-0.92	-1.05	-1.13
200	PWM	-0.03	-0.04	-0.03	-0.05	-0.03
200	ML	0.24	-0.28	-0.47	-0.52	-0.50
$n_y$		Relative RMSE (%)				
10	PWM	27.24	27.60	28.47	29.82	31.93
10	ML	30.00	30.19	32.09	33.43	35.38
20	PWM	18.39	18.61	19.18	20.14	21.42
20	ML	19.47	19.13	19.77	20.40	21.09
50	PWM	11.33	11.43	11.76	12.36	13.19
50	ML	11.59	11.37	11.51	11.90	12.43
100	PWM	7.97	8.04	8.26	8.68	9.28
100	ML	8.13	7.90	7.95	8.24	8.66
200	PWM	5.60	5.67	5.83	6.13	6.52
200	ML	5.75	5.55	5.58	5.79	6.05

Table 4.10 Relative RMSE and bias of the ML and PWM estimators of  $\sigma$  in the GEV model.

$n_y$	GEV				
10	26.18	14.95	4.47	3.55	2.70
20	19.92	2.67	0.60	0.32	0.16
50	23.46	2.56	0.07	0.02	0.01
100	24.52	2.28	0.05	0.01	0.00
200	27.18	1.41	0.04	0.00	0.00

Table 4.11 Failure rate (%) of the ML estimation in the GEV model.

## 5 Finite-sample properties of the two methods with non-stationary and dependent data

### 5.1 Introduction

In this chapter we present the results comparing the finite sample properties of the AM/GEV and of the POT/GPD approaches when applied to simulated time series of dependent and non-stationary observations. As explained in Section 3.3, dependence is introduced by an AR(19) model (cf. Table 3.1) and non-stationarity is imposed by choosing the parameters, other than the tail index  $\xi$ , so as to force the marginal distribution functions  $F_t$  to have mean and variance appropriate to the month in which the measurement at time  $t$  falls (cf. Table 3.2). The simulations were, as in the study presented in the previous chapter, carried out for five sample sizes with different durations in years, namely  $n_y = 10, 20, 50, 100$  and  $200$  years. The PWM method was used in all simulations to estimate the model parameters. Given that the POT/GPD approach simulations are rather expensive, in each case we carry out only  $k = 100$  simulations.

In the next section we present the results of the AM/GEV approach applied to times series possessing tail indices of  $-0.3, -0.2, -0.1, 0$  and  $0.1$ . We also compare, in the case  $\xi=0$ , the results based on 100 simulations with those based on 1,000 simulations. Furthermore, we compare the finite-sample properties of the estimators of the GEV distribution based on i.i.d. data obtained in the preceding section with those of the AM/GEV approach with dependent data.

In Section 5.3 we present the results of the trio of models comprising the Gamma distribution, whose tail index is 0, the Beta distribution function with a tail index of  $-0.1$ , and the Beta distribution function of the second kind with a tail index of  $0.1$ . Because these simulations are much more expensive than those of the AM/GEV approach, the shape parameters of  $-0.2$  and  $-0.3$  are not considered.

### 5.2 Analysis of the results based on time series with marginal GEV distributions

Five simulation scenarios were considered, each with a given shape parameter. Given the complex form of the joint distributions of the non-stationary and dependent processes generated, their high percentiles cannot be obtained analytically. In order to accurately estimate them, 100,000-year long time series were simulated and estimates of the return values of interest obtained by the AM/GEV method. These estimates are given in Table 5.1. together with the empirical return values, i.e., the quantiles of probabilities  $1-1/4,000$  and  $1-1/10,000$  of the empirical distribution of the simulated sample. Note that for low tail indices the parameter estimates differ from the parameter values imposed in the simulations, the larger deviation occurring when  $\xi = -0.3$ , for which an estimate of  $-0.25$  is obtained. The error statistics reported next were all computed assuming that the parameters given in Table 5.1 and the associated percentiles are the true ones.

$\xi$	-0.25	-0.18	-0.10	0.00	0.10
$\sigma$	0.11	0.19	0.33	0.60	1.09
$\mu$	3.45	3.90	4.54	5.46	6.75
<b>4,000-yr rv</b>	3.82	4.71	6.44	10.44	20.79
<b>empirical 4,000-yr rv</b>	3.82	4.71	6.44	10.57	21.26
<b>10,000-yr rv</b>	3.83	4.74	6.57	10.99	23.18
<b>empirical 10,000-yr rv</b>	3.83	4.75	6.56	11.04	23.69

Table 5.1 Estimates of the parameters of the GEV model fitted to 100,000 year maxima.

Table 5.2 presents the error statistics of the POT/GPD and AM/GEV estimators of the 4,000-yr return value, of the 10,000-yr return value and of the tail index. In the case of the POT/GPD approach there is no such thing as a true scale parameter, since this depends on the threshold used; thus Table 5.2 also presents no error statistics on the scale parameter of the GEV.

The following conclusions can be drawn from the tables:

- The error statistics show that, as expected, the errors decrease with increasing sample size.
- Regarding the tail index, the POT/GPD approach yields more accurate estimates than the AM/GEV approach for series of less than 100 years. For series of 100 years the performance of the two approaches is comparable, AM/GEV being slightly better for lighter tails ( $\xi < 0$ ) and POT/GPD slightly better for heavier tails.
- In terms of the percentiles, only for time series of 200 years do both approaches provide comparably low error statistics, the relative RMSEs of the POT/GPD approach being about 2/3 of those of the AM/GEV approach for  $\xi > -0.1$ .
- For time series of less than 200 years the POT/GPD approach is always preferable.

We note that the threshold in the POT/GPD approach is chosen automatically. Figure 5.1 illustrates the choice of the threshold in a simulation with  $n_y = 200$  and  $\xi = 0$ . The figure shows the estimates of  $\xi$ ,  $\sigma^*$  and of the 10,000-yr return value as functions of the threshold. The automatically chosen threshold is indicated by the vertical lines. Although the automatic choice of the threshold seems to be appropriate, the POT/GPD approach is expected to perform even better if visual inspection and/or a formal method (e.g. Einmahl et al., 2009) is used to determine the threshold.

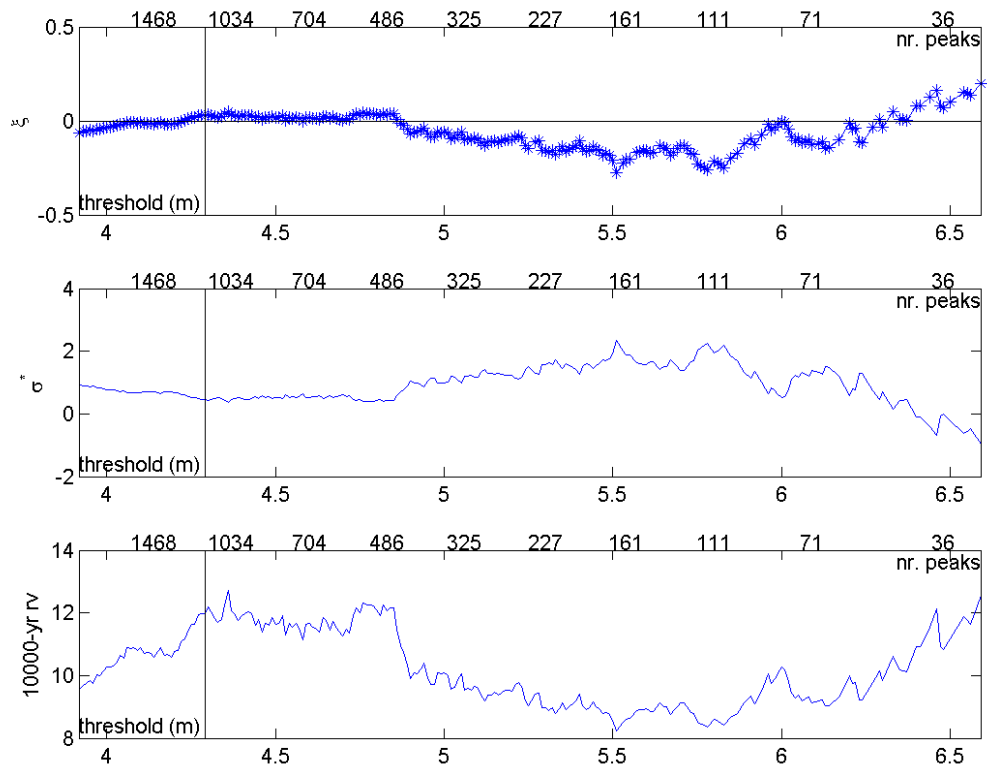


Figure 5.1 *Illustration of the variation of the estimates of the shape parameter, of  $\sigma^*$  and of the 10,000 return value with the threshold in the POT/GPD approach. The vertical line indicates the threshold chosen. The example shown is from a simulation with  $n_y = 200$  and  $\xi = 0$ .*



		4,000-yr rv					10,000-yr rv					$\xi$				
rv		3.82	4.71	6.44	10.44	20.79	3.83	4.74	6.57	10.99	23.18					
$\xi$		-0.25	-0.18	-0.10	0.00	0.10	-0.25	-0.18	-0.10	0.00	0.10	-0.25	-0.18	-0.10	0.00	0.10
$n_y$		Relative bias (%)					Relative bias (%)					Bias				
<b>10</b>	<b>POT</b>	-0.65	2.85	-0.20	1.13	11.06	-0.61	3.78	0.41	2.41	15.74	-0.11	-0.06	-0.08	-0.04	-0.02
<b>10</b>	<b>AM</b>	5.91	6.75	13.90	28.05	43.70	8.39	9.63	20.14	40.98	66.25	0.02	-0.03	-0.02	0.00	-0.02
<b>20</b>	<b>POT</b>	-1.27	-1.01	-1.48	2.42	4.04	-1.35	-1.00	-1.40	3.12	5.61	-0.10	-0.07	-0.05	-0.03	-0.03
<b>20</b>	<b>AM</b>	1.02	1.31	4.59	18.83	34.98	1.38	1.79	6.55	27.22	52.30	-0.01	-0.03	-0.03	0.01	-0.01
<b>50</b>	<b>POT</b>	-1.55	-2.80	-5.51	-2.68	-1.38	-1.65	-3.05	-6.08	-3.09	-1.43	-0.10	-0.09	-0.08	-0.04	-0.03
<b>50</b>	<b>AM</b>	0.19	0.01	-0.69	5.67	9.31	0.27	0.15	-0.39	7.58	13.19	-0.02	-0.02	-0.02	-0.01	-0.02
<b>100</b>	<b>POT</b>	-1.76	-2.35	-3.93	-5.89	-5.75	-1.89	-2.60	-4.43	-6.86	-6.74	-0.10	-0.07	-0.06	-0.05	-0.03
<b>100</b>	<b>AM</b>	0.19	0.99	1.69	2.86	3.77	0.24	1.20	2.14	3.83	5.34	0.00	0.00	0.01	0.00	-0.01
<b>200</b>	<b>POT</b>	-1.58	-2.43	-3.58	-5.52	-6.56	-1.69	-2.67	-4.09	-6.43	-7.77	-0.09	-0.07	-0.05	-0.04	-0.03
<b>200</b>	<b>AM</b>	0.10	-0.02	1.11	3.31	6.06	0.13	0.01	1.35	4.14	7.76	0.00	-0.01	0.00	0.01	0.01
$n_y$		Relative RMSE (%)					Relative RMSE (%)					RMSE				
<b>10</b>	<b>POT</b>	5.43	18.67	21.73	35.21	69.70	5.89	22.80	25.06	44.33	91.30	0.20	0.19	0.17	0.12	0.11
<b>10</b>	<b>AM</b>	19.60	22.23	39.58	73.66	127.36	30.87	30.53	53.15	104.50	182.86	0.28	0.28	0.29	0.25	0.25
<b>20</b>	<b>POT</b>	3.71	7.82	13.54	24.02	40.72	3.97	8.77	15.39	28.14	48.50	0.15	0.13	0.11	0.09	0.10
<b>20</b>	<b>AM</b>	5.14	8.15	20.74	66.11	123.47	6.19	9.50	25.73	94.68	180.78	0.19	0.18	0.21	0.17	0.20
<b>50</b>	<b>POT</b>	2.76	5.43	9.55	14.90	26.20	2.95	5.90	10.57	16.95	31.27	0.13	0.12	0.11	0.08	0.06
<b>50</b>	<b>AM</b>	2.68	5.06	10.11	25.59	51.57	3.01	5.84	11.97	31.98	67.64	0.12	0.11	0.11	0.11	0.12
<b>100</b>	<b>POT</b>	2.61	3.99	7.27	11.39	18.01	2.79	4.35	8.08	12.93	20.57	0.13	0.10	0.08	0.06	0.06
<b>100</b>	<b>AM</b>	1.69	4.24	7.79	16.28	27.78	1.89	4.95	9.28	19.68	33.80	0.08	0.08	0.07	0.08	0.08
<b>200</b>	<b>POT</b>	2.10	3.68	5.66	9.95	14.15	2.25	4.04	6.35	11.35	16.16	0.11	0.09	0.07	0.06	0.05
<b>200</b>	<b>AM</b>	1.21	2.26	5.43	11.03	21.10	1.35	2.54	6.34	13.23	25.77	0.05	0.05	0.05	0.05	0.06

Table 5.2 RMSE and bias of the PWM parameter estimates in the AM/GEV and POT/GPD approaches, computed from 100 simulations

## 5.2.1 Effect of the number of simulations on the error statistics

In order to get an idea of how the number of simulations affects the error statistics we have also carried out 1,000 simulation in the case  $\xi=0$ . Table 5.3 compares the error statistics obtained with 100 and 1,000 simulations. As the table shows, for short time series, of less than 50 years, the error statistics can vary significantly. However, the conclusions regarding the relative performance of the two approaches remain the same.

		4,000-yr rv		10,000-yr rv		$\xi$	
		100	1000	100	1000	100	1000
	rv	10.44		10.99			
	$\xi$	0.00		0.00		0.00	
		Relative bias (%)		Relative bias (%)		Bias	
<b>10</b>	<b>POT</b>	1.13	7.23	2.41	9.94	-0.04	-0.03
<b>10</b>	<b>AM</b>	28.05	29.31	40.98	46.13	0.00	-0.04
<b>20</b>	<b>POT</b>	2.42	1.35	3.12	2.15	-0.03	-0.03
<b>20</b>	<b>AM</b>	18.83	12.54	27.22	18.47	0.01	-0.02
<b>50</b>	<b>POT</b>	-2.68	-4.03	-3.09	-4.59	-0.04	-0.04
<b>50</b>	<b>AM</b>	5.67	4.45	7.58	6.09	-0.01	-0.01
<b>100</b>	<b>POT</b>	-5.89	-6.26	-6.86	-7.25	-0.05	-0.05
<b>100</b>	<b>AM</b>	2.86	2.19	3.83	3.02	0.00	-0.01
<b>200</b>	<b>POT</b>	-5.52	-5.90	-6.43	-6.88	-0.04	-0.04
<b>200</b>	<b>AM</b>	3.31	0.80	4.14	1.14	0.01	0.00
		Relative RMSE (%)		Relative RMSE (%)		RMSE	
<b>10</b>	<b>POT</b>	35.21	44.18	44.33	55.22	0.12	0.14
<b>10</b>	<b>AM</b>	73.66	103.55	104.50	161.97	0.25	0.29
<b>20</b>	<b>POT</b>	24.02	26.62	28.14	31.69	0.09	0.11
<b>20</b>	<b>AM</b>	66.11	54.32	94.68	76.10	0.17	0.18
<b>50</b>	<b>POT</b>	14.90	14.47	16.95	16.59	0.08	0.07
<b>50</b>	<b>AM</b>	25.59	22.44	31.98	27.65	0.11	0.11
<b>100</b>	<b>POT</b>	11.39	11.43	12.93	13.00	0.06	0.06
<b>100</b>	<b>AM</b>	16.28	16.04	19.68	19.42	0.08	0.08
<b>200</b>	<b>POT</b>	9.95	9.76	11.35	11.18	0.06	0.06
<b>200</b>	<b>AM</b>	11.03	10.33	13.23	12.34	0.05	0.05

Table 5.3

RMSE and bias of the PWM parameter estimators of the AM/GEV and POT/GPD approaches computed from 100 and 1,000 simulations. Results only for the type I tail.

5.2.2 Comparison between the performance of the estimators with i.i.d. and non-i.i.d. data

Table 5.4 shows error statistics of the 4,000- and 10,000 -yr return value estimates and of the tail index estimate of the GEV distribution based on 100,000 simulations of i.i.d. time series of varying lengths. These simulations are analogous to those of Section 4.3, except that the parameters are as given in Table 5.1 and only PWM estimators are considered. The results are to be compared with the results obtained with the AM/GEV approach based on the same parameters but with non-i.i.d. time series (Table 5.2).

Taking into account the variation of the error statistics with the number of simulations, it can be concluded that the accuracies of the parameter estimates of the GEV model with i.i.d. and non-i.i.d. data are comparable with each other (cf. Table 5.2 and Table 5.4) and in line with the error statistics presented in the previous chapter (cf. Table 4.3, Table 4.9, Table 5.2 and Table 5.4). This is in contrast with the performance of the POT/GPD approach, which is worse with non-i.i.d. data than with i.i.d. data (cf. Table 4.3 and Table 5.2); this is to be expected given the influence of the threshold (which must be chosen here, but not in the simulations of Chapter 4) on the accuracy of the estimates.

	4,000-yr rv					10,000-yr rv					$\xi$				
rv	3.82	4.71	6.44	10.44	20.79	3.83	4.74	6.57	10.99	23.18					
$\xi$	-0.25	-0.18	-0.10	0.00	0.10	-0.25	-0.18	-0.10	0.00	0.10	-0.25	-0.18	-0.10	0.00	0.10
$n_y$	Relative bias (%)					Relative bias (%)					Bias				
10	3.63	7.18	14.44	26.51	42.50	5.01	10.28	21.62	41.42	68.90	-0.02	-0.03	-0.04	-0.05	-0.07
20	1.26	2.58	5.30	10.80	19.10	1.63	3.48	7.46	15.98	29.49	-0.01	-0.01	-0.02	-0.02	-0.03
50	0.41	0.81	1.77	3.69	7.09	0.52	1.07	2.43	5.30	10.55	0.00	-0.01	-0.01	-0.01	-0.01
100	0.18	0.38	0.84	1.89	3.63	0.23	0.50	1.14	2.67	5.29	0.00	0.00	0.00	0.00	-0.01
200	0.08	0.17	0.38	0.86	1.88	0.10	0.22	0.51	1.23	2.70	0.00	0.00	0.00	0.00	0.00
$n_y$	Relative RMSE (%)					Relative RMSE (%)					RMSE				
10	12.08	23.98	49.50	96.58	172.19	16.64	34.01	72.95	147.06	270.66	0.29	0.29	0.29	0.29	0.30
20	5.26	10.79	22.45	47.55	91.35	6.34	13.50	29.02	64.54	129.93	0.18	0.18	0.18	0.19	0.19
50	2.59	5.32	11.10	23.18	44.37	2.94	6.20	13.35	28.89	57.42	0.11	0.11	0.11	0.11	0.12
100	1.71	3.48	7.32	15.20	28.40	1.91	3.98	8.61	18.43	35.21	0.07	0.07	0.07	0.08	0.08
200	1.16	2.38	4.96	10.24	18.91	1.28	2.70	5.77	12.23	22.90	0.05	0.05	0.05	0.05	0.06

Table 5.4 Error statistics of the PWM estimators of the GEV distribution with i.i.d. data when the parameters correspond to those of the non-i.i.d. data given in Table 5.1. Based on 100,000 simulations.

5.3 Analysis of the results based on time series with marginal Gamma or Beta distributions

In this simulation three scenarios are considered, each with a given tail index. Again, given the complex form of the simulated processes, their high percentiles cannot be obtained analytically and are estimated by applying the AM/GEV approach to a 100,000-year long simulated time series. The estimates are given in Table 5.5.

Figure 5.2 illustrates the automatic choice of the threshold in one of the simulations with  $n_y = 100$  and  $\xi = -0.1$  (marginal Beta distribution functions). The figure shows the estimates of  $\xi$ ,  $\sigma^*$  and of the 10,000-yr return value as functions of the threshold, and indicates the chosen threshold by vertical lines.

$\xi$	-0.11	-0.02	0.10
$\sigma$	0.30	0.53	1.01
$\mu$	4.52	5.27	6.53
<b>4,000-yr rv</b>	<b>6.14</b>	<b>9.38</b>	<b>19.27</b>
<b>empirical 4,000-yr rv</b>	<b>6.18</b>	<b>9.52</b>	<b>18.68</b>
<b>10,000-yr rv</b>	<b>6.25</b>	<b>9.81</b>	<b>21.40</b>
<b>empirical 10,000-yr rv</b>	<b>6.29</b>	<b>10.02</b>	<b>21.47</b>

Table 5.5 Estimates of the parameters of the GEV distribution fitted to 100,000 yearly maxima of time series with Gamma ( $\xi=0$ ) Beta ( $\xi=-0.1$ ) and Beta of the second kind ( $\xi=0.1$ ) marginal distributions.

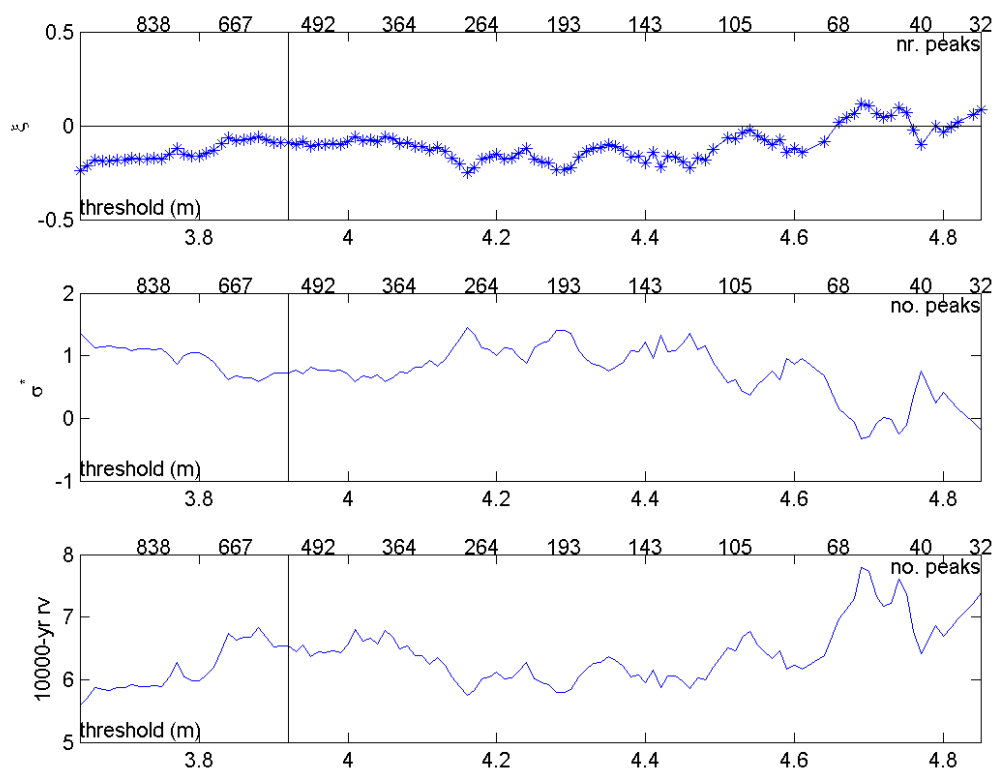


Figure 5.2 Illustration of the variation of the tail index,  $\sigma^*$  and 10,000 return value estimates with the threshold in the POT/GPD approach. The vertical line indicates the automatically chosen threshold. The example shown is from a simulation with  $n_y=100$  and  $\xi=-0.1$ .

Table 5.6 presents the error statistics of the POT/GPD and AM/GEV estimates of the 4,000- and 10,000-year return values and of the tail index. The conclusions of the preceding section concerning the time series with GEV marginal distributions also apply here. However, in this case the GEV/AM approach compares slightly worse with the POT/GPD approach. This is not surprising in view of the fact that the annual maxima from a process with marginal GEV distributions with the same tail index are closer to a limiting GEV distribution of extremes than the annual maxima from a process with Gamma or Beta marginal distributions—and, in fact, it was with the purpose of illustrating this phenomenon that we have chosen marginal distributions other than the GEV.

		4,000-yr rv			10,000-yr rv			$\xi$			
		rv	6.14	9.38	19.27	6.25	9.81	21.40		$\xi$	
		$\xi$	-0.11	-0.02	0.10	-0.11	-0.02	0.10	-0.11	-0.02	0.10
$n_y$		Relative bias (%)			Relative bias (%)			Bias			
10	POT	-0.02	2.72	7.74	0.37	4.29	11.43	-0.06	-0.05	-0.05	
10	AM	17.72	18.73	63.57	25.53	30.16	101.44	0.01	-0.08	-0.06	
20	POT	-1.56	-5.25	-2.67	-1.68	-5.69	-2.36	-0.06	-0.06	-0.04	
20	AM	5.91	5.31	20.78	8.13	8.89	31.22	0.00	-0.03	-0.03	
50	POT	-2.88	-8.09	-7.30	-3.20	-9.14	-8.34	-0.06	-0.07	-0.04	
50	AM	2.80	-0.11	1.80	3.46	0.45	4.11	0.01	-0.02	-0.04	
100	POT	-3.97	-6.74	-7.77	-4.47	-7.75	-9.07	-0.06	-0.05	-0.04	
100	AM	0.80	0.29	6.11	1.08	0.69	8.24	0.00	-0.01	0.00	
200	POT	-3.61	-4.81	-9.94	-4.09	-5.66	-11.68	-0.06	-0.04	-0.04	
200	AM	0.59	1.81	3.19	0.74	2.25	4.28	0.00	0.00	0.00	
$n_y$		Relative RMSE (%)			Relative RMSE (%)			RMSE			
10	POT	14.18	33.13	61.61	16.58	41.06	75.59	0.15	0.15	0.15	
10	AM	47.26	72.91	195.11	68.24	107.78	303.64	0.26	0.31	0.34	
20	POT	11.16	17.44	31.75	12.48	19.83	37.02	0.11	0.11	0.10	
20	AM	20.56	37.09	83.76	26.44	49.54	114.19	0.17	0.17	0.19	
50	POT	8.08	13.78	22.22	9.05	15.45	25.40	0.10	0.09	0.07	
50	AM	8.90	15.11	42.03	10.64	18.38	53.35	0.09	0.09	0.13	
100	POT	6.28	10.05	15.85	7.02	11.38	18.10	0.08	0.07	0.06	
100	AM	6.94	11.48	28.02	8.13	13.79	34.45	0.07	0.07	0.08	
200	POT	5.34	7.81	13.63	5.95	8.95	15.69	0.07	0.05	0.05	
200	AM	4.27	10.33	18.57	4.92	12.33	22.59	0.05	0.06	0.06	

Table 5.6 RMSE and bias of the PWM estimates of the 4,000- and 10,000-yr return value and of  $\xi$  in the AM/GEV and POT/GPD approaches applied to non-stationary and dependent time series with Gamma, Beta and Beta of the second kind marginal distributions. Computed from 100 simulations.

Table 5.7 shows the same statistics as Table 5.6 but for the ML estimates instead of PWM estimates. Comparing the two tables we can say that the conclusions drawn in Chapter 4 concerning the relative performance of the two estimation methods apply here as well.

		4,000-yr rv			10,000-yr rv			$\xi$		
		rv								
		$\xi$								
$n_y$		Relative bias (%)			Relative bias (%)			Bias		
10	POT	-4.04	-5.78	1.19	-4.25	-5.88	4.02	-0.10	-0.10	-0.08
10	AM	—*	—*	—*	—*	—*	—*	—*	—*	—*
20	POT	-4.19	-5.93	-4.15	-4.67	-6.50	-3.54	-0.08	-0.07	-0.06
20	AM	14.58	25.08	290.14	24.53	42.49	674.08	-0.02	-0.03	-0.04
50	POT	-4.11	-6.53	-9.74	-4.65	-7.52	-11.26	-0.07	-0.06	-0.05
50	AM	-1.48	-0.69	4.33	-1.40	0.07	8.17	-0.04	-0.03	-0.03
100	POT	-2.56	-4.55	-8.08	-2.98	-5.38	-9.52	-0.05	-0.04	-0.04
100	AM	0.62	2.85	6.27	0.87	3.83	8.47	-0.00	0.00	0.00
200	POT	-1.97	-3.62	-7.08	-2.31	-4.32	-8.43	-0.04	-0.03	-0.03
200	AM	0.17	1.15	2.03	0.26	1.55	2.88	-0.00	-0.00	-0.00
$n_y$		Relative RMSE (%)			Relative RMSE (%)			RMSE		
10	POT	13.62	27.82	63.93	15.15	32.04	78.51	0.17	0.17	0.17
10	AM	—*	—*	—*	—*	—*	—*	—*	—*	—*
20	POT	8.95	19.95	46.08	9.80	22.75	55.73	0.11	0.12	0.12
20	AM	94.64	156.53	2444.40	164.12	273.10	5977.56	0.22	0.24	0.29
50	POT	6.37	12.23	19.92	7.04	13.81	22.53	0.09	0.08	0.07
50	AM	8.62	21.32	56.48	9.97	26.45	76.44	0.12	0.12	0.13
100	POT	4.40	8.18	14.91	4.93	9.33	17.07	0.06	0.05	0.05
100	AM	6.06	14.30	27.23	7.08	17.40	33.60	0.07	0.08	0.08
200	POT	3.60	6.78	11.79	4.10	7.81	13.64	0.05	0.05	0.04
200	AM	4.23	9.61	18.72	4.90	11.48	22.72	0.05	0.05	0.06

Table 5.7 RMSE and bias of the ML estimates of the 4,000- and 10,000-yr return value and of  $\xi$  in the AM/GEV and POT/GPD approaches applied to non-stationary and dependent time series with Gamma, Beta and Beta of the second kind marginal distributions. Computed from 100 simulations. \* No estimate available due to a high failure rate.

## 6 Conclusions

The objective of this study was to provide a global assessment and comparison of the AM/GEV and POT/GPD approaches on the basis of small to large time series of simulated data possessing realistic features such as non-stationarity and serial dependence.

Two types of simulation studies were carried out, both taking into account the characteristics of the data on the basic variables currently available. The first study focused on the finite-sample properties of the estimators of the GEV and the GPD distributions. The second study focused on the finite-sample properties of the AM/GEV and the POT/GPD approaches applied to non-stationary and dependent data.

The length of the simulated time series ranged from 10 to 200 years, and the tail indices ranged from -0.3 to 0.1. These were found to represent the availability of the data used in the determination of the HBC as well as their tail characteristics.

In the first study both the ML and PWM methods of estimation were considered. The conclusions of the study were that for POT samples with an average of two or more observations per year, the estimators of the GPD are more accurate than the corresponding estimators of GEV distribution, and that with more than 200 years of data the errors of the estimates of the GEV model are low. Furthermore, it was found that with less than 50 years of data the PWM method, thanks to its error characteristics and robustness against nonexistence of solutions, is preferable to the ML method, and that with bigger data sets the two methods provide comparably accurate estimates.

In the second study the simulated data incorporated non-stationarity and serial dependence. The results showed that the POT/GPD estimates of the shape parameter are more accurate than those of the AM/GEV approach for time series with less than 100 years. With 100-year long time series the performance of the two approaches is comparable, AM/GEV being slightly better with lighter tails (negative tail index) and POT/GPD being slightly better with the heavier tails. In terms of the 4,000- and 10,000-year return value estimates, the POT/GPD approach is significantly better. Only for time series of 200 years do the two approaches provide comparably low error statistics. Still, even with 200-year long time series with a tail index greater than -0.1 the relative RMSEs of the POT/GPD approach are about 2/3 of those of the AM/GEV approach. We note that in our simulations the threshold in the POT/GPD approach was chosen automatically; if visual inspection or a formal method is used to choose the threshold the POT/GPD approach is expected to perform even better.

Based on the results of this study, we recommend that the POT/GPD approach be used for the extreme value analyses required for the determination of HBC, irrespective of the basic variable being considered. Furthermore, we recommend that the GPD be estimated with the method of PWM.

## References

- Caires, S., 2007: *Extreme wave statistics. Methodology and applications to North Sea wave data*. WL | Delft Hydraulics Report H4803.30.
- Caires, S., 2009: *Extreme wind statistics for the Hydraulic Boundary Conditions for the Dutch primary water defences*. SBW-Belastingen: Phase 2 of subproject 'Wind modelling'. Deltares Report 1200264-005.
- Coles, S., 2001: *An Introduction to the Statistical Modelling of Extreme Values*. Springer Texts in Statistics, Springer-Verlag: London.
- Cunnane, C., 1973: A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *J. Hydrol.*, 18, pp. 257–271.
- de Haan, L. and A. Ferreira, 2006: *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering, Springer, New York.
- de Haan, L. and C. Zhou, 2009: Notes for the discussion Deltares 15/10/08. *Evaluation of the statistical methods currently used to determine boundary conditions: Workshop proceedings* (Deltares Report H5098.70, Eds. S. Caires and D. Dillingh), 38-40.
- de Valk, C., 1993: *Estimation of marginals from measurements and hindcast data*. WL | Delft Hydraulics Report H1700.
- de Waal, H., E. H. Chbab and J. Groeneweg, 2009: *ACTION PLAN SBW-Belastingen*. Deltares Report H5098.30.
- Dillingh, D., L. de Haan, R. Helmers, G.P. Können and J. van Malde, 1993: De basispeilen langs de Nederlandse kust; statistisch onderzoek (In Dutch). Rijkswaterstaat, Dienst Getijdenwateren /RIKZ, Report DGW-93.023.
- Einmahl, J.H.J., J. Li and R.Y. Liu, 2009: Thresholding events of extreme in simultaneous monitoring of multiple risks, *J. of the American Statistical Association*, 104 (487), 982-992 (with a supplement available in <http://center.uvt.nl/staff/einmahl/AppELL.pdf>).
- Embrechts, P., C. Klüppelberg, and T. Mikosch, 1997: *Modelling Extremal Events for Insurance and Finance*. Springer.
- Guedes Soares, C., A.M. Ferreira, C. Cunha, 1996: Linear models of the time series of significant wave height on the southwest coast of Portugal. *Coastal Eng.*, 1–19.
- 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.
- Tavares, L.V. and J. E. da Silva, 1983: Partial duration series method revisited. *J. Hydrol.*, 64, 1-14.
- van den Brink, H.W., G.P. Können and J.D. Opsteegh, 2005: Uncertainties in extreme surge level estimates from observational records. *Phil. Trans. R. Soc. A*, 363, 1377-1386, doi: 10.1098/rsta.2005.1573.
- Yevjevich, V. and V. Taesombut, 1978: Information on flood peaks in daily flow series. *Proc. Int. Symp. on Risk and Reliability in Water Resources*, University of Waterloo, Waterloo.



## A Further details on the simulation of the time series

### A.1 GEV distribution

The expressions for the mean and variance of a GEV random variable are given for example in Embrechts et al. (1997). From these expressions it is seen that the scale and location parameters that yield a GEV model with monthly variance  $S_i^2$  and mean  $M_i$  are given by

$$\sigma_i = \sqrt{\frac{S_i^2 \xi^2}{\Gamma(1-2\xi) - 2\Gamma(1-\xi)}} \quad (1.1)$$

and

$$\mu_i = M_i - \frac{\sigma_i}{\xi} \Gamma(1-\xi). \quad (1.2)$$

The  $U_i$  variable falling in month  $i$  can then be transformed into a GEV variable with shape parameter  $\xi$ , monthly variance  $S_i^2$  and mean  $M_i$  by

$$V_i = G^{-1}(U_i; \xi, \mu_i, \sigma_i), \quad (1.3)$$

where  $G^{-1}(\cdot; \mu, \sigma)$  is the inverse of the GEV distribution function defined in Eq. (2.1).

### A.2 Gamma distribution

The *Gamma distribution* has density function given by

$$f_1(x; \alpha, \beta) = \frac{\alpha^\beta x^{\beta-1} e^{-\alpha x}}{\Gamma(\beta)}, \quad x > 0, \quad (1.4)$$

where  $\alpha, \beta > 0$  and  $\Gamma(\beta) = \int_0^\infty y^{\beta-1} e^{-y} dy$  is the *Gamma function*. The subscript 1 in  $f_1$  is a reminder that this model has a type I tail.

It can be seen from the expressions of the mean and variance of  $f_1$  that in order that the Gamma distribution have variance  $S_i^2$  and mean  $M_i$  the scale and shape parameters  $\alpha$  and  $\beta$  must be set equal to

$$\alpha_i = \frac{S_i^2}{M_i} \quad (1.5)$$

and

$$\beta_i = \frac{M_i^2}{S_i^2}. \quad (1.6)$$

The  $U_i$  data falling in month  $i$  can be transformed into a Gamma variable with monthly variance  $S_i^2$  and mean  $M_i$  by

$$V_i = F_1^{-1}(U_i; \alpha_i, \beta_i), \quad (1.7)$$

where  $F_1^{-1}(\cdot; \alpha_i, \beta_i)$  is the inverse of the Gamma distribution function with parameters  $\alpha_i$  and  $\beta_i$ , given by

$$F_1(x; \alpha_i, \beta_i) = \int_0^x f_1(y; \alpha_i, \beta_i) dy, \quad x \geq 0. \quad (1.8)$$

### A.3 Beta distribution of the first kind

The Beta distribution of the first kind with a scale parameter  $\alpha > 0$  and shape parameters  $\beta > 0$  and  $\xi < 0$  has density function given by

$$f_3(x; \alpha, \beta, \xi) = \frac{(-\xi\alpha)(-\xi\alpha x)^{\beta-1}(1+\xi\alpha x)^{-1/\xi-1}}{B(-\xi^{-1}, \beta)}, \quad 0 < x < -(\xi\alpha)^{-1}, \quad (1.9)$$

where  $B(a, b) = \int_0^1 y^{a-1}(1-y)^{b-1} dy$ ,  $a, b > 0$ , is the *Beta function*. As the subscript in  $f_3$  indicates, this model has a type III tail. Indeed, the distribution function of the excesses from this Beta distribution above a threshold  $t$  and normalized (i.e. divided) by  $g(t) = 1 + \xi\alpha t$ , which is given by

$$1 - \frac{1 - F_3(t + xg(t))}{1 - F_3(t)}, \quad (1.10)$$

where  $F_3(x) \equiv F_3(x; \alpha, \beta, \xi) = \int_0^x f_3(y; \alpha, \beta, \xi) dy$  for  $0 \leq x \leq -(\alpha\xi)^{-1}$ , converges, as  $t \rightarrow -(\alpha\xi)^{-1}$  from the left, to  $1 - (1 + \xi\alpha x)^{-1/\xi}$ , which is the distribution function of a GPD with shape parameter  $\xi < 0$ .

Moreover, as  $\xi \rightarrow 0$  from the left  $f_3(\cdot; \alpha, \beta, \xi)$  approaches the Gamma density  $f_1(\cdot; \alpha, \beta)$ , just as the GPD approaches the exponential distribution and the GEV approaches the Gumbel distribution.

From the expressions of the mean and variance of  $f_3$  it is seen that in order that the Beta distribution with fixed parameter  $\xi$  have variance  $S_i^2$  and mean  $M_i$  its scale and shape parameters  $\alpha$  and  $\beta$  must be set equal to

$$\beta_i = 0.5 \left[ \left( \frac{1}{\xi} - 1 \right) + \sqrt{\left( \frac{1}{\xi} - 1 \right)^2 - \frac{4M_i^2}{\xi S_i^2}} \right], \text{ and} \quad (1.11)$$

$$\alpha_i = -\frac{\beta_i}{\xi(\beta_i - \xi^{-1})} M_i. \quad (1.12)$$

The  $U_i$  data falling in month  $i$  can thus be transformed into a Beta variable with a negative shape parameter  $\xi$ , and monthly variance  $S_i^2$  and mean  $M_i$  by

$$V_i = F_3^{-1}(U_i; \alpha_i, \beta_i, \xi), \quad (1.13)$$

where  $F_3^{-1}(\cdot; \alpha_i, \beta_i, \xi)$  is the inverse of  $F_3(\cdot; \alpha_i, \beta_i, \xi)$ , already defined above.

For completeness, let us give an expression for  $F_3^{-1}(\cdot; \alpha, \beta, \xi)$  that is useful in actual computations. The *incomplete Beta function* is a function of  $x$  defined on the interval  $[0, 1]$  by

$$B_x(a, b) = \int_0^x y^{a-1} (1-y)^{b-1} dy \quad (1.14)$$

for  $a > 0$  and  $b > 0$ . Clearly,  $B_1(a, b) = B(a, b)$ , i.e., the incomplete Beta function at  $x=1$  equals the ('complete') Beta function. For convenience, instead of  $B_x(a, b)$  we shall consider in what follows its normalized version, namely  $\bar{B}_x(a, b) = B_x(a, b) / B(a, b)$ , which is the Beta distribution function with parameters  $a$  and  $b$ . The numerical computation of  $\bar{B}_x(a, b)$  is available in most mathematical software packages, including MATLAB. Its inverse function, defined for all  $y$  in the interval  $[0, \infty)$  and denoted by  $\bar{B}_y^{-1}(a, b)$ , is also available in such packages, so it is useful to know that  $F_3^{-1}(\cdot; \alpha, \beta, \xi)$  can be expressed in terms of it as

$$F_3^{-1}(u; \alpha, \beta, \xi) = (-\xi\alpha)^{-1} \bar{B}_u^{-1}(\beta, -1/\xi) \text{ for } 0 \leq u \leq 1. \quad (1.15)$$

#### A.4 Beta distribution function of the second kind with a scale parameter

The Beta distribution of the second kind with a scale parameter  $\alpha > 0$  and shape parameters  $\beta > 0$  and  $\xi > 0$  has density function given by

$$f_2(x; \alpha, \beta, \xi) = \frac{\xi\alpha}{B(-\xi^{-1}, \beta)} \frac{(\xi\alpha x)^{\beta-1}}{(1 + \xi\alpha x)^{1/\xi + \beta}}, \quad x > 0, \quad (1.16)$$

The notation indicates that this model has a type II tail: the distribution function of the excesses from this distribution above a threshold  $t$  and normalized by  $g(t) = \xi\alpha t$ , namely

$$1 - \frac{1 - F_2(t + xg(t))}{1 - F_2(t)}, \quad (1.17)$$

where  $F_2(x) \equiv F_2(x; \alpha, \beta, \xi) = \int_0^x f_2(y; \alpha, \beta, \xi) dy$  for  $x \geq 0$ , converges, as  $t \rightarrow \infty$  to  $1 - (1 + \xi \alpha x)^{-1/\xi}$ , which is the distribution function of a GPD with shape parameter  $\xi > 0$ .

Moreover, as  $\xi \rightarrow 0$  from the right  $f_2(\cdot; \alpha, \beta, \xi)$  approaches the Gamma density  $f_1(\cdot; \alpha, \beta)$ , just as the GPD approaches the exponential distribution and the GEV approaches the Gumbel distribution.

From the expressions of the mean and variance of  $f_3$  it is seen that in order that the Beta distribution of the second kind with fixed parameter  $\xi$  have variance  $S_i^2$  and mean  $M_i$  its scale and shape parameters  $\alpha$  and  $\beta$  must be set equal to

$$\beta_i = \left( \frac{S_i^2 + M_i^2}{M_i^2} \frac{1 - 2\xi}{1 - \xi} - 1 \right)^{-1} \quad (1.18)$$

and

$$\alpha_i = - \frac{\beta_i}{M_i(1 - \xi)} \quad (1.19)$$

The  $U_i$  data falling in month  $i$  can be transformed into a Beta variable of the second kind with a positive shape parameter  $\xi$ , monthly variance  $S_i^2$  and mean  $M_i$  by

$$V_t = F_2^{-1}(U_t; \alpha_i, \beta_i, \xi), \quad (1.20)$$

where  $F_2^{-1}(\cdot; \alpha_i, \beta_i, \xi)$  is the inverse of  $F_2(\cdot; \alpha_i, \beta_i, \xi)$ . As in Section A.3 above, it is convenient for purposes of numerical computations to know that this last function can be expressed in terms of the inverse of the normalized incomplete Beta function  $\bar{B}_x(a, b)$  as

$$F_2^{-1}(u; \alpha, \beta, \xi) = (\xi \alpha)^{-1} \left( \frac{1}{\bar{B}_{1-u}^{-1}(1/\xi, \beta)} - 1 \right) \text{ for } 0 \leq u \leq 1. \quad (1.21)$$

## **B Report review**

### **B.1 Introduction**

Prof. Laurens de Haan was asked to review this report according to the following guidelines:

The reviewer must write an English review covering the following aspects:

- Are the background and objectives given in the report clear and unequivocal?
- Is the chosen approach suitable to fulfil the objectives?
- Are the objectives met?
- Does the report provide the desired insight?
- Are the results reasonably concrete and applicable?

### **B.2 Review report**

Review of  
A comparative simulation study  
of the annual maxima and peaks-over-threshold methods  
by Sofia Caires

**OVERVIEW**

The aim of the report is to make a recommendation - based on computer simulations - on the subject whether to choose the “annual maxima” or the “peaks over threshold” approach in problems concerning extreme values for the “basic variables”. A second aim is to give a guideline whether to use the “maximum likelihood” or the “probability weighted moment” estimators in the endeavour.

I shall make some comments about the various sections of the report and some general comments at the end.

§2: Extreme value analysis

The account of extreme value theory is quite good. It is concerned only with *GEV* and *GPD* distribution as such. The nice feature of extreme value theory is that it can be applied to many other distributions, the ones that are in the domain of attraction of *GEV* (or equivalently *GPD*) distribution. The tail of such a distribution resembles the one of the limit distribution and this fact is sufficient for applying similar methods. Since this is the situation encountered in applications and since this is also used in the present report (section 5 and appendix), I feel that at least some information should be included in this section. It can be found in several books.

I could not understand the second half of page 6 and the first 3 lines of the page 7. There is a somewhat similar text in the book of Coles (Chapter 4) that I do not understand either.

§3: Setup of the study

I understand that the decision to carry out one and the same simulation program for all basic variables is motivated by the fact that they all depend on the same weather or climate conditions.

The choices made for the simulation program and explained in this section seem very sensible in relation to the characteristics of the “basic variables”. I understand that the lag 19 has been chosen for the AR process since it corresponds more or less to independence for observations 3 days apart (if I am not mistaken). By the way, an AR process can have many possible marginal distributions, the normal distribution is a choice not a consequence. I suppose that a stationary version of the AR process has been taken.

In general *EVT* statistical procedures apply without change if some dependence between the observations is allowed as it is the case here. Non-stationarity seems a bigger problem not the least because one does not know what one is trying to estimate anymore. However in the present case the goal is clear. Then the problem is whether the non-stationarity within a year affects the extreme value approximation. Not much is known about this. I do not expect big problems. It would be helpful in relation to table 3.2 to mention the average number of peaks for each month. Probably the winter months dominate anyway. Is  $\xi$  really expected to be the same for all months?

#### § 4 : Finite-sample properties of the ML and PWM estimators of the GEV and GPD models

The comparison between two models and two estimators has been done under very clean “laboratory” conditions: a *GPD* distribution for the *GPD* method and a *GEV* distribution for the *GEV* method. This is certainly valuable and enlightening. In reality the underlying distribution is such that only its tail resembles that of the *GEV* and *GPD* distributions. In that case the annual maximum distribution may be better approximated by the *GEV* distribution than the peak distribution is approximated by the *GPD* distribution counterweighing somewhat the advantage of a bigger sample size for the *GPD* approximation. It would be advisable in my view to repeat the exercise - *mutatis mutandis* - for the three distributions from the appendix in order to create a more equal level playing field and a more realistic setting. It should be mentioned that the *ML* and *PWM* estimators in the two cases (*GEV* and *GPD*) are different (Hosking e.a. 1985 and Hosking e.a. 1987).

#### § 5 : Finite-sample properties of the two methods with non-stationary and dependent data

This section is the most interesting one in my view. Section 5.2 seems to show that despite the fact that the parent distribution is *GEV*, the *GEV* method seems to perform less well in general than the *GPD* approach. Apparently the non-i.i.d. character of the observations obscures the convergence to *GEV* more than the convergence to *GPD*.

I shall comment on the exclusive use of the *PWM* estimators at the end of the report.

In relation to table 5.1 (or rather table 5.6) it would be interesting to report the non-parametric estimator of the 4,000 and 10,000 years quantile i.e., the observations ranked 25 and 10 in size.

Figures 5.1 and 5.2 are nice and they illustrate the choice of threshold well. With respect to section 5.2.2 I would expect not much difference between the i.i.d. case and the particular non-i.i.d. case chosen. The fact that the *POT* method does not perform as well as before may be due to the fact that the underlying distribution is no longer *GPD* but *GEV*. This creates a systematic bias (cf. my remarks at the end).

The same remark applies to section 5.3. In this section the situation considered is the most realistic in view of the appropriate applications. Indeed it looks like the *POT* approach is a bit superior in a situation that is realistic for the “basic variables”.

**FINAL COMMENTS**

It seems to me that the scenario of section 5.3 - no *GEV* or *GPD* parent distributions and non i.i.d. random variables - offers the right platform for comparing different estimation methods given the fact that it seems to approach the real world situation of the basic variables the best. The case for the POT approach is convincing. It would be good not to restrict this exercise to the *PWM* estimator.

**APPENDIX**

Finally I would like to remark that a theoretical comparison between the various methods (*GEV* vs. *GPD* and *PWM* vs. *ML*) is also possible : virtually all distributions in the domain of attraction have a so-called second order parameter  $\rho$  measuring the closeness to the limit *GPD* distribution (besides the first order parameter  $\xi$ ). The minimal asymptotic mean square error of the estimators for  $\xi$  and a high quantile can be expressed in terms of those two parameters alone. The parameter  $\rho$  is non-positive. The larger (negative)  $\rho$  is, the better the approximation.

We (PhD student Juan Juan Cai and me) have calculated for which combinations of values of  $\xi$  and  $\rho$  the *PWM* estimator is better and for which ones the *ML* estimator is better based on the asymptotic mean-square criterion for i.i.d. random variables. For  $\gamma < 0.05$  *ML* is better. For  $\rho \geq -1$  *ML* is also better as well as for large negative  $\rho$ . There is an area, roughly  $0.05 < \gamma < 0.35$  and  $M < \rho < -1$  with  $M < -10$ , where *PWM* is better. We also calculated the  $\rho$  - values for the probability distributions used in the report:

*GPD* :  $\rho = -\infty$

*GEV* :  $\rho = -1$

*gamma* distribution :  $\rho = 0$

*beta* distribution of both kinds :  $\rho = -|\xi|$ .

These results have not been 100% rechecked. They are mainly meant to show that not one estimator dominates the other ones. And there are also other estimators that produce good results.

All five questions mentioned in the “guidelines” should be answered affirmatively.

Laurens de Haan  
 Professor, University of Tilburg  
 Emeritus professor, Erasmus University Rotterdam  
 Associate Scientist, University of Lisbon

November 24, 2009



### B.3 Reply to the review

I am very grateful to Prof. Laurens de Haan for his review of a previous version of this report reproduced in the preceding section. His advice and suggestions have led to an improvement of the exposition and of the analyses presented in the report. Below a reply is given to his comments using the same headings. Adjustments carried out on an earlier version of the report motivated by the review are also described. His review also complements the analyses given in the report.

#### §2

Section 2.2.3 was added to address the first point, and the explanation around Eq. (2.6) was rewritten and expanded.

#### §3

Yes, in principle we could have also used ARMA processes with other marginal distributions. However, as far as stationarity properties and simulation we are really only familiar with Gaussian processes, and it would have taken me some time to find out the relevant properties of non-gaussian processes. The processes we have simulated are indeed stationary.

Unfortunately, it is impossible to provide the average number of peaks for each month without redoing all the simulations, which would take me a considerable amount of time and would go beyond the time available for this project. However, as you say, the winter months dominate.

The assumption of a single tail index for all months may not be completely realistic. The reason why I took a single tail index is that it facilitates the interpretation of the results; in particular, if we know that the tail index is, say, equal to 0.1, then we can compare our estimates of  $\xi$  with that number, while if the tail index varied we would only know that the estimates of  $\xi$  should be around some average of the different tail indices. However, it seems realistic to me that the extremes should be dominated by a single tail index, and the simulation is intended to approximate that situation.

#### §4

Table 5.7, showing the performance of the ML estimates, was added.

#### §5

The non-parametric estimators of the quantiles have been added to tables 5.1 and 5.5.

### FINAL COMMENTS

As indicated in 4. above, the ML method was also considered in the case of the Gamma and Beta distributions.

## **APPENDIX**

We appreciate the new information provided on this topic. The results you mention are asymptotic, but they are of course qualitatively useful (e.g. no method is uniformly better than the other). Our conclusions in Chapter 4 have been updated in the light of your comments.