

# Uncertainties in extreme value modelling of wave data in a climate change perspective

Erik Vanem<sup>1</sup>

Received: 13 February 2015 / Accepted: 15 April 2015 / Published online: 8 May 2015  
© Springer International Publishing AG 2015

**Abstract** The extreme values of wave climate data are of great interest in a number of different applications, including the design and operation of ships and offshore structures, marine energy generation, aquaculture and coastal installations. Typically, the return values of certain met-ocean parameters such as significant wave height are of particular importance. In a climate change perspective, projections of such return values to a future climate are of great importance for risk management and adaptation purposes. However, there are various ways of estimating the required return values, which introduce additional uncertainties in extreme weather and climate variables pertaining to both current and future climates. Many of these approaches are investigated in this paper by applying different methods to particular data sets of significant wave height, corresponding to the historic climate and two future projections of the climate assuming different forcing scenarios. In this way, the uncertainty due to the extreme value analysis can also be compared to the uncertainty due to a changing climate. The different approaches that are considered in this paper are the initial distribution approach, the block maxima approach, the peak over threshold approach and the average conditional exceedance rate method. Furthermore, the effect of different modelling choices within each of the approaches will be explored. Thus, a range of different return value estimates for the different data sets is obtained. This exercise reveals that the uncertainty due to the extreme value analysis method is notable and, as expected, the variability of the estimates increases for higher return periods. Moreover, even though the variability due to the extreme value analysis is greater than the climate variability, a shift towards higher extremes in a future wave climate

can clearly be discerned in the particular datasets that have been analysed.

**Keywords** Ocean and coastal engineering · Wave climate · Extreme value analysis · Climate change · Significant wave height · Environmental loads

## 1 Introduction

Extreme value analysis of wave climate parameters is an important part of ocean and coastal engineering where the extreme loads from extreme environmental conditions need to be taken into account. However, there are large uncertainties associated with extreme value analyses, and the uncertainties generally increase for higher return periods. Ideally, time series that are long compared to the desired return periods should be available to reliably extract return values. In practice, however, the opposite is true and return values corresponding to return periods much longer than the length of recorded data are needed. Therefore, there is a need to extrapolate to obtain estimates of the tail behaviour of the underlying statistical distributions. Intuitively, the further away from the data one has to extrapolate, the larger the uncertainties of the resulting estimates will be. As a rule of thumb, for example, the ISO standard ISO 19901-1 (ISO 2005) recommends to not use return periods more than a factor of four beyond the length of the data set when deriving return values for design of offshore structures. Hence, for the datasets analysed in this paper, covering a period of 30 years, the longest return periods that should be investigated are 120 years. Adhering to this rule of thumb, return values for 20- and 100-year return periods will be estimated in this paper.

✉ Erik Vanem  
Erik.Vanem@dnvgl.com

<sup>1</sup> DNV-GL Strategic Research and Innovation, Høvik, Norway

There are a number of different approaches to extreme value analysis and return value estimation, which all rely on a set of assumptions. The initial distribution approach fits a statistical model to all the data under the assumption of independent and identically distributed (iid) observations and estimate high return values by extrapolating the fitted distribution to high quantiles corresponding to the desired return periods. However, one fundamental problem with this approach is that most of the data used to fit the model will lie near the mode of the distribution and hence quite remote from the tail area of interest. As a consequence, such models will typically often be able to capture the area close to the mode of the distribution quite well, but may give poor fit to the tail of the distribution. Another source of uncertainty encountered with this approach, as indeed with all statistical model fits, is the fitting procedures. Even after having selected a parametric model to fit to the data, there are several methods to estimate the model parameters, such as the maximum likelihood, the method of moments, the least squares method and other approaches. Some methods for the initial distribution approach will be investigated in this paper and compared to other means of estimating extreme values.

Some of the classical approaches to extreme value analysis rely on assumptions on the asymptotic behaviour of the extremes as the number of observations approaches infinity. These methods will typically also assume that the data are iid, i.e. that the observations are realizations from the same stationary process and can be construed as independent samples drawn from the same probability distribution. Two commonly used approaches to extreme value analysis are the block maxima (BM) approach and the peaks over threshold (POT) approach. An obvious drawback with these approaches is that they are wasteful and only exploits a small subset of all the data available. As will be demonstrated in this study, this also significantly increases the statistical uncertainty of the resulting return value estimates. An introduction to these methods, along with a general introduction to the theory behind extreme value analysis, can be found in [Coles \(2001\)](#). Both the block maxima approach and the POT method will be explored in this paper. Recent applications of the POT approach to analyse the extremes of ocean waves are presented in e.g. [Caires and Sterl \(2005\)](#) and [Thevasiyani and Perera \(2014\)](#).

A more recent method for extreme value analysis that allows for the assumptions of independence to be relaxed is proposed in [Naess and Gaidai \(2009\)](#), i.e. the average conditional exceedance rate method (ACER). It was initially proposed only for the asymptotic Gumbel case, but was later extended to apply in more general cases ([Naess et al. 2013](#)). A further generalization to the bivariate case has also been presented in [Naess and Karpa \(2013\)](#). In the study presented in this paper, the univariate ACER approach will be applied

and compared to the other methods for estimating extreme return values.

A full Bayesian approach to extreme value modelling is set forth in [Coles et al. \(2003\)](#), and a Bayesian hierarchical model is presented in [Oliver et al. \(2014\)](#) for estimating extremes from climate model output. A review of Bayesian approaches to extreme value methods is given in [Coles and Powell \(1996\)](#). An alternative approach to estimate return values of significant wave height, referred to as the (modified) Rice method is proposed in [Rychlik et al. \(2011\)](#). See also the review of extreme value modelling for marine design in [Jonathan and Ewans \(2013\)](#). Possibly, results from these and other approaches could be applied in future work and compared to the results in this paper.

Several previous studies have discussed the bias and uncertainty of extreme value prediction of metocean parameters, see, e.g. [Gibson et al. \(2009\)](#), [Hagen \(2009\)](#), and [Aarnes et al. \(2012\)](#). Uncertainty of extreme value prediction is divided into statistical uncertainty and modelling uncertainty in [Li et al. \(2014\)](#) which also investigates the impact on structural reliabilities. The uncertainty of design values from extreme value analyses is also addressed in [Harris \(2001\)](#) and different methods of extreme value estimation of wind speeds are compared in [An and Pandey \(2005\)](#). Some fundamental problems in extreme value analysis such as uncertainties due to plotting position, the fitting method and due to the fact that the asymptotic conditions are never fulfilled in practice are discussed in [Makkonen \(2008\)](#). In design problems, joint distributions of several metocean parameters are often needed, and the extremal dependence of those variables becomes very important. This is addressed in [Towe et al. \(2013\)](#), but the present study is limited to univariate extreme value analyses. It is out of scope of this paper to provide a comprehensive literature survey of uncertainties in extreme value analysis, and the references above are included simply to demonstrate that this is an important issue that has been discussed at length in the academic and technical literature without arriving at final conclusions.

It is noted that there are several sources of uncertainties of future climate projections that are not investigated in this paper. The climate scenario is obviously important, and only two future scenarios are considered in this study. However, studies have demonstrated that the choice of climate model might contribute more to the overall uncertainty of the future wave climate than the climate scenarios themselves ([Wang and Swail 2006](#); [Grabemann and Weisse 2008](#); [Grabemann et al. 2015](#); [Wang et al. 2012, 2014](#); see also [de Winter et al. 2013](#)). Future projections of waves are typically obtained using wind output from climate models as input to numerical wave models, and also the choice of wave model, the downscaling method and the model resolution will have a big impact on the results. Hence, the uncertainties associated with future wave climate extremes are not restricted to

the uncertainties due to the statistical extreme value analysis which is the focus of this paper. Nevertheless, these uncertainties remain important and will also be equally present in the estimation of extremes in historical, present and future climates.

In the following, several approaches to univariate extreme value analysis are applied to wave climate data for a historical period and for two future climate projections. First, the wave data will be described and then the results from the different extreme value analyses will be presented. The estimated return values from the different methods are then compared and the differences are discussed. Preliminary results from this study will be presented at the OMAE 2015 conference (Vanem 2015).

## 2 The wave data

For the purpose of this study, a particular set of three datasets generated by running the numerical wave model WAM (Reistad et al. 2011) for an area in the North Atlantic with forcings from the climate model GFDL-CM3 (Donner et al. 2011) are analysed. One particular location is analysed, disregarding the spatial variability, i.e. the location at 59.28°N and 11.36°W in the North Atlantic Ocean. The three datasets correspond to a 30-year historical period from 1970 to 1999 and two future periods consistent with the RCP 4.5 and RCP 8.5 scenarios, respectively (Moss et al. 2010; van Vuuren et al. 2011), from 2071 to 2100. The variable that will be considered is the significant wave height,  $H_s$ , of the total sea, i.e. including both the wind sea and the swell compo-

nent. The temporal resolution of the data is 3 h, resulting in very long time series. A total of 87,599 data are included in the historical dataset and the future projections contain 87,597 data points. Summary statistics from the datasets are presented in Table 1. These summary statistics indicate that the distributions for the different scenarios are different. An interesting observation is that even though the extremes and all quantiles higher than 95 % are higher in both the future scenarios, the mean value and actually all quantiles up to at least the median are higher for the historical period. However, the data for the future climate scenarios have higher standard deviation (larger spread), higher positive skewness (longer upper tails) and higher kurtosis (fatter upper tails), which would all contribute to give higher extreme values in a future scenario.

## 3 Extreme value analyses of significant wave height

In the following, the datasets for significant wave height presented above will be subject to different extreme value analysis methods and it will be investigated how sensitive the results are to the choice of method and different choices within each method. Obviously, the actual results are conditioned on this dataset, but it is believed that it will still give a good indication of the modelling uncertainties in extreme value analysis. The approaches that have been applied are different variations of the initial distribution approach, the block maxima approach, the peaks-over-threshold approach and the ACER method.

**Table 1** Summary statistics and quantiles

Statistic	Historic period (1970–1999)	RCP 4.5 (2071–2100)	RCP 8.5 (2071–2100)
Number of data points ( $N$ )	87,599	87,597	87,597
Mean	3.232	3.158	3.122
Standard deviation	1.754	1.862	1.979
Skewness	1.216	1.395	1.413
Kurtosis	1.996	3.177	2.688
Minimum value	0.3	0.2	0.2
10 % quantile	1.4	1.2	1.1
25 % quantile	1.9	1.8	1.7
50 % quantile (median)	2.9	2.7	2.7
75 % quantile	4.1	4.1	4.1
90 % quantile	5.7	5.6	5.8
95 % quantile	6.6	6.7	7.0
99 % quantile	8.7	9.2	9.6
99.9 % quantile	11.5	12.6	12.8
99.99 % quantile	14.5	17.7	15.9
Maximum value	15.6	21.7	17.5

### 3.1 Initial distribution approach

One approach to extreme value modelling is the initial distribution approach, where all the data are used to fit a probability density function and then the extremes are estimated from the fitted distribution by extrapolation to higher return periods. The obvious advantage of this method is that it exploits all the available information, i.e. data. However, as most data will lie close to the mode of the distribution, it might not give a very good representation of the tail area of the distribution, which is the interesting part for extreme value analysis.

Often, the Weibull distribution is assumed appropriate for significant wave height, and in this paper the results of fitting both the 2- and 3-parameter Weibull distributions to the data by way of maximum likelihood are reported, in Tables 2, 3, respectively. The 3-parameter Weibull distribution, sometimes referred to as the translated Weibull distribution, is a generalization of the 2-parameter Weibull distribution where an additional location parameter,  $\mu$ , has been added. The location parameter allows a lower limit for the support of the distribution and the density function will be zero for values less than this location parameter. The probability distribution function of the 3-parameter Weibull distribution is defined as follows, with location parameter  $\gamma$ , scale parameter  $\alpha$  and shape parameter  $\beta$ :

$$f(x) = \frac{\beta}{\alpha} \left( \frac{x - \gamma}{\alpha} \right)^{\beta-1} e^{-\{(x-\gamma)/\alpha\}^\beta} \quad x \geq \gamma \quad (1)$$

The shape and scale parameters need to take positive values;  $\alpha > 0$ ,  $\beta > 0$ , and the location parameter can in principle

take any value, but  $f(x) = 0$  for  $x \leq \gamma$ . The 2-parameter Weibull distribution function is easily obtained by letting the location parameter be zero.

The results differ considerably and illustrate the sensitivity due to the choice of distribution to use. Obviously, other distributions could also have been assumed to give different return value estimates. It is also observed that the extremes tend to be underestimated by these models. In fact, except for the RCP 8.5 data with the 3-parameter Weibull model, all the 100-year return value estimates are lower than the 99.99 %-tile in the original data. For 3-hourly data, the 99.99 %-tile corresponds to a return period of less than 3.5 years.

Even if the parametric model has been decided, there are several fitting methods that can be employed that will yield different parameter estimates and consequently different return value estimates. In order to illustrate this, the results of fitting the 3-parameter Weibull distribution to the data by way of the method of moments (MoM), the method of L-moments, minimization of the Cramer von Mises distance and by minimizing the second order Anderson–Darling statistic, respectively. The results are presented in Tables 4, 5, 6 and 7.

The fits based on the second-order Anderson–Darling statistic are found to consistently yield higher return value estimates, and this statistic is known to give more weight on the tail of the distribution. Hence, this might be the preferred method in this case. Nevertheless, this study demonstrates that there are large uncertainties associated with return value estimates for high return periods, even with a single approach (initial distribution approach) and with a single parametric model. The different fitted distributions are illustrated in

**Table 2** Parameter and return value estimates for 2-parameter Weibull distributions fitted to all the data by maximum likelihood

	Historical period	RCP 4.5	RCP 8.5
Shape parameter ( $\beta$ )	1.9670	1.8173	1.6955
Scale parameter ( $\alpha$ )	3.6628	3.5720	3.5206
20-year return value	12.38 m [0.0395]	13.35 m [0.0451]	14.46 m [0.0534]
95 % CI	(12.31, 12.46)	(13.26, 13.44)	(14.36, 14.56)
100-year return value	13.27 m [0.0444]	14.39 m [0.0512]	15.68 m [0.607]
95 % CI	(13.19, 13.36)	(14.29, 14.49)	(15.57, 15.80)
Standard errors in brackets			

**Table 3** Estimated parameters and return values for 3-parameter Weibull distributions fitted to all the data by maximum likelihood

	Historical period	RCP 4.5	RCP 8.5
Shape parameter ( $\beta$ )	1.7763	1.6965	1.5757
Scale parameter ( $\alpha$ )	3.3098	3.3322	3.2732
Location parameter ( $\mu$ )	0.2995	0.1995	0.1989
20-year return value	13.05 m [0.0458]	13.88 m [0.0499]	15.17 m [0.0613]
95 % CI	(12.95, 13.14)	(13.78, 13.97)	(15.05, 15.30)
100-year return value	14.07 m [0.0520]	15.03 m [0.0569]	16.53 m [0.0703]
95 % CI	(13.96, 14.17)	(14.91, 15.14)	(16.39, 16.67)

**Table 4** Estimated parameters and return values for 3-parameter Weibull distributions fitted to all the data by the method of moments (MoM)

	Historical period	RCP 4.5	RCP 8.5
Shape parameter ( $\beta$ )	1.3875	1.2703	1.2597
Scale parameter ( $\alpha$ )	2.6337	2.5311	2.6642
Location parameter ( $\mu$ )	0.8283	0.8087	0.6453
20-year return value	15.63 m [0.109]	17.49m [0.144]	18.49 m [0.154]
95 % CI	(15.42, 15.85)	(17.20, 17.76)	(18.20, 18.81)
100-year return value	17.17 m [0.130]	19.39m [0.174]	20.54 m [0.185]
95 % CI	(16.91, 17.43)	(19.04, 19.72)	(20.18, 20.92)

**Table 5** Estimated parameters and return values for 3-parameter Weibull distributions fitted to all the data by the method of L-moments (LM)

	Historical period	RCP 4.5	RCP 8.5
Shape parameter ( $\beta$ )	1.411	1.347	1.269
Scale parameter ( $\alpha$ )	2.672	2.686	2.681
Location parameter ( $\mu$ )	0.780	0.694	0.633
20-year return value	15.40 m [0.0840]	16.60 m [0.0929]	18.33 m [0.112]
95 % CI	(15.23, 15.56)	(16.41, 16.77)	(18.11, 18.56)
100-year return value	16.89 m [0.0986]	18.30 m [0.110]	20.35 m [0.133]
95 % CI	(16.68, 17.08)	(18.07, 18.50)	(20.08, 20.62)

**Table 6** Estimated parameters and return values for 3-parameter Weibull distributions fitted to all the data by minimizing the Cramer von Mises distance

	Historical period	RCP 4.5	RCP 8.5
Shape parameter ( $\beta$ )	1.4383	1.3788	1.2960
Scale parameter ( $\alpha$ )	2.6612	2.7020	2.6855
Location parameter ( $\mu$ )	0.8042	0.6770	0.6262
20-year return value	14.88 m [0.0947]	16.03 m [0.110]	17.68 m [0.126]
95 % CI	(14.68, 15.07)	(15.80, 16.23)	(17.43, 17.93)
100-year return value	16.28 m [0.111]	17.63 m [0.129]	19.58 m [0.149]
95 % CI	(16.05, 16.51)	(17.36, 17.86)	(19.28, 19.88)

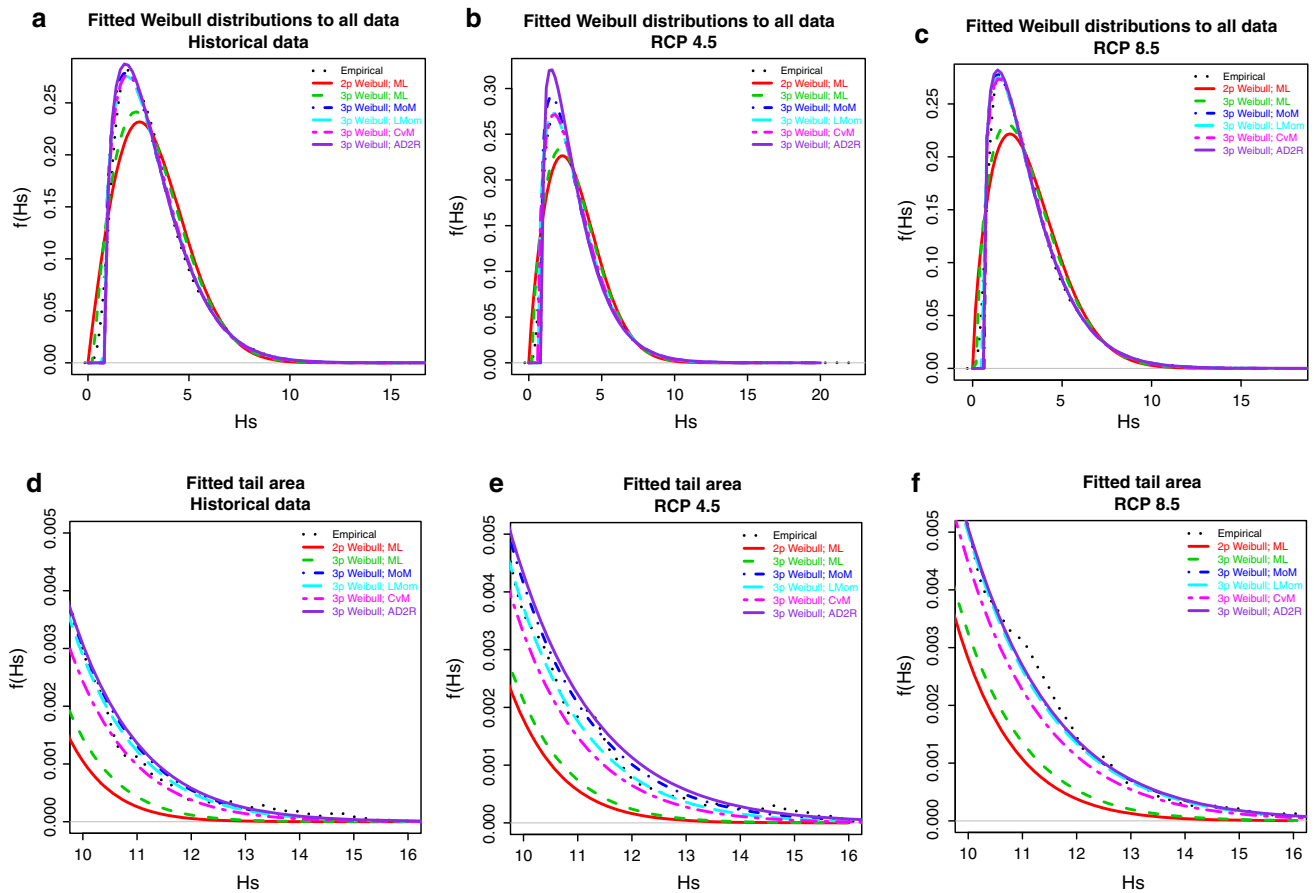
**Table 7** Estimated parameters and return values for 3-parameter Weibull distributions fitted to all the data by minimizing the second-order Anderson–Darling statistic

	Historical period	RCP 4.5	RCP 8.5
Shape parameter ( $\beta$ )	1.3564	1.2012	1.2485
Scale parameter ( $\alpha$ )	2.5522	2.3424	2.6301
Location parameter ( $\mu$ )	0.9032	0.9783	0.6765
20-year return value	15.83 m [0.162]	18.19m [0.194]	18.60 m [0.208]
95 % CI	(15.49, 16.08)	(17.74, 18.54)	(18.10, 18.89)
100-year return value	17.41 m [0.197]	20.27 m [0.237]	20.67 m [0.255]
95 % CI	(16.99, 17.71)	(19.72, 20.69)	(20.06, 21.02)

Fig. 1, which also includes a close-up of the tail areas. It is observed that in spite of its theoretical advantages, the maximum likelihood estimates seem to yield the worst fit to the data.

Included in the tables are also bootstrap estimates of the standard errors (in brackets) and the 95 % confidence intervals of the return value estimates. These are obtained by parametric bootstrap with  $B = 1000$  bootstrap samples and are conditioned on the estimated model. It is observed that the

statistical uncertainty conditioned on the established models is quite small. This is due to the quite large datasets that are available for the initial distribution approach. The confidence intervals for the historical period are mostly not overlapping with the corresponding confidence intervals for the future scenarios. This indicates that there is a statistically significant trend, at the 5 % confidence level, in the extreme wave climate described by these particular datasets. However, the different results from the different fitting methods illustrate



**Fig. 1** The alternative fitted Weibull densities for the historical data (*left*), RCP 4.5 (*middle*) and RCP 8.5 (*right*). The *upper row* shows the complete distributions and the *lower row* shows a close-up on the tail areas

that the unconditional uncertainty is quite large due to the problem of accurately fitting the tail area of the distribution.

### 3.2 The block maxima approach

One common approach to extreme value analysis is to fit a parametric model such as the generalized extreme value (GEV) model to block maxima (Coles 2001). Block sizes of one year is typical, but the results might be sensitive to the block size. In this paper, different block sizes are tried out to investigate this. Furthermore, it is implicitly assumed that the maxima are independent and identically distributed (iid). Within a climate change perspective, this approach needs to assume that the extremes can be considered stationary within each of the time intervals, i.e. that the extremes are stationary during the 30-year reference period and the 30-year projection period. If the effect of any long-term trend is small compared to the other variability, this might not be a very unrealistic approximation.

The density function of the GEV distribution is as follows, with location parameter  $\mu \in \mathbb{R}$ , scale parameter  $\sigma > 0$  and shape parameter  $\xi \in \mathbb{R}$ :

$$g(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi - 1} e^{-\left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi}}$$

$$\text{for } 1 + \xi \left( \frac{x - \mu}{\sigma} \right) > 0 \tag{2}$$

In the following, different GEV-models will be fitted to each data set, with block sizes corresponding to 1 year, 2 years and 6 months, respectively. The different block sizes give different subsets of the data for fitting; small block sizes give a larger sample to fit to the model, but the iid assumption becomes less obvious. For example, using the semiannual maxima over the 30-year periods gives 60 samples and assumes that the distributions of wave heights are the same for the first 6 months of the year as for the last 6 months. Using the biennial maxima, on the other hand, reduces the sample size to 15.

This exercise yields a total of 9 different estimated GEV-distributions, and the model parameters are presented in Table 8 together with the corresponding 20-year and 100-year return periods. Again, standard error and 95% confidence intervals for the return values are estimated by parametric bootstrap. Note that the parameter estimates for the different

**Table 8** Estimated GEV-models and associated return value estimates for different block sizes

Block size	Historical period			RCP 4.5			RCP 8.5		
	12m	24m	6m	12m	24m	6m	12m	24m	6m
Shape ( $\xi$ )	-0.0377	-0.595	-0.0629	0.272	-0.0472	0.142	-0.0711	-0.393	-0.219
Scale ( $\sigma$ )	1.420	1.950	1.520	1.717	2.324	1.717	1.533	1.766	1.977
Location ( $\mu$ )	11.341	12.619	10.210	11.821	13.699	10.563	12.633	13.946	11.403
20-Year return value	15.33 m [0.867]	15.04 m [0.353]	15.20 m [0.811]	19.66 m [3.068]	18.66 m [1.489]	18.85 m [1.866]	16.74 m [0.844]	16.58 m [0.468]	16.67 m [0.585]
95 % CI	(13.60, 16.88)	(14.60, 16.02)	(13.55, 16.80)	(12.05, 23.54)	(15.70, 16.75)	(14.56, 21.85)	(15.03, 18.34)	(15.93, 17.76)	(15.58, 17.89)
100-Year return value	17.34 m [2.311]	15.58 m [0.538]	17.05 m [1.552]	27.55 m [12.617]	21.98 m [6.365]	24.11 m [4.634]	18.65 m [1.850]	17.47 m [0.938]	17.94 m [0.962]
95 % CI	(12.57, 20.15)	(15.12, 16.75)	(13.67, 19.57)	(-5.69, 36.99)	(8.86, 27.17)	(12.92, 30.45)	(14.35, 21.57)	(16.05, 19.12)	(16.02, 19.77)

**Table 9** Estimated Gumbel models and associated return value estimates for different block sizes

Block size	Historical period			RCP 4.5			RCP 8.5		
	12m	24m	6m	12m	24m	6m	12m	24m	6m
Scale ( $\sigma$ )	1.399	1.655	1.494	1.977	2.291	1.827	1.491	1.646	1.977
Location ( $\mu$ )	11.312	12.078	10.160	12.093	13.640	10.698	12.576	13.595	11.167
20-Year return value	15.47 m [0.741]	15.80 m [0.972]	15.65 m [0.639]	17.96 m [1.037]	18.80 m [1.335]	17.41 m [0.794]	17.00 m [0.763]	17.30 m [0.972]	18.43 m [0.860]
95 % CI	(14.01, 16.92)	(13.91, 17.73)	(14.37, 16.85)	(15.97, 20.08)	(16.26, 21.51)	(15.77, 18.86)	(15.47, 18.54)	(15.44, 19.18)	(16.67, 20.09)
100-Year return value	17.75 m [1.061]	18.53 m [1.481]	18.07 m [0.873]	21.19 m [1.478]	22.58 m [2.028]	20.37 m [1.082]	19.43 m [1.087]	20.02 m [1.491]	21.64 m [1.173]
95 % CI	(15.73, 19.84)	(15.64, 21.56)	(16.34, 19.74)	(18.32, 24.26)	(18.88, 26.77)	(18.17, 22.36)	(17.31, 21.63)	(17.08, 23.02)	(19.33, 23.88)

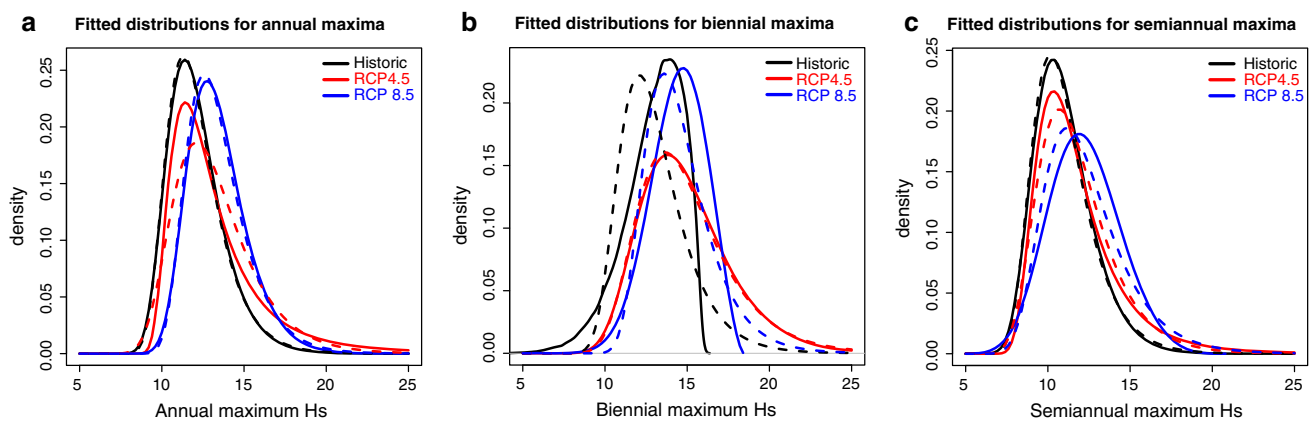
models are not comparable since the data sets are different, but the resulting return values should in principle be comparable—they are different estimates of the same quantity. All the fits are obtained by the maximum likelihood method.

These results indicate that the block size has an influence on the estimated distributions and hence on the estimated return values. It is observed that using the 1-year extremes to fit a GEV-model gives generally higher return values. It is interesting to observe that the statistical uncertainty of the return value estimates are much higher for the block maxima approach compared to the initial distribution approach. This is natural since the number of data samples is much lower. Whereas the initial datasets had almost 90,000 samples, the annual maxima data, for example, contain only 30 yearly maxima. It is also observed that the estimates pertaining to the RCP 4.5 data are particularly uncertain. Indeed, the 100-year return value 95 % confidence interval for this dataset ranges from negative values to almost 37 m. The confidence intervals for the return value estimates in the historical data overlap with the confidence intervals for the future projections and, therefore, the block maxima approach is in fact not able to detect any statistically significant shift due to climate change. This is in contrast to the initial distribution approach, where the climatic shifts were found to be statistically significant.

The Gumbel distribution is a special case of the GEV-distribution where the shape parameter,  $\xi$ , is 0. It can be seen from the table above that in fact  $\xi$  is not significantly different from 0 in some cases. Hence, one could assume

$\xi = 0$  and fit a Gumbel distribution to the extremes and compare the results with the full GEV-model. The resulting parameter estimates and return value estimates are presented in Table 9, with standard error estimates of the return values included in brackets. In practice, the effect of assuming a Gumbel distribution will be that for GEV-models with a positive shape parameter, the return values will be reduced, whereas the GEV-models with negative shape parameters will get increasing return values by enforcing  $\xi = 0$ . The uncertainties associated with this choice illustrate additional model uncertainties that contribute to the overall uncertainty when modelling the extremes. However, other choices than the Gumbel model are available, so the difference between the fitted GEV-model and the Gumbel model does not reflect the overall extreme value model uncertainty. It is also noted that some researchers have warned against applying the Gumbel reduction, even in cases where the data support such a model reduction, as this generally tends to overestimate the precision (the uncertainty in the  $\xi$ -parameter is essentially eliminated) which may lead to anti-conservative results, see, e.g. Coles et al. (2003). On the other hand, if the data really belong to the Gumbel domain of attraction, as they seem to do in this case, the Gumbel distribution will be the appropriate asymptotic extreme value distribution.

The estimated uncertainty of the return values is based on parametric bootstrap and it is observed that the statistical uncertainty of the return value estimates is considerably large, even if conditioned on the fitted models. However, the estimates are more precise than those obtained by the



**Fig. 2** Fitted extreme value distributions to the historic and projected block maximum data; annual maxima (*left*), biennial maxima (*centre*) and semiannual maxima (*right*). *Solid lines* represent full GEV models and *dashed lines* represent Gumbel models (assuming  $\xi=0$ )

**Table 10** Estimated parameters and return values for 2-parameter Weibull distributions fitted to annual maxima

	Historical period	RCP 4.5	RCP 8.5
Shape parameter ( $\beta$ )	7.4206	4.6562	7.7574
Scale parameter ( $\alpha$ )	12.8938	14.4703	14.2489
20-Year return value	14.95 m [0.408]	18.32 m [0.804]	16.41 m [0.437]
95 % CI	(14.27, 15.90)	(16.79, 19.91)	(15.66, 17.39)
100-Year return value	15.84 m [0.500]	20.09 m [1.031]	17.35 m [0.538]
95 % CI	(14.99, 16.99)	(18.19, 22.12)	(16.40, 18.55)

**Table 11** Estimated parameters and return values for 3-parameter Weibull distributions fitted to annual maxima

	Historical period	RCP 4.5	RCP 8.5
Shape parameter ( $\beta$ )	1.7322	1.0932	1.661
Scale parameter ( $\alpha$ )	3.2699	3.3722	3.3446
Location parameter ( $\mu$ )	9.2102	10.05	10.4397
20-Year return value	15.37 m [0.940]	19.25 m [11.86]	16.91 m [1.022]
95 % CI	(13.00, 16.81)	(−23.26, 24.22)	(14.47, 18.52)
100-Year return value	17.11 m [1.675]	23.68 m [24.308]	17.35 m [1.808]
95 % CI	(12.38, 19.19)	(−63.80, 32.92)	(14.14, 21.21)

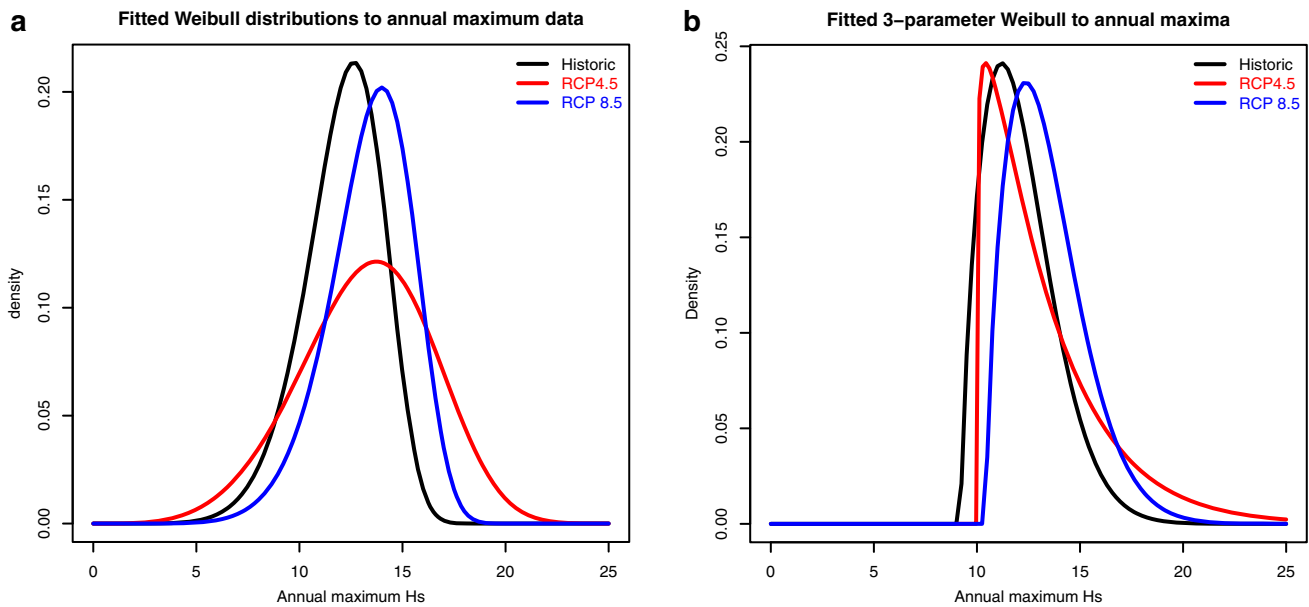
more flexible GEV model, most notably for the RCP 4.5 data. However, this can be explained by the stronger assumptions associated with the Gumbel reduction, which camouflages some of the uncertainties. Nevertheless, also the confidence intervals of the return value estimates based on the Gumbel model are overlapping, and the climatic shifts that are detected are not statistically significant. By comparing the full GEV models with the reduced Gumbel model, it is observed that the full GEV model yields higher return values if using annual extremes and that the Gumbel model gives higher return values if using biennial extremes. Hence, it is not straightforward to assign a particular bias for this model and block size uncertainty.

The fitted extreme value distributions are shown in Fig. 2. The figure to the left shows fitted distributions to the annual maxima (a total of 30 data points); in the middle are the

fitted distributions for the biennial maxima (a total of 15 data points) and to the right the semiannual maxima (a total of 60 data points). The solid lines represent the full GEV models whereas the dashed lines are the Gumbel models. All the models were obtained by maximum likelihood fitting, but obviously other fitting methods are available as well. See Wang et al. (2013) for a discussion on uncertainties related to different estimators for the GEV models.

Another modelling option is to assume that the extremes are Weibull distributed and fit a (2- or 3-parameter) Weibull distribution to the annual maxima. The results from fitting the annual maxima (the same data that were used in fitting the GEV-models above) to Weibull distributions by maximum likelihood are presented in Tables 10 and 11, respectively, including model parameters and return value estimates. The uncertainty estimates of the return values are again calcu-





**Fig. 3** Fitted Weibull distributions to the annual maxima for the historic period and future projections; 2-parameter (*left*) and 3-parameter (*right*) distributions

lated by parametric bootstrap. It is interesting to observe that assuming 2-parameter Weibull distributions yields a statistically significant increase in the return values for the RCP 4.5 scenario but not for RCP 8.5. Assuming the 3-parameter Weibull, on the other hand, yields much wider confidence intervals and no statistically significant trend. The fitted density functions are shown in Fig. 3. Compared to the fitted full GEV-models, it is observed that the similar general results are obtained, i.e. that there will be an expected increase in the return value estimates for future climates and that the increase is higher for RCP 4.5 than for RCP 8.5. It is also observed that the estimated return values using a 2-parameter Weibull distribution are lower than the estimates from the full GEV model, whereas that the estimates from a 3-parameter distribution are comparable.

It is noted that for the RCP 4.5 data, there was a need to specify an upper limit for the location parameter in order for the optimization procedure to converge, and the estimated location parameter would take this value when fitting the 3-parameter Weibull distribution. Hence, there might be additional numerical uncertainties in these estimates. This is confirmed by the uncertainty estimates of the return values obtained by bootstrapping, which are very large for the RCP 4.5 data. Indeed, the 95% confidence intervals are very wide and do not make much sense. It is emphasized, however, that it might be artificial to try to fit the data for the RCP 4.5 scenario to a 3-parameter Weibull distribution. The Weibull distribution is a special case of the GEV-model for negative shape parameters  $\xi < 0$ , and the fitted shape parameter for these data was indeed positive (see Table 8). Hence, one would expectedly run into problems

**Table 12** Estimated parameters and return values for the Fréchet distribution fitted to annual maxima of the RCP 4.5 data

	RCP 4.5
Location parameter	5.5019
Scale parameter	6.3190
Shape parameter	3.6816
20-Year return value	19.66 m [1.897]
95% CI	(15.89, 23.37)
100-Year return value	27.55 m [4.415]
95% CI	(18.88, 35.97)

when trying to fit a Weibull distribution to data that do not support such an assumption. Therefore, this particular data set will be tried to be fitted to another special case of the GEV-distribution, i.e. the Fréchet distribution. The results are presented in Table 12 and it is observed that this model yields the same return value estimates as the full GEV model. In fact, the fitted Fréchet distribution is nearly identical to the estimated GEV-model and the density function is not shown.

### 3.3 The peaks-over-threshold approach

A different approach to extreme value analysis is to fit statistical models to any data points above a specified threshold, the so-called peaks-over-threshold (POT) approach (Coles 2001). One of the benefits of the peaks-over threshold method compared to the block maxima method is that the amount of data available for fitting the model will gen-

**Table 13** Estimated GPD-models and associated return value estimates for the threshold exceedances with  $u = 10$  m for different cluster distances

	Historical period			RCP 4.5			RCP 8.5		
	No clusters	2-day clusters	4-day clusters	No clusters	2-day clusters	4-day clusters	No clusters	2-day clusters	4-day clusters
Threshold ( $u$ )	10	10	10	10	10	10	10	10	10
Scale ( $\sigma$ )	1.460	1.428	1.510	1.509	1.684	1.748	1.673	2.466	2.620
Shape ( $\xi$ )	-0.097	-0.0189	-0.491	0.0841	0.112	0.109	-0.127	-0.258	-0.285
20-Year return value	18.76 m [0.542]	15.28 m [0.886]	15.18 m [0.824]	21.14 m [1.280]	18.80 m [1.718]	18.89 m [1.805]	17.12 m [0.470]	16.51 m [0.482]	16.52 m [0.482]
95 % CI	(14.93, 17.04)	(13.44, 16.94)	(13.55, 16.81)	(18.29, 23.51)	(14.92, 21.77)	(14.96, 21.93)	(16.16, 18.05)	(15.74, 17.57)	(15.79, 17.61)
100-Year return value	17.27 m [0.889]	17.38 m [1.946]	17.12 m [1.682]	25.36 m [2.445]	23.51 m [4.178]	23.67 m [4.527]	18.24 m [0.696]	17.55 m [0.726]	17.50 m [0.717]
95 % CI	(15.50, 18.91)	(12.84, 20.36)	(13.35, 20.12)	(19.48, 29.56)	(13.05, 29.50)	(12.56, 29.88)	(16.86, 19.60)	(16.35, 19.12)	(16.32, 19.10)

**Table 14** Estimated GPD-models and associated return value estimates for the threshold exceedances with  $u = 12$  m for different cluster distances

	Historical period			RCP 4.5			RCP 8.5		
	No clusters	2-day clusters	4-day clusters	No clusters	2-day clusters	4-day clusters	No clusters	2-day clusters	4-day clusters
Threshold ( $u$ )	12	12	12	12	12	12	12	12	12
Scale ( $\sigma$ )	1.887	2.924	3.171	2.112	2.711	2.711	1.669	2.112	2.189
Shape ( $\xi$ )	-0.489	-0.795	-0.871	-0.0364	-0.110	-0.110	-0.196	-0.294	-0.313
20-Year return value	15.24 m [0.182]	15.14 m [0.270]	15.18 m [0.264]	20.57 m [1.171]	18.71 m [1.181]	18.71 m [1.180]	17.12 m [0.413]	16.56 m [0.481]	16.57 m [0.461]
95 % CI	(15.02, 15.73)	(14.90, 15.93)	(14.95, 15.97)	(18.21, 22.94)	(16.48, 21.32)	(16.56, 21.11)	(16.42, 18.03)	(15.81, 17.73)	(15.93, 17.71)
100-Year return value	15.58 m [0.235]	15.53 m [0.266]	15.53 m [0.257]	23.39 m [2.260]	21.62 m [2.707]	21.62 m [2.560]	18.04 m [0.649]	17.54 m [0.758]	17.53 m [0.729]
95 % CI	(15.34, 16.20)	(15.45, 16.44)	(15.47, 16.43)	(18.68, 27.46)	(16.60, 26.34)	(16.79, 26.23)	(16.84, 19.38)	(16.35, 19.33)	(16.47, 19.29)

erally increase—one is allowed to include more than the maximum from each block. However, the results are known to be sensitive to the selection of threshold, adding additional uncertainties, and one should also take care to avoid dependent data, for example, using two extremes from one and the same storm. This can be ensured by adequately applying de-clustering techniques, where the duration of a cluster must be specified. Having selected an exceedance threshold and a minimum cluster separation, the cluster maxima can be modelled according to the Generalized Pareto Distribution (GPD) by estimating the scale and shape parameters.

The density function of the GPD is defined as follows, with location  $\mu \in \mathbb{R}$ , scale  $\sigma > 0$  and shape  $\xi \in \mathbb{R}$ :

$$g(x; \sigma, \xi) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi - 1}$$

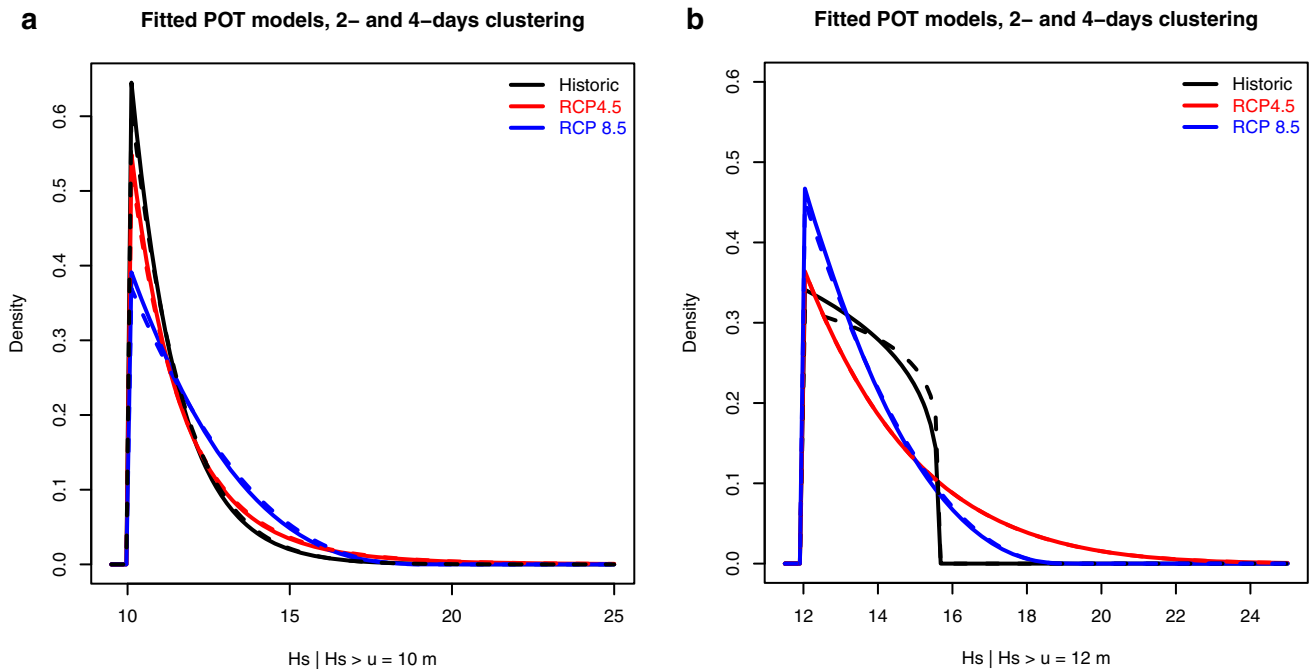
$$\text{for } 1 + \xi \left( \frac{x - \mu}{\sigma} \right) > 0 \tag{3}$$

In the following, the results from applying the peaks-over-threshold analysis to the significant wave height data will be reported, where different values of both the threshold and the cluster separation distance have been applied. Again, only the maximum likelihood method has been used for the fitting. It is noted that the exponential distribution emerges as a special case of the GPD when the shape parameter  $\xi \rightarrow 0$ . However,

this model reduction has not been applied in this study, and only the full GPD has been assumed.

Two different threshold values,  $u = 10$  m and  $u = 12$  m, and three cluster separation distances, corresponding to 2-day and 4-day cluster separation and no de-clustering, have been applied. Results for even longer clusters of 10 days are presented in Vanem (2015). This yields different datasets to be fitted to the GPD-model; when applying a 10 m threshold value and no clustering the average number of events per year were 8.9, 15.6 and 22.5 for the historic, RCP 4.5 and RCP 8.5 data, respectively, which represents a significant increase in sample size compared to the block maxima approach. The sample sizes decrease when the threshold value is increased and when a minimum cluster separation distance is introduced, but on the other hand this makes the iid assumption more realistic.

In order to complete the description of the POT models, a model for the frequency of extreme events, i.e. threshold excesses, must be specified. Often, a Poisson process with rate parameter  $\lambda$  is assumed for such models, but in this study the annual frequency of extreme events has simply been estimated as the average annual number of excesses in the data. The results from the POT-modelling are presented in Tables 13 and 14 and the fitted distributions for the two threshold values and the two cluster distances are shown in Fig. 4. It is observed that the fitted densities for the historical data looks strange for  $u = 12$  m but this can be explained by the very low number of cluster maxima in this case, indicat-



**Fig. 4** Estimated POT models for the historic period and future projections for different thresholds [ $u = 10$  m (left) and  $u = 12$  m (right)] and different clustering distances. *Solid lines* correspond to a clustering of 2 days and the *dotted lines* correspond to clustering of 4 days

ing that a threshold of 12 m is probably a bit too high for this particular dataset.

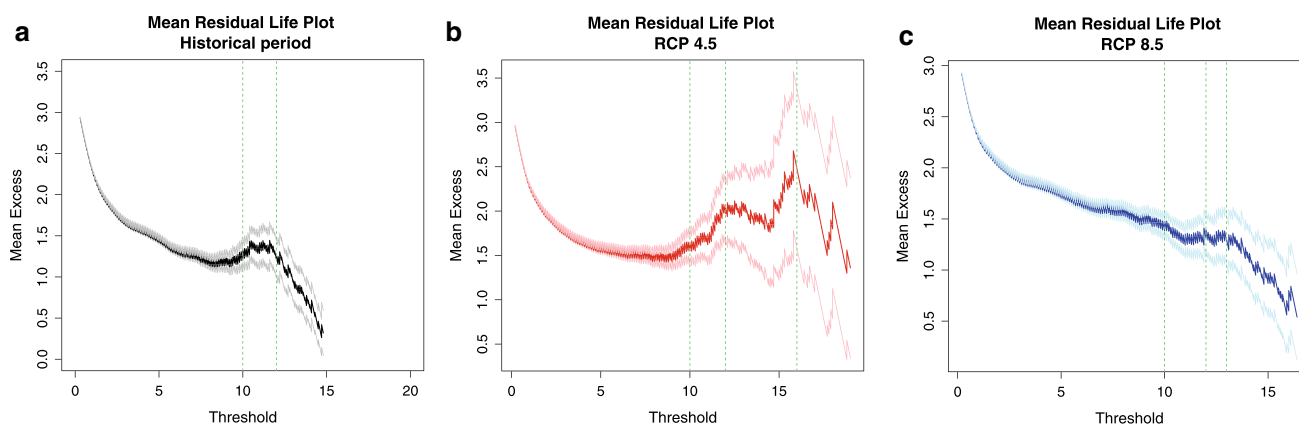
In summary, it is clear that the selection of threshold value influences the resulting model fits and the corresponding return value estimates. Differences of up to 2 m can be observed for the 100-year return value when changing the threshold value from 10 to 12 m (RCP 4.5 data). Also the length of the cluster separation distance influences the results somewhat, but the biggest difference is between models fitted with some clustering and models without clustering (Tables 13, 14). Nevertheless, both the threshold and the clustering distance are parameters that need to be specified and that introduce uncertainties in the results.

Even though the value estimates are different, the POT analyses are in general agreement with the block maxima approach in that the expected return values will be increasing in a future wave climate. However, the uncertainties are large and confidence intervals obtained by parametric bootstrapping indicate that the expected increases are statistically significant at the 5% level for most models with  $u = 12$  m, whereas the results from using  $u = 10$  m mostly indicate non-significant trends. The results also suggest that there is less variability between the different POT return value estimates compared to the block maxima method, indicating that the POT approach might be more accurate. However, it cannot be established whether it is in fact more correct without much longer time series.

### 3.3.1 Threshold selection with the POT-approach

In the POT analyses above, the same thresholds have been assumed for all three datasets, but it is not obvious that this is appropriate. The choice of threshold is a trade-off between having a large enough sample to get stable fits (reduce the variance) of the statistical model and being far enough out in the tail to ensure that only tail observations are included (reduce the bias). There are different methods for aiding the selection of thresholds and there are, for example, different graphical tools available. However, a more straightforward way to select threshold is to simply use a specified high percentile of the empirical distribution.

The threshold of 10 m used above corresponds to the 99.7-percentile in the historical dataset, the 99.5-percentile of the RCP 4.5 data and the 99.2-percentile of the RCP 8.5 data. The slightly higher threshold of 12 m corresponds to the 99.9-, 99.9- and 99.8-percentiles of the data, respectively. However, rather than selecting a threshold value directly, one could select an appropriate percentile and calculate the threshold in each dataset accordingly. Again, it is not obvious which percentile to use, but it should arguably be a high quantile. Selecting, for example, the 97-percentile would yield thresholds of 7.3, 7.5 and 7.9 m for the historical, RCP 4.5 and RCP 8.5 data, respectively. Increasing this to the 99.5-percentile would give thresholds at 9.5, 10.1 and 10.8 m, respectively. This is close to the initially chosen threshold of 10 m, which could still be a bit low for the data at hand. At any rate,

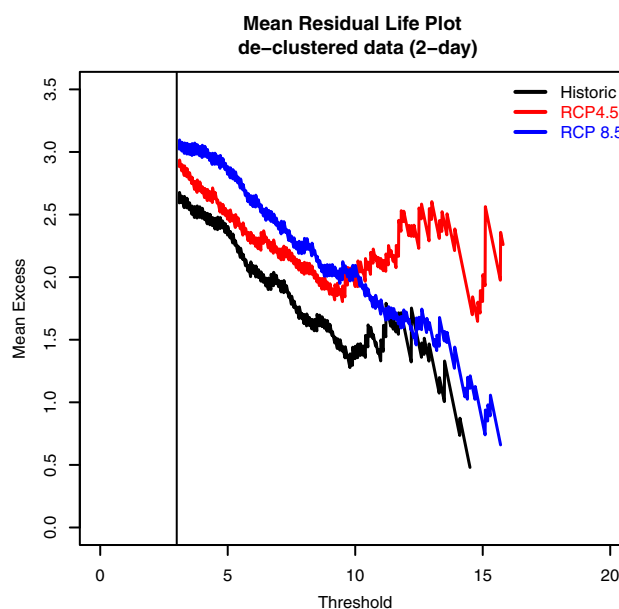


**Fig. 5** Mean residual life plots; historical data (*left*), scenario RCP 4.5 (*centre*) and scenario RCP 8.5 (*right*)

merely selecting a high percentile is as arbitrary as selecting a threshold directly and there is no universal guidance as to what percentiles are adequate.

The mean residual life plot plots the threshold  $u$  against the mean excess above that threshold. Such plots should be linear above the threshold for which the GPD model becomes appropriate. The mean residual life plots for the three datasets are shown in Fig. 5 for the original data without de-clustering. The thresholds used in the analyses above, i.e.  $u = 10$  m and  $u = 12$  m are indicated in the figures. It can be observed that even though a threshold of 12 m might be appropriate for the historical data, higher threshold would be more adequate for the future projections. For the RCP 4.5 data, a threshold as high as 16 m would be required to get approximately linear behaviour of the plot, whereas a threshold of 13 m seems adequate for the RCP 8.5 data. If these thresholds were selected for the data without clustering, one would have  $n = 64$  data points for the modelling of the historical data,  $n = 14$  for the RCP 4.5 data and  $n = 81$  for the RCP 8.5 data. However,  $n = 14$  is obviously too few to get a stable fit for the RCP 4.5 data, so the threshold selection is particularly challenging for this dataset. Taking the rather wide confidence bands of the RCP 4.5 plot into account, however, a threshold of 12 m could not have been rejected even for this dataset.

Using all the data above a certain threshold without clustering results in dependent data and is not appropriate. Hence, de-clustering should be performed to reduce the sample size. For example, applying a 2-day cluster length would effectively reduce the samples sizes of the excesses to  $n = 17$ ,  $n = 3$  and  $n = 25$ , respectively, for the historical, RCP 4.5 and RCP 8.5 data with the above thresholds. This is quite low and renders too few samples to get a stable fit. It was also observed that  $u = 12$  m might be too high threshold for the historical data. At any rate, if GPD models are fitted to excesses above 13 m and with a 2-day clustering, the estimated 20-year return values would be 15.20, 18.88 and 16.65 m, and the 100-year return value estimates would

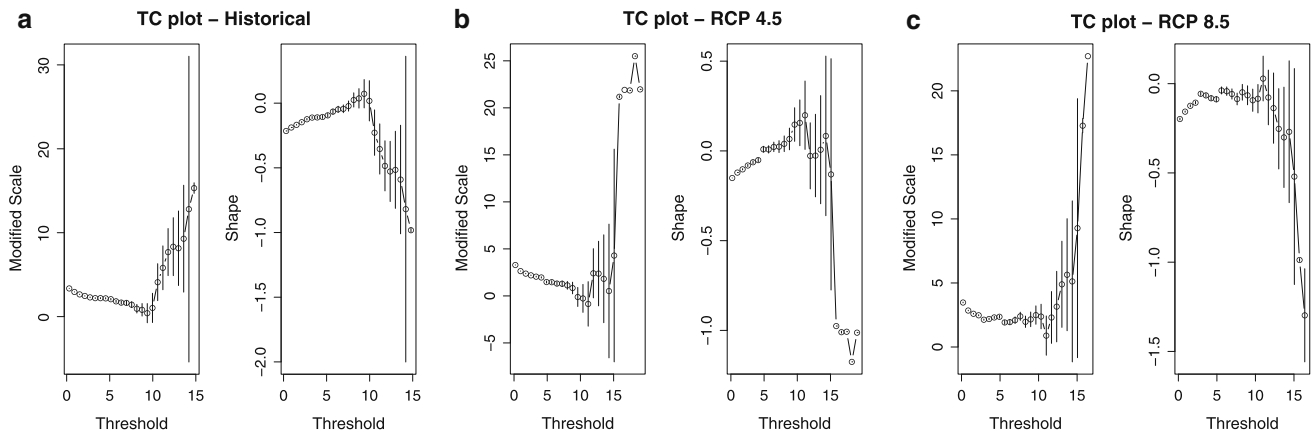


**Fig. 6** Mean residual life plot for the three datasets on de-clustered (independent) data with cluster separation of 2 days and cluster maxima above 3 m

be 15.52, 21.45 and 17.41 m, respectively, for the historical, RCP 4.5 and RCP 8.5 data. This is some centimetres different from the estimates obtained by  $u = 12$  m and again illustrates that the results are sensitive to this choice.

Figure 6 shows the mean residual life plot for applied to de-clustered data, where all cluster maxima with separation distance greater than 2 days for excesses above 3 m are included. The confidence bands are not included, but it can be seen that thresholds of 12 and 13 m for the historical data and the RCP 8.5 data still look reasonable but that it is different to read a reasonable threshold for the RCP 4.5 data.

Another tool for threshold selection is the threshold choice plots or tc-plots, sometimes also referred to as parameter stability plots. These plot the modified scale and the

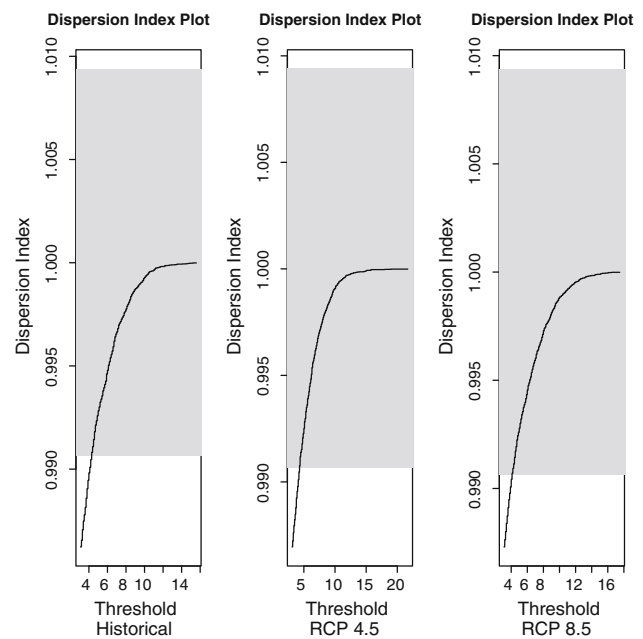


**Fig. 7** Threshold choice plots; historical data (left), scenario RCP 4.5 (centre) and scenario RCP 8.5 (right)

shape parameter for different thresholds and these should be constant for any values above the correct threshold value. Figure 7 shows tc-plots for the three datasets analysed in this report. Unfortunately, these plots are not very helpful in this case and there are no values for when the plots level out. This might indicate that there are no thresholds for which the GPD distribution is suitable, possibly due to violation of the iid assumption. Notwithstanding, taking the confidence intervals into account, these plot would not reject thresholds above approximately 12m for the historical and RCP 8.5 data, although a threshold above 15 m would be needed for the RECP 4.5 data. Such a high threshold would, however, render very few points for model fitting, as seen above.

Finally, the dispersion index plot plots values of the dispersion index, i.e. the ratio of the variance to the mean, for different threshold. For the Poisson distribution, which is often associated with the probability of threshold exceedance in a GPD model, this ration should be 1 and the appropriate threshold should correspond to a dispersion index not too far from this value. The dispersion index is plotted for each of the datasets in Fig. 8. The shaded area corresponds to the confidence interval and it can be seen that no threshold above a value of about 5 m would be rejected. However, the plots indicate that the dispersion index is very close to 1 for all thresholds above approximately 10 m for all datasets. Hence, from these plots a threshold of about 10 m would be selected.

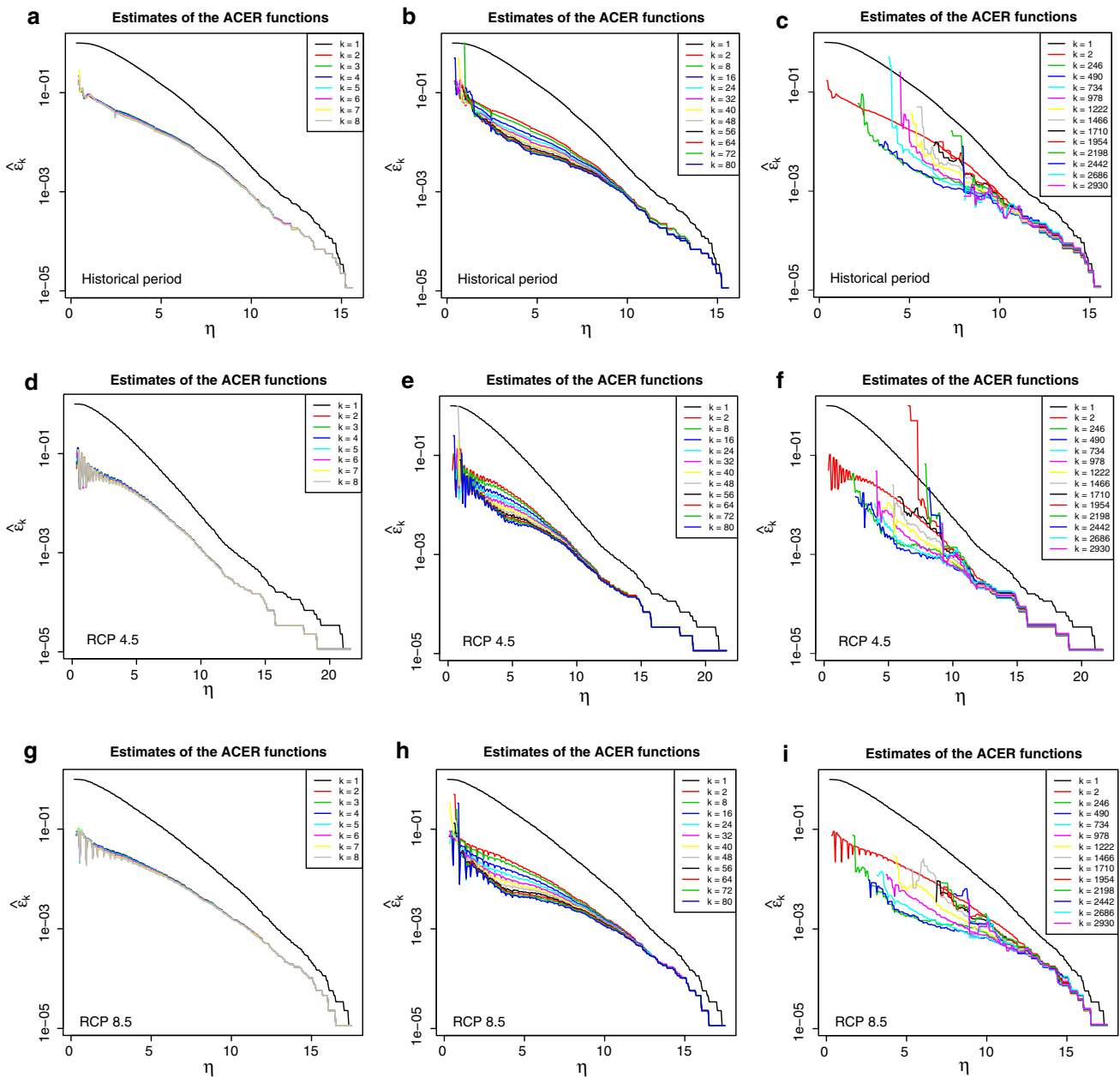
Some useful graphical tools for threshold selection are reviewed above, and it has been seen that they could all be used to guide in the selection of a proper threshold. However, the methods are not exact, and the exact value of an optimal threshold is not straightforward. Indeed, the different methods would suggest different thresholds. When one takes into account that the final results are very sensitive to the choice of threshold, this obviously adds uncertainty in the analyses. Notwithstanding, it can be concluded that the thresholds at 10 and 12 m used in the analysis are not unreasonable, and is supported by different threshold selection approaches.



**Fig. 8** Dispersion index plots for the three datasets on de-clustered (independent) data with cluster separation of 2 days and cluster maxima above 3 m; historical data (left), RCP 4.5 (centre) and RCP 8.5 (right)

### 3.4 The ACER method

A final approach to analysing the extreme values on the data that was tried out in this study is the average conditional exceedance rate method or the ACER method (Naess and Gaidai 2009; Naess et al. 2013). The ACER method relaxes the independent data assumptions and accounts for the dependence by conditioning on previous data points in the time series, where  $k$  is a parameter to be chosen reflecting the  $(k-1)$ -step memory of the data. A value of  $k = 1$  corresponds to independent data,  $k = 2$  corresponds to conditioning on the preceding value only (1-step memory) and a value of  $k = 3$

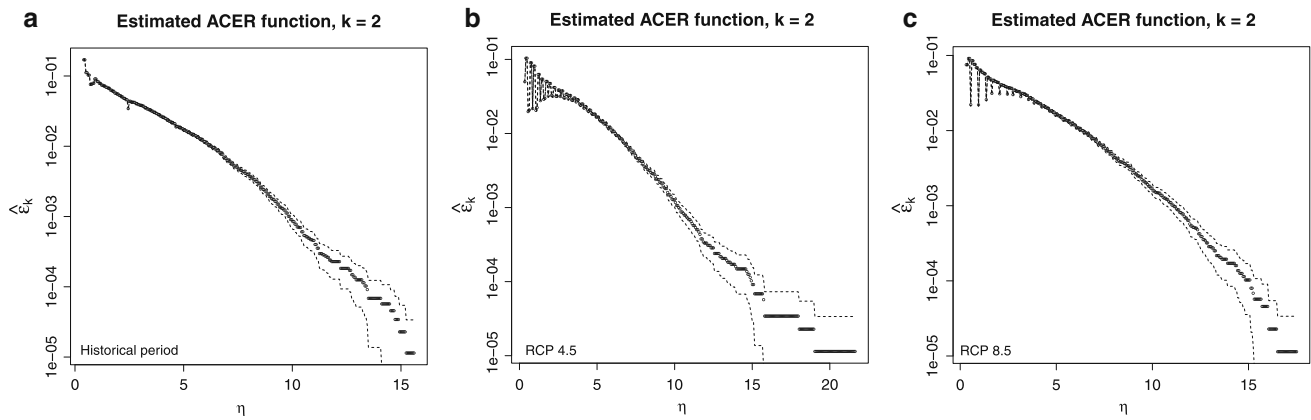


**Fig. 9** Estimated ACER functions for different  $k$ -step memory assumptions; historical data (*top*), scenario RCP 4.5 (*middle*) and scenarios RCP 8.5 (*bottom*). The  $x$ -axis corresponds to significant wave height and the  $y$ -axis is the estimated ACER functions

corresponds to conditioning on the two preceding values (2-step memory) and so on. Typically  $k \geq 2$  will be assumed.

One of the first tasks in an ACER analysis is to determine the value of  $k$ . Different values of  $k$  will give different empirical estimations of the average conditional exceedance rates (ACER). Initial empirical estimates of the ACER-functions of different orders  $k$  are shown in Fig. 9, for the historical time series and for the future projections. The leftmost figures show estimated ACER functions for  $k$ -values ranging from 1 to 8, corresponding to modelling the dependence conditional on 0, 1, ..., 7 preceding values. The estimated functions are

very different for  $k = 1$  and the others, but there is not much difference between  $k = 2$  and  $k > 2$ . This indicates that the dependence structure is well approximated by a one-step memory process. The time series investigated in this analysis have a temporal resolution of 3 h, meaning that 8 observations correspond to 1 day. However, in climate-related data, there can be expected to be diurnal or seasonal dependence structures, and it will be necessary to investigate whether higher values of  $k$  are appropriate. Hence, the middle column of plots in Fig. 9 shows the estimated ACER-functions for  $k$ -values up to 80, corresponding to a memory of up to 10 days and



**Fig. 10** The estimated ACER functions for  $k = 2$  with 95 % confidence bands; historic period (*left*), RCP 4.5 (*middle*) and RCP 8.5 (*right*)

the right column shows the ACER functions for  $k$ -values corresponding to about 1 month to 1 year memory. This shows somewhat larger differences between the estimated ACER-functions, indicating that the dependence is not completely captured by an ACER function of order 2. However, for the larger  $\eta$ -values, the estimated ACER functions behave rather similarly. It is, therefore, assumed that a value of  $k = 2$  is sufficient for the ACER analysis, and this value will be adopted in the following. The estimated ACER functions with  $k = 2$  are shown in Fig. 10, including the 95 % confidence intervals.

In order to estimate return values for high return periods, there is a need to extrapolate the tail behaviour, and a nonlinear function is fitted to the estimated empirical ACER functions, parameterized by the parameters  $q, a, b, c$  (see (Naess et al. 2013) for details). In order to do this, there is a need to specify a tail marker for when the tail of the distribution begins. In this paper, analyses for tail markers at 8, 10 and 12 m are presented. The fitted ACER functions for the different data sets are shown in Fig. 11. The estimated parameters of the nonlinear function and the resulting return value estimates are presented in Table 15. Even if there are large uncertainties in the estimated return value estimates, and the wide confidence intervals suggest that the climatic trends are not statistically significant, there seems to be a fairly consistent positive trend in the future projections compared to the historical data.

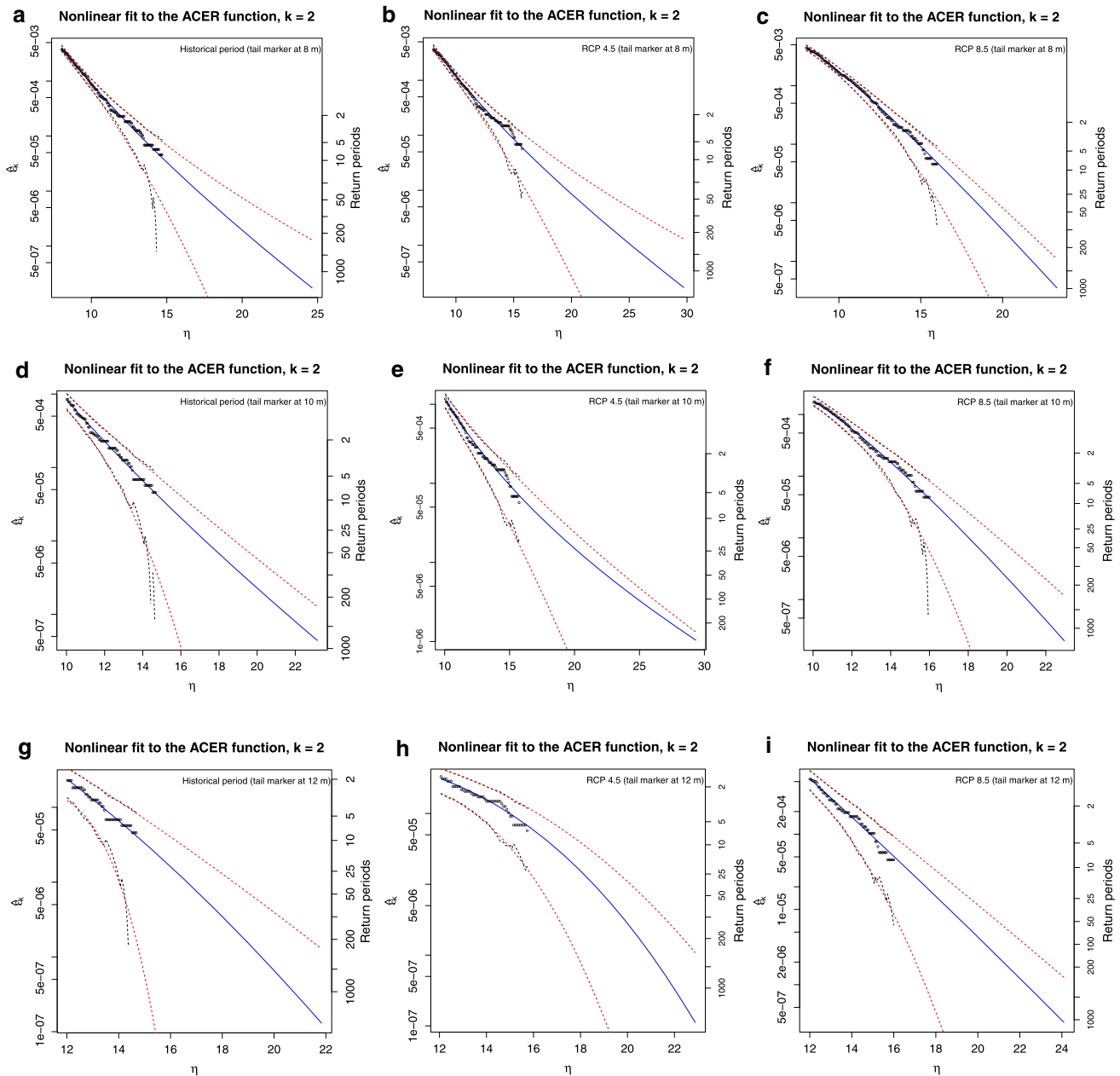
It appears that the results are quite sensitive to the chosen value of the tail marker, and this is somewhat troublesome. The estimated return values are different and it can be seen in Fig. 11 that the curvature of the estimated ACER functions are indeed different depending on the tail marker. The choice of tail marker would be a trade-off between making sure that the estimated nonlinear function is indeed fitted to the tail area (the value of the tail marker should be sufficiently high) and having enough samples to get a stable fit (the value of the tail marker should not be too high). It is noted that the sensitivity due to the parameter  $k$  has not been evaluated;

only  $k = 2$  has been applied. Presumably, the results would display some variability for different values of this parameter as well.

#### 4 Comparison of results

To summarize this investigation on the uncertainty related to the estimation of extreme values, conditioned on a particular data-set, the different return value estimates are repeated in Tables 16 and 17. The first table shows the different results from the initial distribution method, i.e. fitting a model to all the data and estimating return values by extrapolation, and the second table includes the various estimates from dedicated extreme value analyses. Some additional results from analyses not covered in this paper are also included, i.e. using 10-day clusters in the POT approach and increasing the tail marker to 14 m for the ACER method. Included in the tables are also the minimum and maximum estimates for each return value as well as the arithmetic mean and the standard error of the various estimates. The minimum and maximum estimates for each dataset are presented in bold in the tables to clearly demonstrate which models give the highest and lowest estimates. It is easily observed that the uncertainties of the return value estimates are considerable. As could be expected, the variability of the estimates is higher for the 100-year return value estimates than for the 20-year return values.

First, it is noted that there are large variabilities depending on which approach to extreme value analysis is chosen. It is interesting to observe that there are no systematic biases regarding the methods. For the historical data, the ACER method seems to give the highest return value estimates, for the RCP 4.5 scenario both the POT and the GEV models give the highest return value estimates and for the RCP 8.5, the Gumbel model gives the highest estimates. However, for some data sets and some return values (20- or 100-years), the POT-, the ACER and the Weibull-



**Fig. 11** Fitted nonlinear ACER-functions for different values of the tail marker; 8 m (*top row*), 10 m (*second row*) and 12 m (*bottom row*); historical period (*left*), RCP 4.5 (*middle*) and RCP 8.5 (*right*)

models yield the lowest return value estimates. Hence, it is difficult to associate particular bias with particular approaches. For the various initial distribution models it seems clear that the 2-parameter Weibull distribution consistently gives the lowest return values, whereas the distribution fitted by means of the Anderson–Darling statistic gives the highest estimates. At any rate, the bottom line is that there is great variability according to which approach is adopted.

Considering the block maxima approach, it is seen that the return value estimates are sensitive to the block-size. The

variability due to the block size appears larger if one assumes a Gumbel model compared to the full GEV model and is again larger for the 100-year return value than for the 20-year return value. For the 100-year return value estimate for the RCP 4.5 data, the different block sizes give rise to quite significant differences of up to 3.5 m. Furthermore, whether one uses the full GEV model or the reduced Gumbel model appears to have a notable influence on the estimates, even with the same block size. Most extreme is again the RCP 4.5 data set, where a reduction from the full GEV model to the Gumbel model for annual maximum data yields a reduction



**Table 15** Estimated parameters for the nonlinear fit to the ACER functions and associated return value estimates with different values for the tail markers

Tail marker	Historical period			RCP 4.5			RCP 8.5		
	8 m	10 m	12 m	8 m	10 m	12 m	8 m	10 m	12 m
q	0.6896	0.3814	0.0083	0.4926	0.8352	0.0007	0.0045	0.0027	0.1299
b	3.6151	3.3060	2.7906	2.9829	5.3405	3.9721	7.9730	8.7995	1.1569
a	1.8375	1.5061	0.1102	1.6129	3.3229	0.0025	0.4234	0.3698	0.3555
c	0.6937	0.7381	1.5642	0.6760	0.4433	2.7708	1.1375	1.2226	1.1470
20-year return value	16.09 m	16.28 m	15.88 m	18.38 m	19.53 m	17.83 m	17.57 m	17.26 m	17.74 m
95 % CI	(13.73, 17.97)	(13.70, 17.69)	(13.75, 17.47)	(15.38, 20.88)	(15.68, 20.92)	(15.67, 19.30)	(15.58, 18.70)	(15.22, 18.63)	(15.43, 19.20)
100-year return value	18.95 m	19.23 m	18.01 m	22.14 m	24.84 m	19.78 m	20.01 m	19.45 m	20.36 m
95 % CI	(15.13, 21.99)	(14.77, 21.06)	(14.44, 20.23)	(17.26, 26.21)	(17.71, 26.07)	(17.03, 21.53)	(17.07, 21.44)	(16.37, 21.25)	(16.67, 22.24)

**Table 16** Summary of estimated return values (m) from initial distribution methods

Model	Historical period		RCP 4.5		RCP 8.5	
	20-year return value	100-year return value	20-year return value	100-year return value	20-year return value	100-year return value
2p Weibull	<b>12.38</b>	<b>13.28</b>	<b>13.35</b>	<b>14.40</b>	<b>14.46</b>	<b>15.58</b>
3p Weibull (ML)	13.05	14.07	13.88	15.03	15.17	16.53
3p Weibull (MoM)	15.63	17.17	17.49	19.40	18.49	20.54
3p Weibull (LM)	15.39	16.89	16.59	18.30	18.33	20.35
3p Weibull (CvM)	14.88	16.28	16.03	17.63	17.68	19.58
3p Weibull (AD)	<b>15.83</b>	<b>17.41</b>	<b>18.19</b>	<b>20.27</b>	<b>18.60</b>	<b>20.67</b>
Minimum	12.38	13.28	13.35	14.40	14.46	15.58
Average	14.53	15.85	15.92	17.51	17.12	18.88
Maximum	15.83	17.41	18.19	20.27	18.60	20.67
Standard error	1.45	1.74	1.94	2.35	1.83	2.24

in the 20-year return value estimate of 1.7 m and a reduction of almost 6.4 m for the 100-year return value. It should be noted, however, that the data did not support the Gumbel reduction in this case.

Regarding the POT-approach, variations related to the threshold value and the clustering separation were considered. It is observed that the estimates are quite sensitive to the threshold selection, with differences up to more than 2 m for different thresholds (RCP 4.5 data). Furthermore, whether a de-clustering technique is applied or not leads to quite different results. Again, the largest difference is for the RCP 4.5 data, with estimates differing up to 2.4 m depending on whether clustering has been considered or not. Assuming that de-clustering is included, different cluster separation distances also give different results even though the sensitivity to the exact choice of cluster separation distance seems to be smaller than the sensitivity to the threshold value.

Finally, the ACER method is found to be sensitive to the choice of value for the tail marker. For some of the estimates, a low value of the tail marker gives higher return

value estimates and for some of the estimates it is opposite. Again, the variability is greatest for the RCP 4.5 data set, with differences up to 3 m for the 20-year return value and 7.6 m for the 100-year return value for different values of the tail marker. The variability due to the parameter *k* has not been evaluated, and the results would presumably also display some variability for different values of this parameter.

An interesting observation regarding the future scenarios is that whether one analyses all the data or only the extremes, i.e. block maxima or values above a certain threshold, influences the relative trends in the two scenarios. By comparing the estimated return periods in Tables 16 and 17 it is observed that based on all the data, the estimated increase in return values is largest for the RCP 8.5 scenario. However, if only the extreme data are considered, the RCP 4.5 scenario is ascribed the largest increase in the return values. This could be an effect of a few extreme outliers in the RCP 4.5 dataset, which may have a much stronger influence on the fits based on the reduced dataset of extremes compared to the full dataset. Whether it is correct or not to put a strong

**Table 17** Summary of estimated return values (m) from extreme value analysis methods

Model	Historical period		RCP 4.5		RCP 8.5	
	20-year return value	100-year return value	20-year return value	100-year return value	20-year return value	100-year return value
GEV (annual)	15.33	17.34	19.66	<b>27.55</b>	16.74	18.65
GEV (biennial)	15.04	15.58	18.66	21.98	16.58	17.47
GEV (semi-annual)	15.20	17.05	18.85	24.11	16.67	17.94
Gumbel (annual)	15.47	17.75	17.96	21.19	17.00	19.43
Gumbel (biennial)	15.80	18.53	18.80	22.58	17.30	20.02
Gumbel (semi-annual)	15.65	18.07	17.41	20.37	<b>18.43</b>	<b>21.64</b>
Weibull (2-par)	<b>14.95</b>	15.84	18.32	20.09	<b>16.41</b>	<b>17.35</b>
Weibull (3-par)	15.37	17.11	19.25	23.68	16.91	18.83
Fréchet			19.66	<b>27.55</b>		
POT, $u = 10$ m, 10-day cluster	15.13	16.88	18.76	23.13	16.54	17.45
POT, $u = 12$ m, 10-day cluster	15.18	<b>15.53</b>	18.72	21.66	16.56	17.51
POT, $u = 10$ m, 4-day cluster	15.18	17.12	18.89	23.67	16.52	17.50
POT, $u = 12$ m, 4-day cluster	15.18	<b>15.53</b>	18.71	21.62	16.57	17.53
POT, $u = 10$ m, 2-day cluster	15.28	17.38	18.80	23.51	16.51	17.55
POT, $u = 12$ m, 2-day cluster	15.14	15.53	18.71	21.62	16.56	17.54
POT, $u = 10$ m, no cluster	15.95	17.27	<b>21.14</b>	25.36	17.12	18.24
POT, $u = 12$ m, no cluster	15.24	15.58	20.57	23.39	17.12	18.04
ACER, $\eta = 10$ m	16.28	<b>19.23</b>	19.53	24.84	17.26	19.45
ACER, $\eta = 12$ m	15.88	18.01	17.83	19.78	17.74	20.36
ACER, $\eta = 8$ m	16.09	18.95	18.38	22.14	17.57	20.01
ACER, $\eta = 14$ m	<b>16.36</b>	18.80	<b>16.48</b>	<b>17.21</b>	16.88	18.38
Minimum	14.95	15.53	16.48	17.21	16.41	17.35
Average	15.49	17.15	18.77	22.47	16.95	18.54
Maximum	16.36	19.23	21.14	27.55	18.43	21.64
Standard error	0.43	1.23	1.01	2.26	0.51	1.23

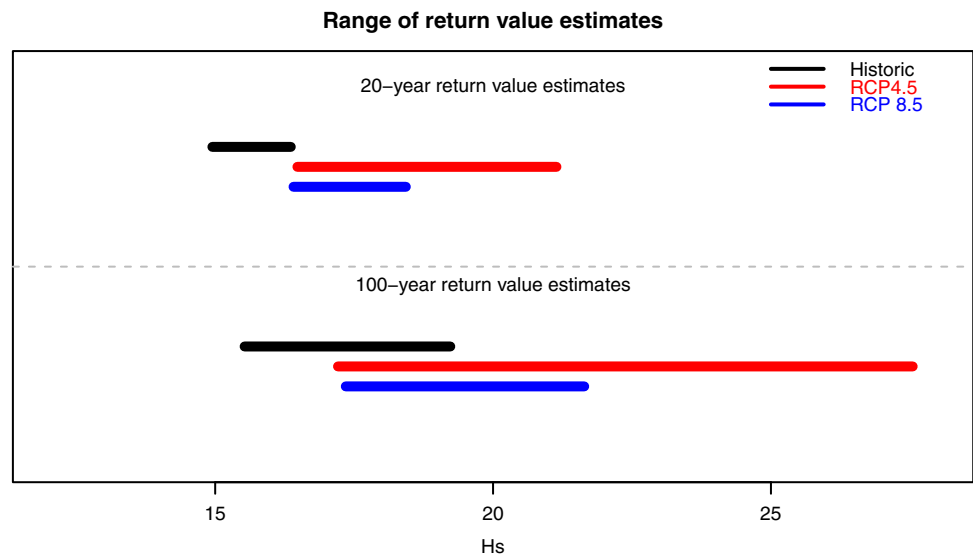
weight on these extreme observations is not straightforward to determine without knowing more about how the data are obtained.

It is also noted that the statistical uncertainty of the return value estimates generally increases significantly for the extreme value analyses techniques compared to the initial distribution approach. This can be seen from the various parametric bootstrap estimates of the uncertainties. This is presumably due to the notable decrease in sample size when only using the extreme data points and should, therefore, be expected; see also the discussion on uncertainties as a function of sample size in Wang et al. (2013). However, one implication of this is that the climate change signal in the return values are statistically significant at the 5% level for estimates obtained from all the data (initial distribution approach) but ceases to be so for most of the dedicated extreme value analysis approaches. This makes it even more difficult to conclude on whether there is a statistically significant increase in

high return values of significant wave height in a future scenario.

It may be illustrative to plot the range of point estimates for the 20- and 100-year return values from the different methods, and this is shown in Fig. 12. As can be seen, there are quite large spreads of the different estimates, and the variability is larger for the 100-year return values than for the 20-year return values. Interestingly, for the 20-year return value estimates there is little overlap between the historical period and the future projections, and even though there are large modelling uncertainties with regard to the actual return values, there seem to be higher 20-year return values in a future climate. This trend is also evident for the 100-year return value, although there is much more variability associated with these estimates and the estimates for future climates are partly overlapping with estimates for the historical (current) climate. Moreover, the return value estimates for the RCP 4.5 scenario has more variability than the estimates for the RCP 8.5 scenario and the historical period.

**Fig. 12** Range of estimates for the different return values from the different extreme value analysis methods



### 5 Discussion

According to this study, there are large uncertainties involved when extreme values corresponding to high return periods are extracted from a finite set of data. In the most extreme case, estimates of the same return value using the very same set of data differed by as much as 13.15 m (100-year return value for the RCP 4.5 data). This is troublesome and it is deemed difficult to establish which estimates are most accurate without much longer data records. Indeed, many of the extreme value methods are only valid asymptotically and hence only approximately for finite time series. Moreover, the statistical uncertainty of return value estimates is much larger for the methods that only utilize the extreme data points since the sample size is drastically reduced compared to the methods that use all the data.

Another important issue is that many of the methods assume that the data, or at least the extremes, are iid. In particular, for the initial distribution approach this assumption is clearly unrealistic. Moreover, assuming that there are long-term trends in the wave climate due to climate change this may no longer be the case also for the extremes and the strong theoretical foundation of the extreme value theories falls apart. This suggests that non-stationary modelling would be more appropriate, see, e.g. Wang et al. (2004, 2013), and Vanem (2013). Nevertheless, often these changes are assumed to be negligible over limited time periods and standard extreme value analysis methods are still applied. In this paper it is tacitly assumed that the effect of climate change is negligible within the individual 30-year periods of data that have been analysed, even though it might not be negligible over longer time periods. If the effect of the long-term trend is small compared to the other variability, this might not be an overly unrealistic approximation but it

should be mentioned that assuming a stationary model might influence the results.

It should be noted that some of the spread in the return value estimates presented in this paper might be artificial and that some of the estimates might have been eliminated by further scrutiny and inspection of the model fits.

This paper is only concerned with the extremes of significant wave height, which is a parameter describing a short-term sea state rather than individual waves. In many applications, one would also need to know the extreme values of individual waves or wave crests, and this adds further uncertainty to the return value estimates. One often assume that the sea surface can be considered a stationary process over a limited period of time, say from 20 min to a few hours, and apply a statistical model for the individual wave or crest heights conditioned on the sea state, for example, the Rayleigh distribution (Rydén 2006). One could then use a certain quantile of this distribution to estimate extreme individual wave heights. However, it is not obvious which quantile to use and the fact that the most extreme individual wave might not appear in the most extreme sea state should be accounted for. Indeed, the probability distribution of extreme individual waves will typically have contributions from a range of different sea states. This has not been considered in this study, but would presumably add to the uncertainty of the extreme value estimation. See, e.g. Forristall (2008) for discussions on how short-term variability can be included in extreme value estimates.

In many marine engineering applications there will be several environmental parameters that need to be taken into account jointly (Bitner-Gregersen 2015) and estimates of multivariate extremes associated with long return periods will be needed. Obviously, this would complicate the picture even more, and there are even different ways of defining

what is meant by a multivariate extreme. Hence, uncertainties would presumably increase as the complexity increases with the dimensions of the multivariate extreme value problem.

Finally, it is emphasized that the results presented in this paper are conditioned on a particular dataset for a particular location in the North Atlantic. Hence, the identified trends in the extreme wave climate must be verified by other datasets before conclusions can be made. In particular, it has been demonstrated in previous studies that the effect of climate change on the wave climate is highly location dependent (Vanem 2014; Wang et al. 2014). Therefore, the results from one arbitrary location cannot be used to confidently inform about climatic trends. The aim of this paper, however, was to highlight the uncertainty in estimating weather and climate extremes associated with long return periods from a finite dataset by applying different methods to the same dataset of historical data and future projections.

## 6 Summary and conclusions

This paper has revealed that there are large uncertainties associated with the estimation of return values for high return periods in weather and climate data. It is difficult to single out one method or approach that is best overall, and this is out of scope for the current study. Presumably, longer data records would be needed to investigate this.

In spite of the large variability of the different return value estimates, there seems to be evidence in the data for a notable increase in the extreme wave heights for both of the future scenarios that have been considered compared to the historical period. Hence, these data indicate that there will be a general trend towards more extreme wave events in a future climate. Obviously, this result is conditioned on the particular dataset that has been analysed and for a particular location in the North Atlantic Ocean. Further studies are needed to confirm or refute these trends.

**Acknowledgments** The work presented in this paper has been carried out with support from the Research Council of Norway (RCN) within the research project ExWaCli. The data used in the analysis have been kindly provided by the Norwegian Meteorological Institute.

## References

- Aarnes O, Breivik Ø, Reistad M (2012) Wave extremes in the Northeast Atlantic. *J Clim* 25:1529–1543
- An Y, Pandey M (2005) A comparison of methods of extreme wind speed estimation. *J Wind Eng* 93:535–545
- Bitner-Gregersen EM (2015) Joint met-ocean description for design and operation of marine structures. *Appl Ocean Res*. doi:10.1016/j.apor.2015.01.007
- Caires S, Sterl A (2005) 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data. *J Clim* 18:1032–1048
- Coles S (2001) An introduction to statistical modeling of extreme values. Springer, London
- Coles SG, Powell EA (1996) Bayesian methods in extreme value modelling: a review and new developments. *Int Stat Rev* 64:119–136
- Coles S, Pericchi LR, Sisson S (2003) A fully probabilistic approach to extreme rainfall modeling. *J Hydrol* 273:35–50
- de Winter R, Sterl A, Ruessink B (2013) Wind extremes in the North Sea Basin under climate change: an ensemble study of 12 CMIP5 GCMs. *J Geophys Res: Atmos* 118:1601–1612
- Donner LJ, Wyman BL, Hemler RS, Horowitz LW, Ming Y, Zhao M, Golaz J-C, Ginoux P, Lin SJ, Schwarzkopf MD, Austin J, Alaka G, Cooke WF, Delworth TL, Freidenreich SM, Gordon CT, Griffies SM, Held IM, Hurlin WJ, Klein SA, Knutson TR, Lagenhorst AR, Lee H-C, Lin Y, Magi BI, Malyshev SL, Milly PCD, Naik V, Nath MJ, Pincus R, Ploshay JJ, Ramaswamy V, Seman CJ, Shevliakova E, Sirutis JJ, Stern WF, Stouffer RJ, Wilson RJ, Winton M, Wittenberg AT, Zeng F (2011) The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *J Clim* 24:3484–3519
- Forristall G (2008) How should we combine long and short term wave height distributions? In: Proceedings of 27th international conference on Offshore mechanics and Arctic engineering (OMAE 2008). American Society of Mechanical Engineers (ASME), Portugal, Estoril
- Gibson R, Forristall GZ, Owrid P, Grant C, Smyth R, Hagen Ø, Leggett I (2009) Bias and uncertainty in the estimation of extreme wave heights and crests. In: Proceedings of the 28th international conference on ocean, offshore and Arctic Engineering (OMAE 2009). American Society of Mechanical Engineers (ASME), Honolulu, HI, USA
- Grabemann I, Weisse R (2008) Climate change impact on extreme wave conditions in the North Sea: an ensemble study. *Ocean Dyn* 58:199–212
- Grabemann I, Groll N, Möller J, Weisse R (2015) Climate change impact on North Sea wave conditions: a consistent analysis of ten projections. *Ocean Dyn* 65:255–267
- Hagen Ø (2009) Estimation of long term extreme waves from storm statistics and initial distribution approach. In: Proceedings of the 28th international conference on ocean, offshore and Arctic Engineering (OMAE 2009). American Society of Mechanical Engineering (ASME), Honolulu, HI, USA
- Harris R (2001) The accuracy of design values predicted from extreme value analysis. *J Wind Eng* 89:153–164
- ISO (2005) Petroleum and natural gas industries—specific requirements for offshore structures—part 1: metocean design and operating considerations. International Standards Organization, Geneva
- Jonathan P, Ewans K (2013) Statistical modelling of extreme ocean environments for marine design: a review. *Ocean Eng* 62:91–109
- Li L, Li P, Liu Y (2014) How we determine the design environmental conditions and how they impact the structural reliabilities. In: Proceedings of the 33rd international conference on ocean, offshore and Arctic Engineering (OMAE 2014). American Society of Mechanical Engineers (ASME), San Francisco, CA, USA
- Makkonen L (2008) Problems in extreme value analysis. *Struct Saf* 30:405–419
- Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, Meehl GA, Mitchell JFB, Nakicenovic N, Riahi N, Smith SJ, Stouffer RJ, Thomson AM, Weyant JP (2010) The next generation of scenarios for climate change research and assessment. *Nature* 463:747–756
- Naess A, Gaidai O (2009) Estimation of extreme values from sampled time series. *Struct Saf* 31:325–334
- Naess A, Karpa O (2013) Statistics of extreme wind speeds and wave heights by the bivariate ACER method. In: Proceedings of 32nd international conference on ocean, offshore and Arctic Engineer-

- ing (OMAE 2013). American Society of Mechanical Engineers (ASME), Nantes, France
- Naess A, Gaidai O, Karpa O (2013) Estimation of extreme values by the average conditional exceedance rate method. *J Probab Stat* 2013. Art. ID 797014. <http://www.hindawi.com/journals/jps/2013/797014/>
- Oliver EC, Wotherspoon SJ, Holbrook NJ (2014) Estimating extremes from global ocean and climate models: a Bayesian hierarchical model approach. *Prog Oceanogr* 122:77–91
- Reistad M, Breivik Ø, Haakenstad H, Aarnes O, Furevik B, Bidlot J-R (2011) A high-resolution hindcast of wind and waves for the North Sea, the Norwegian Sea and the Barents Sea. *J Geophys Res* 116:C05019
- Rychlik I, Rydén J, Anderson CW (2011) Estimation of return values for significant wave height from satellite data. *Extremes* 14:167–186
- Rydén J (2006) A note on asymptotic approximations of distributions for maxima of wave crests. *Stoch Environ Res Risk Assess* 20:238–242
- Thevasiyani T, Perera K (2014) Statistical analysis of extreme ocean waves in Galle, Sri Lanka. *Weather Clim Extrem* 5–6:40–47
- Towe R, Eastoe E, Tawn J, Wu Y, Jonathan P (2013) The extremal dependence of storm severity, wind speed and surface level pressure in the northern North Sea. In: *Proceedings of the 32nd international conference on ocean, offshore and Arctic Engineering (OMAE 2013)*. American Society of Mechanical Engineers (ASME), Nantes, France
- van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK (2011) The representative concentration pathways: an overview. *Clim Chang* 109:5–31
- Vanem E (2013) Bayesian hierarchical space–time models with application to significant wave height. Springer, Heidelberg
- Vanem E (2014) Spatio-temporal analysis of NORA10 data of significant wave height. *Ocean Dyn* 64:879–893
- Vanem E (2015) Uncertainties in extreme value analysis of wave climate data and wave climate projections. In *press for Proceedings of the 34rd international conference on ocean, offshore and Arctic Engineering (OMAE 2015)*. St. John's, NL, Canada
- Wang XL, Swail VR (2006) Climate change signal and uncertainty in projections of ocean wave heights. *Clim Dyn* 26:109–126
- Wang XL, Feng Y, Swail V (2012) North Atlantic wave height trends as reconstructed from the 20th century reanalysis. *Geophys Res Lett* 39:L18705
- Wang XL, Feng Y, Swail VR (2014) Changes in global ocean wave heights as projected using multimodel CMIP5 simulations. *Geophys Res Lett* 41:1026–1034
- Wang XL, Trewin B, Feng Y, Jones D (2013) Historical changes in Australian temperature extremes as inferred from extreme value distribution analysis. *Geophys Res Lett* 40:1–6
- Wang XL, Zwiers FW, Swail VR (2004) North Atlantic ocean wave climate change scenarios for the twenty-first century. *J Clim* 17:2368–2383