



# On the use of discrete seasonal and directional models for the estimation of extreme wave conditions

Edward B. L. Mackay<sup>c,\*</sup>, Peter G. Challenor<sup>b</sup>, AbuBakr S. Bahaj<sup>a</sup>

<sup>a</sup> Sustainable Energy Research Group, School of Civil Engineering and the Environment, University of Southampton, Southampton SO17 1BJ, UK

<sup>b</sup> Ocean Observing and Climate Group, National Oceanography Centre, Southampton SO14 3ZH, UK

<sup>c</sup> Garrad Hassan & Partners Ltd., St Vincent's Works, Silverthorne Lane, Bristol BS2 0QD, UK

## ARTICLE INFO

### Article history:

Received 18 July 2009

Accepted 29 January 2010

Available online 4 February 2010

### Keywords:

Extremes

Peaks-over-threshold

Generalised Pareto distribution

Seasonal model

Directional model

## ABSTRACT

Extreme value theory is commonly used in offshore engineering to estimate extreme significant wave height. To justify the use of extreme value models it is of critical importance either to verify that the assumptions made by the models are satisfied by the data or to examine the effect violating model assumptions. An important assumption made in the derivation of extreme value models is that the data come from a stationary distribution. The distribution of significant wave height varies with both the direction of origin of a storm and the season it occurs in, violating the assumption of a stationary distribution. Extreme value models can be applied to analyse the data in discrete seasons or directional sectors over which the distribution can be considered approximately stationary. Previous studies have suggested that models which ignore seasonality or directionality are less accurate and will underestimate extremes. This study shows that in fact the opposite is true. Using realistic case studies, it is shown that estimates of extremes from non-seasonal models have a lower bias and variance than estimates from discrete seasonal models and that estimates from discrete seasonal models tend to be biased high. The results are also applicable to discrete directional models.

© 2010 Elsevier Ltd. All rights reserved.

## 1. Introduction

Accurate estimates of extreme wave conditions are required for the design of offshore structures such as platforms, wind turbines or wave energy converters. One of the most common approaches used at present for modelling extreme wave conditions is the peaks over threshold (POT) method (e.g. Mathiesen et al., 1994; Ferreira and Guedes Soares, 1998; Naess, 1998; Caires and Sterl, 2005).

The POT model is based on results from extreme value theory. A good introduction to extreme value theory is given by Coles (2001) and the reader is referred here for the mathematical details. Roughly speaking, the theory states that for any random variable from a distribution which satisfies certain regularity conditions, the distribution of exceedances of a high enough threshold will tend to a member of the generalised Pareto family. In the POT method the distribution of storms where the significant wave height,  $H_s$ , exceeds a threshold is modelled by the generalised Pareto distribution (GPD). To justify the use of the POT model for extrapolating the data outside the range of observations it is important to verify the assumptions made by the model or to examine the effect of violating the assumptions.

The POT model assumes that the threshold crossings and the values of the exceedances are independent and identically distributed, and that the threshold is sufficiently high so that the asymptotic argument used to derive the model is a reasonable approximation.

Wave conditions exhibit strong serial correlation and exceedances of a high threshold tend to occur in clusters. The assumption of independence can be relaxed by requiring independence of extremes which are sufficiently separated in time (see e.g. Leadbetter et al., 1983). This is entirely plausible for wave data, with the occurrence of storms separated by several days being roughly independent. Short-term dependence is dealt with by declustering. That is, only the maximum value in a single storm is used, rather than all data points within a storm above the threshold.

The distribution of  $H_s$  varies with other variables, known as covariates, such as the time of year, direction of origin of the storm, or climatic variables such as the North Atlantic Oscillation index. This violates the assumption that the data are identically distributed. Many studies ignore covariate effects and use a stationary model to estimate extremes, but it is also possible to use models which take into account covariate effects. It has been suggested by several authors (e.g. Morton et al., 1997; Anderson et al., 2001; Jonathan et al., 2008) that covariate models will give more accurate results since they explain more of the variability in the data and because the data violates assumptions made in the

\* Corresponding author.

E-mail address: [ed.mackay@garradhassan.com](mailto:ed.mackay@garradhassan.com) (E.B. L. Mackay).

stationary model. To date, however, it has not been demonstrated conclusively which approach gives more accurate results.

Carter and Challenor (1981) considered estimation of return values from a population composed of a number of distinct homogeneous sub-populations. They prove that when the distribution in each sub-population is known, the return values calculated from a random sampling of the entire population are less than or equal to those obtained when the sub-populations are sampled proportionally, with equality only when the distributions in each sub-population are equal. This argument has often been used to support the use of covariate models. However, in practice the distribution of each sub-population is not known and must be estimated from the data. Moreover, over a long enough period, sampling from each sub-population will be approximately proportional.

The simplest approach to modelling seasonal or directional effects is to analyse the data in discrete seasons or directional sectors over which the distribution can be considered approximately stationary. Morton et al. (1997) analysed data from the North Sea in four separate seasons and compared the results to those from a non-seasonal analysis. They show that the value of the likelihood function for the seasonal model is higher than that for the non-seasonal model, indicating that the seasonal model provides a better fit for the data. This is to be expected since modelling seasonal effects explains more of the variability in the data. However, it does not immediately follow from this that the seasonal model will give a more accurate estimate of extremes.

The alternative to modelling the data in discrete seasons or directional sectors is to assume that the distribution parameters vary smoothly with season or direction and to use a Fourier expansion to describe the variation (e.g. Anderson et al., 2001; Jonathan and Ewans, 2007, 2008; Ewans and Jonathan, 2008). Threshold selection is problematic when estimating the variation of parameters using a Fourier expansion. When the data are split into separate seasons or directional sectors the variation of certain statistics with threshold can be examined in order to select an appropriate threshold (see Section 3) for each season or directional sector. As far as the present authors are aware, no objective method has been proposed for selecting a threshold when the parameters are modelled as varying smoothly throughout the year. Anderson et al. (2001) chose arbitrarily to set the threshold as the 90th percentile of the data for each month. Jonathan and Ewans (2008) use a variable threshold, estimated for each day of the year as the 50th or 80th percentile of the nearest 300 storm peaks (in terms of season). However, they note that the effect of the choice of threshold appears more influential than incorporation of seasonally varying extreme value parameters. Moreover, there is no *a priori* reason to assume that the threshold for which the GPD can be considered a reasonable fit will correspond to the same percentile of the data or storm peaks in each season.

Another approach to modelling the seasonality of extreme sea states has been taken by Stefanakos and Athanassoulis (2006). They assume that the time series of  $H_s$  admits the representation  $H_s(t) = \mu(t) + \sigma(t)W(t)$  where  $\mu(t)$  and  $\sigma(t)$  are deterministic time-dependent periodic functions, with a period of 1 year, representing the seasonal mean and standard deviation of the process.  $W(t)$  is a zero-mean stationary stochastic process, referred to as the residual stochastic process. They then show that extremal properties of the process  $H_s(t)$  can be calculated from extremal properties of  $W(t)$ . The model is fitted to the data using the method described by Athanassoulis and Stefanakos (1995). They show that mean, standard deviation and frequency spectrum of the fitted model closely match those of the original data. However, they do not show whether the extremal properties of their fitted model are actually a good match to those of the

original data, so we cannot draw conclusions about the efficacy of seasonal modelling from this method.

Reports of the performance of analyses which model directionality or seasonality have been mixed. Morton et al. (1997), Anderson et al. (2001), and Jonathan and Ewans (2007) all report higher return values from non-stationary models, whereas Stefanakos and Athanassoulis (2006) report lower return values. In these studies the true return values are not known, so it is not possible to determine which models are more accurate, only that the results differ. Jonathan et al. (2008) fit directional and non-directional models to simulated data where the true return values are known. In their study data comes from two separate distributions, representing storms from two directions, but which could equally be interpreted as storms from different seasons. They demonstrate that in the cases they consider the non-directional models under-estimate return values. However, in the present authors' opinion the examples presented in this study are rather artificial since in real situations the distribution of storm peak wave height will vary smoothly with direction, season or other covariate, rather than changing sharply at the boundary of two sectors. Therefore, the examples presented by Jonathan et al. (2008) do not answer the question of whether it is preferable to use non-stationary models in real situations.

To answer this question, in this paper we construct more realistic simulations, where the distribution of storm peak wave height varies smoothly throughout the year. The simulations are based on the seasonal variation observed from long records of data from 14 wave buoys in the North-East Pacific. The paper is organised as follows: In Section 2 the theory required for the use of the POT model is presented. Section 3 describes the data used to construct the case studies. In Section 4 the method used to fit and check the model to the data is described and the results of the analysis are presented. In Section 5 realistic case studies are constructed based on the results established in Section 4. These are used to compare the accuracy of estimates of return values from seasonal and non-seasonal models. Finally, conclusions are presented in Section 6.

## 2. Theory

Let  $X$  denote the peak  $H_s$  in a storm and  $u$  denote a high threshold. Then the generalised Pareto distribution for  $X$ , conditional on exceeding  $u$ , is given by

$$F(x) = \Pr\{X \leq x | X > u\} = \begin{cases} 1 - [1 + \zeta(x-u)/\sigma]_+^{-1/\zeta} & \text{for } \zeta \neq 0 \\ 1 - \exp[-(x-u)/\sigma]_+ & \text{for } \zeta = 0 \end{cases} \quad (1)$$

where  $\sigma > 0$  and  $[\cdot]_+$  denotes the positive part of the term in the square brackets. The parameters  $\sigma$  and  $\zeta$  are called the scale and shape parameters, respectively.

Some authors have proposed estimators for the threshold,  $u$ , at which the GPD can be considered a valid model for the data (e.g. Moharram et al., 1993; Singh and Guo, 1995). However, the method preferred by many practitioners is to estimate the scale and shape parameters for a range of thresholds and chose the threshold by examining the behaviour of certain statistics with the threshold. This approach makes the threshold choice somewhat subjective, but it illustrates whether the GPD can be considered a valid model for the data. The statistics which are plotted against threshold are described below.

If the exceedances of a threshold  $u_0$  follow a generalised Pareto distribution with scale parameter  $\sigma_{u_0}$  and shape parameter  $\zeta$ , it can be shown that (e.g. Coles, 2001) the exceedances of any threshold  $u > u_0$ , follow a generalised Pareto distribution with the

same shape parameter  $\xi$  and scale parameter given by

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0) \tag{2}$$

So the variable

$$\sigma^* = \sigma_u - \xi u, \tag{3}$$

known as the normalised scale parameter, is constant with respect to  $u$ . Therefore, if estimates of  $\xi$  and  $\sigma^*$  are plotted against  $u$  a minimum threshold,  $u_0$ , should be observed above which the estimates are constant. In practice finite sampling effects lead to variability of these parameters with threshold, so it is rare to see a completely straight line. Moreover, as the threshold increases the number of samples will decrease, increasing the variance of the estimates.

We also have that the expected value of the exceedance over a threshold  $u$  is given by (Coles, 2001)

$$E(X - u | X > u) = \frac{\sigma_{u_0} + \xi u}{1 - \xi} \tag{4}$$

So if the GPD is a valid model for exceedances over  $u_0$  then the mean of the exceedances over threshold  $u > u_0$  are a linear function of  $u$  with slope  $\xi / (1 - \xi)$  and intercept  $\sigma_{u_0} / (1 - \xi)$ .

Finally, if the GPD is a valid model for exceedances above a threshold then the estimates of return values should be constant above the threshold. The  $N$  year return value is the value that, on average, is exceeded once every  $N$  years. It is the solution of

$$\frac{1}{Nm} = \Pr\{X > x_N\} = \Pr\{X > u\} \Pr\{X > x_N | X > u\} \tag{5}$$

where  $m$  is the number of observations per year. Substituting (1) into (5) and rearranging gives

$$x_N = \begin{cases} u + \frac{\sigma}{\xi} [(Nm \zeta_u)^\xi - 1] & \text{for } \xi \neq 0 \\ u + \sigma \log(Nm \zeta_u) & \text{for } \xi = 0 \end{cases} \tag{6}$$

where  $\zeta_u = \Pr\{X > u\}$ . For dependent data,  $\zeta_u$  is estimated by  $n_c/n$ , where  $n_c$  is the number of clusters above  $u$ , and  $n$  is the total number of samples.

So if we plot estimates of  $\xi$ ,  $\sigma^*$ ,  $x_N$  and  $E(X - u | X > u)$  against  $u$ , the threshold is selected as the lowest value for which the estimates of  $\xi$ ,  $\sigma^*$  and  $x_N$  approach constant values and  $E(X - u | X > u)$  satisfies (4).

### 3. Data

In order to form realistic case studies to compare seasonal and non-seasonal models we will use data from several buoys on the west coast of the USA, operated by the National Data Buoy Centre (NDBC). The locations of these buoys are shown in Fig. 1 and information about the buoy types, locations and coverage are given in Table 1. The buoys selected for use here are those for which the record length exceeds 20 years. The west coast of the USA is an area with an energetic wave climate that is not affected by tropical cyclones. This second point is important since it simplifies the analysis. Storms generated by tropical cyclones cannot be considered as coming from the same population as those from extratropical cyclones. Tropical cyclones are much smaller than extratropical storms, so a 20- or 30-year record for one location may only contain one or two tropical cyclone events. Moreover, these events are typically much more intense than those from extratropical cyclones. Methods for estimating the distribution of extremes from tropical cyclones are discussed by Hardy et al. (2003) and Jonathan and Ewans (2006).

One-dimensional spectra have been retrieved from the US National Oceanic Data Centre. The data has already had some quality control applied, as detailed in NDBC (1996, 2003).

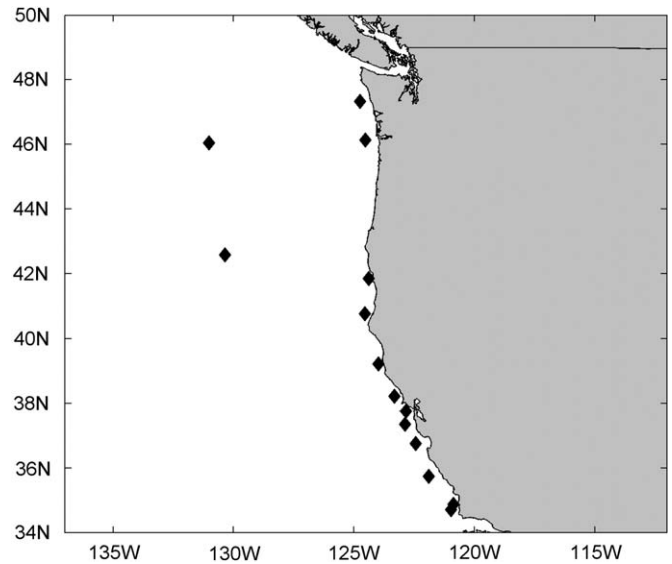


Fig. 1. Locations of the 14 buoys used in this study.

Table 1  
Details of the NDBC buoys used in this study.

Buoy number	Lat. (N)	Lon. (W)	Data period	Water depth (m)	Buoy size (m)
46002	42.60	130.27	1978–2008	3374	6
46005	46.05	131.02	1978–2008	2780	6
46011	34.87	120.86	1980–2008	204	3
46012	37.36	122.88	1980–2008	88	3
46013	38.23	123.32	1981–2008	123	3
46014	39.20	123.97	1981–2008	274	3
46022	40.78	124.54	1982–2008	509	3
46023	34.71	120.97	1982–2008	384	10
46026	37.76	122.83	1982–2008	52	3
46027	41.85	124.38	1983–2008	48	3
46028	35.74	121.89	1983–2008	1112	3
46029	46.14	124.51	1984–2008	128	3
46041	47.35	124.73	1987–2008	132	3
46042	36.75	122.42	1987–2008	2115	3

However, a number of clear outliers remained for some buoy records. These were confirmed by inspection of the individual spectra and removed from the datasets. The data has been interpolated for gaps less than 3 h and then smoothed using a 3 h moving average filter. Smoothing measured data in this way is recommended by Forristall et al. (1996). They note that if there are several samples of  $H_s$  at times near the peak of a storm, then as a result of sampling variability the maximum of these samples is very likely to be larger than the expected value of  $H_s$  at the peak of the storm. Averaging the data over longer periods reduces sampling variability and therefore minimises this effect. Three hours is recommended as a compromise between statistical stability and adequate sampling of changes in sea state.

The data have been declustered to pick out the maximum value of  $H_s$  in each storm. The notion of a storm has a direct physical interpretation: a peak in the time history of  $H_s$  at a particular location is associated with winds generated by a weather system. In practice winds may vary as the weather system evolves, causing several peaks over a number of days. If synoptic pressure charts were available it may be relatively easy to see that several consecutive peaks in a record were the result of a single storm, but this may not be so clear from the time history of  $H_s$  alone. From a physical point of view, a pressure system which causes a large peak in the  $H_s$  record will generally be large

in extent and persist for a number of days. It is therefore unlikely that peaks close to each other come from separate storms. Moreover, if two separate storms occur in close proximity then it may not be realistic to assume that their characteristics are independent. Tawn (1988) uses a minimum separation of

30 h between storms peaks in the southern North Sea, whereas Anderson et al. (2001) use an interval of 24 h for a location in the northern North Sea. The choice is somewhat subjective, but a separation of 36 h appears appropriate for the data used here.

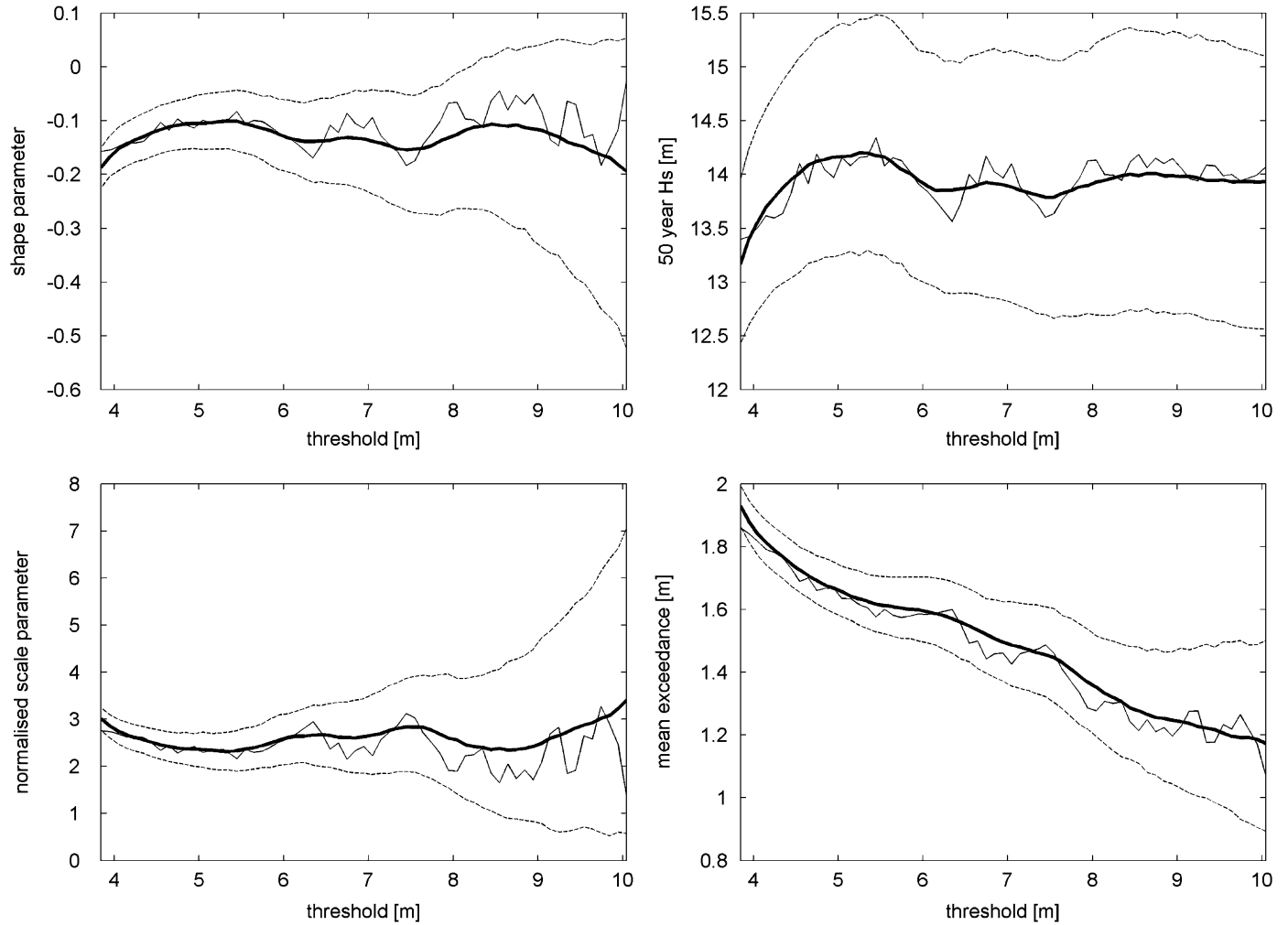


Fig. 2. Threshold plots for data from buoy 46002. Thin solid line: estimates without resampling. Bold line: mean of modified bootstrap estimates. Dotted line: 95% confidence bounds from modified bootstrap trials.

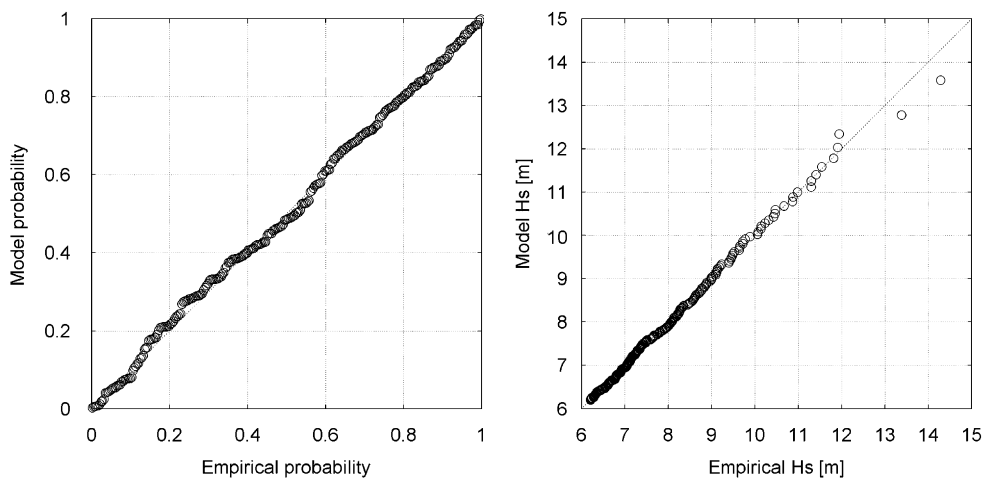


Fig. 3. Diagnostic plots for model fitted using LM estimators with a threshold of 6.2 m. Left: Probability plot; right: quantile plot.

#### 4. Analysis of seasonal distributions of extreme $H_s$

There have been many methods proposed to estimate the parameters of GPD. A review of many of these methods has been presented by de Zea Bermudez and Kotz (2009). Some of the commonly used estimators for the GPD such as maximum likelihood (ML) (Grimshaw, 1993; Chaouche and Bacro, 2006) and probability weighted moments (PWM) (Hosking and Wallis, 1987) have drawbacks. The ML estimator is difficult to compute and for small sample sizes may not exist at all. It is asymptotically efficient, but Hosking and Wallis (1987) showed that the ML estimates do not display this asymptotic property for sample sizes up to 500 and that PWM estimates have lower bias and variance for smaller sample sizes. In this study the likelihood-moment estimator (Zhang, 2007), using the hybrid probability weighted moment estimate (Dupuis and Tsao, 1998) as a first guess, has been used to estimate the parameters of the GPD from the buoy data. Mackay et al. (2009) have shown that this method (henceforth referred to as simply the LM estimator) outperforms other methods commonly used to estimate the parameters of the GPD. The LM estimators always exist and are easy to compute and are much less sensitive to threshold choice. The LM estimator also performs as well or better than both the ML and PWM estimates in terms of bias and variance.

To aid the identification of a valid threshold, Naess and Clausen (2001) have suggested smoothing the threshold plots using a

moving average filter. This method does produce smoother plots, but does not provide any further information about the data. Moreover, estimates for higher thresholds will have greater uncertainty due to fewer data, and using a moving average does not take this into account. We use a modified bootstrap method to aid threshold selection which takes into account the uncertainty inherent in the data. An introduction to the bootstrap method is given by Efron and Tibshirani (1993). When using a bootstrap technique it is assumed that the data represent a realisation of the underlying population. If it is further assumed that the data are independent and identically distributed then it follows that a random sample of the data, drawn with replacement, is an equally likely realisation of the same process. By resampling the data a large number of times we can estimate the effect of sampling variability on estimates of  $\xi$ ,  $\sigma^*$ , return values and the mean exceedance. This can be used to add confidence bounds to the plots. The bootstrap is used by numerous authors in this context (e.g. Elsinghorst et al., 1998; Naess and Clausen, 2001; Jonathan and Ewans, 2007).

Resampling in this way does not smooth the threshold plots, since the mean bootstrap estimate will tend to the original non-bootstrapped estimate. Incorporating information about the uncertainty of each estimate of storm peak  $H_s$  can smooth the plots to a certain extent. The distribution of the estimate of maximum  $H_s$  in a storm has been discussed in detail by Forristall et al. (1996). From their results we estimate that the coefficient of

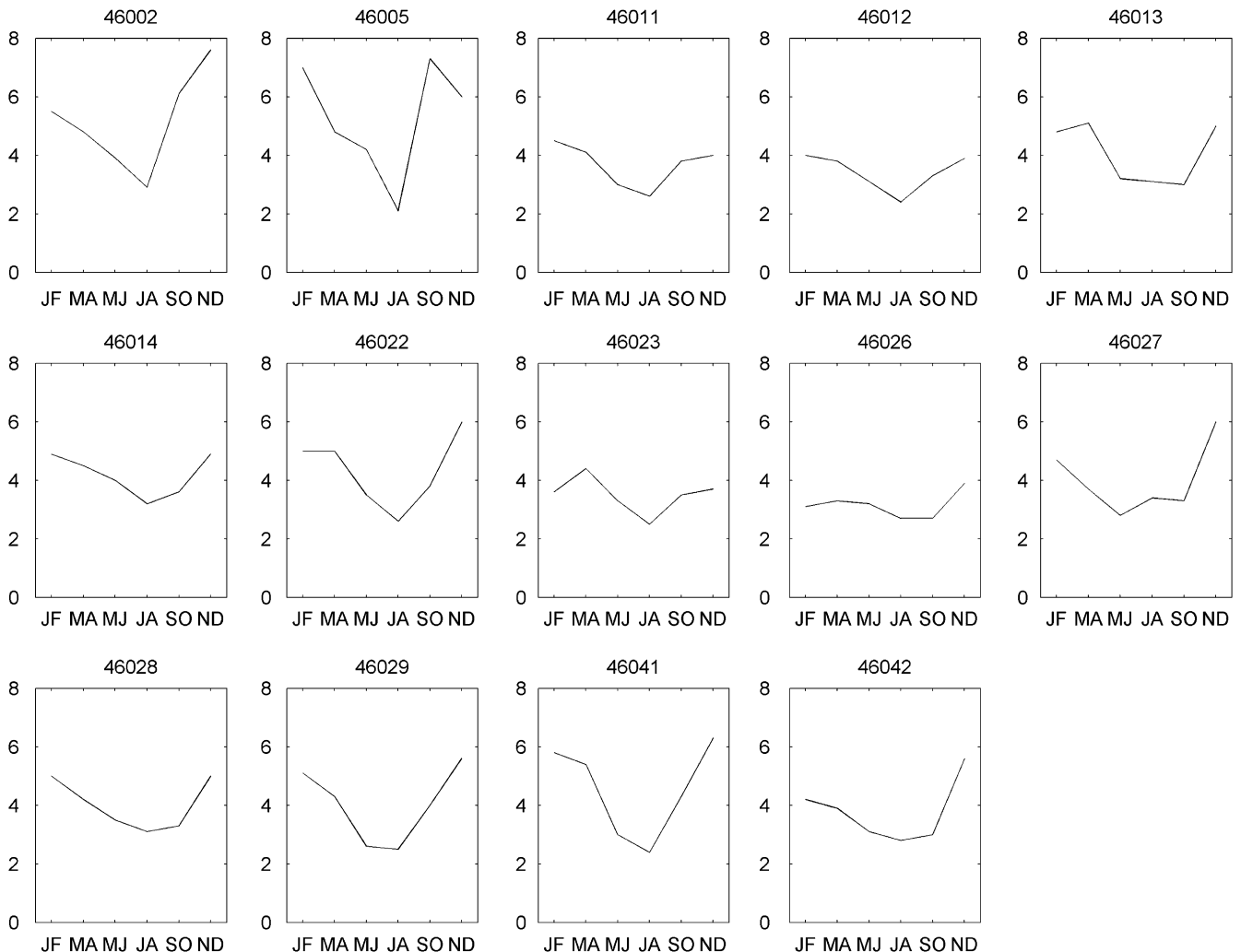


Fig. 4. Estimated thresholds (m) for the GPD.

variation in the peak  $H_s$  in our dataset is around 3%. This information is incorporated into the bootstrap in the following way. For each bootstrap trial we draw a random sample  $X_{(j,1)}, \dots, X_{(j,n)}$  from the original data  $X_1, \dots, X_n$  and generate  $n$  standard normal variables  $e_1, \dots, e_n$ . The bootstrap sample is defined to be  $X_{(j,1)}(1+0.03e_1), \dots, X_{(j,n)}(1+0.03e_n)$ . This is a natural extension of the bootstrap method, since samples generated in this way are an equally likely realisation of the underlying process. The fact that the data are not identically distributed can be dealt with by dividing the data into seasons, and resampling each season separately. The use of seasons of various lengths has been tested. It was found that 2 months is reasonable compromise between sample size and the distribution being reasonably stationary within each season.

The use of this modified bootstrap technique is illustrated in Fig. 2, using data from buoy 46002. The mean bootstrap estimate is considerably smoother than the original non-bootstrapped estimates. Bootstrapping cannot rectify any inherent bias in the empirical distribution (i.e. differences between the shape of the sample distribution and the underlying population distribution), but uncertainty can be gauged through the confidence bounds. From these plots the threshold has been selected as 6.2 m. After this point the mean exceedance becomes approximately linear, with slope  $-0.11$  and intercept 2.2. From Eq. (4) this is equivalent to  $\hat{\xi} = -0.12$  and  $\hat{\sigma}_{u_0} = 2.5$ . The estimates of  $\xi$ , and  $\sigma^*$  display stability and from around 6.2 m until around 7.5 m, with values consistent with those from the mean excess plot. The variability

displayed after this point is likely to be due to decreasing sample size. Despite the variability in the estimates of  $\xi$  and  $\sigma^*$  above 7.5 m the estimate of the 50-year return value of  $H_s$  is reasonably constant from 6.2 m onwards.

The fit of the model for a threshold of 6.2 m is checked using probability and quantile plots, shown in Fig. 3. The quantile plot shows excellent agreement for all but the highest 3 points. The probability plot shows good agreement for the higher probability points, with some small deviations below a probability of 0.6. Overall it appears that the fit is good and that the GPD can be considered as a valid model for the data.

This method has been used to select thresholds for fitting the GPD to data from the 14 NDBC buoys listed in Table 1, both for the data considered as a whole and in separate seasons. Again, a season length of 2 months has been selected as a reasonable compromise between sample size and stationarity. The GPD was found to provide a good fit for both annual and seasonal data in most cases, although in some cases it was more difficult to identify a valid threshold for the seasonal data due to the smaller sample sizes. The estimated parameters of the seasonal distributions are shown in Figs. 4–7.

Fig. 4 shows the thresholds estimated for each season and each buoy. The smallest range of thresholds was for buoy 46026 with a minimum of 2.7 m and maximum of 3.9 m. The largest range was for buoy 46005 with a minimum of 2.1 m and maximum of 7.0 m. As would be expected, the estimated threshold varies with season. Estimated scale parameters are shown in Fig. 5. These

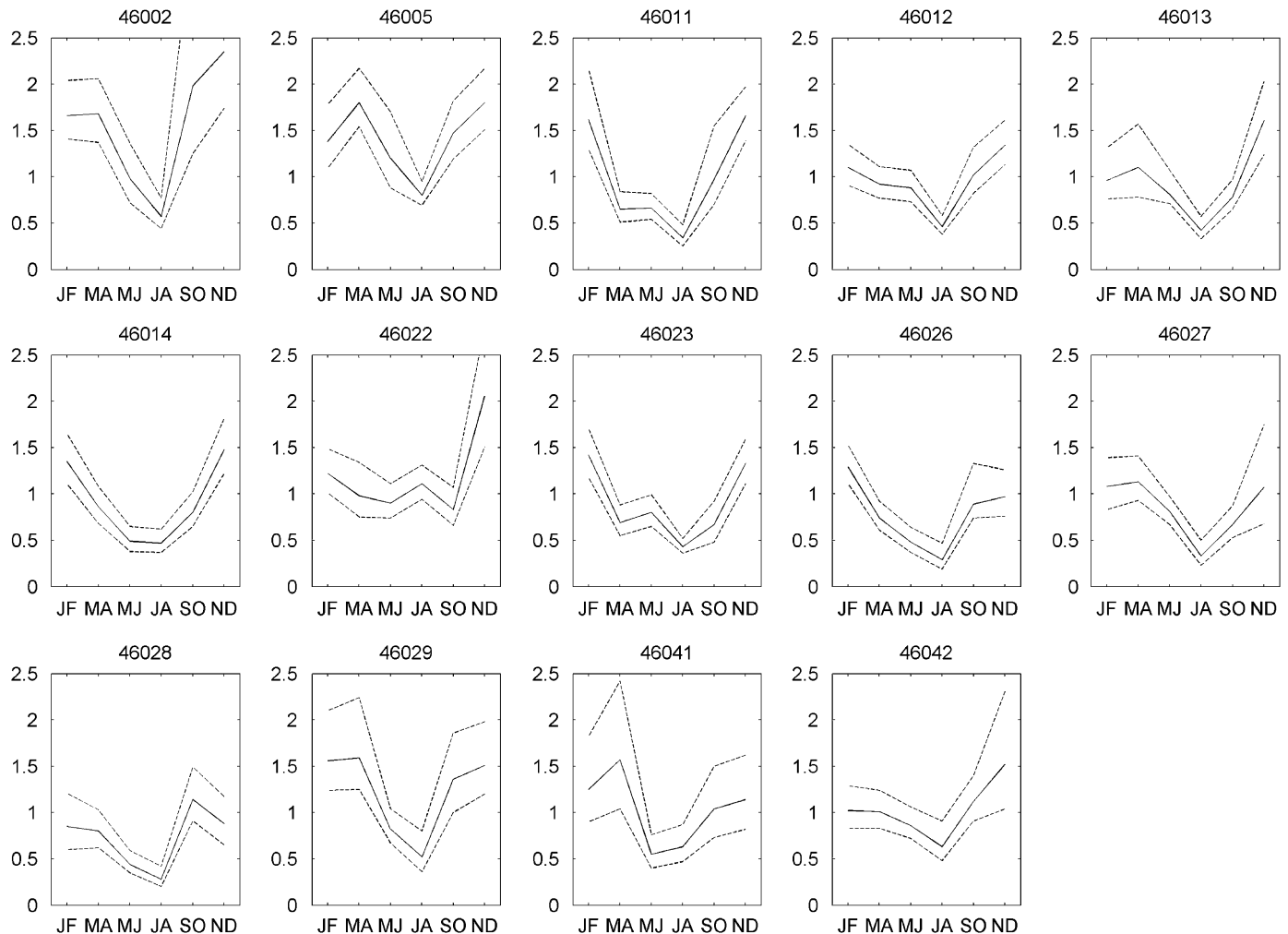


Fig. 5. Estimated scale parameters with 95% confidence bounds (dashed lines).

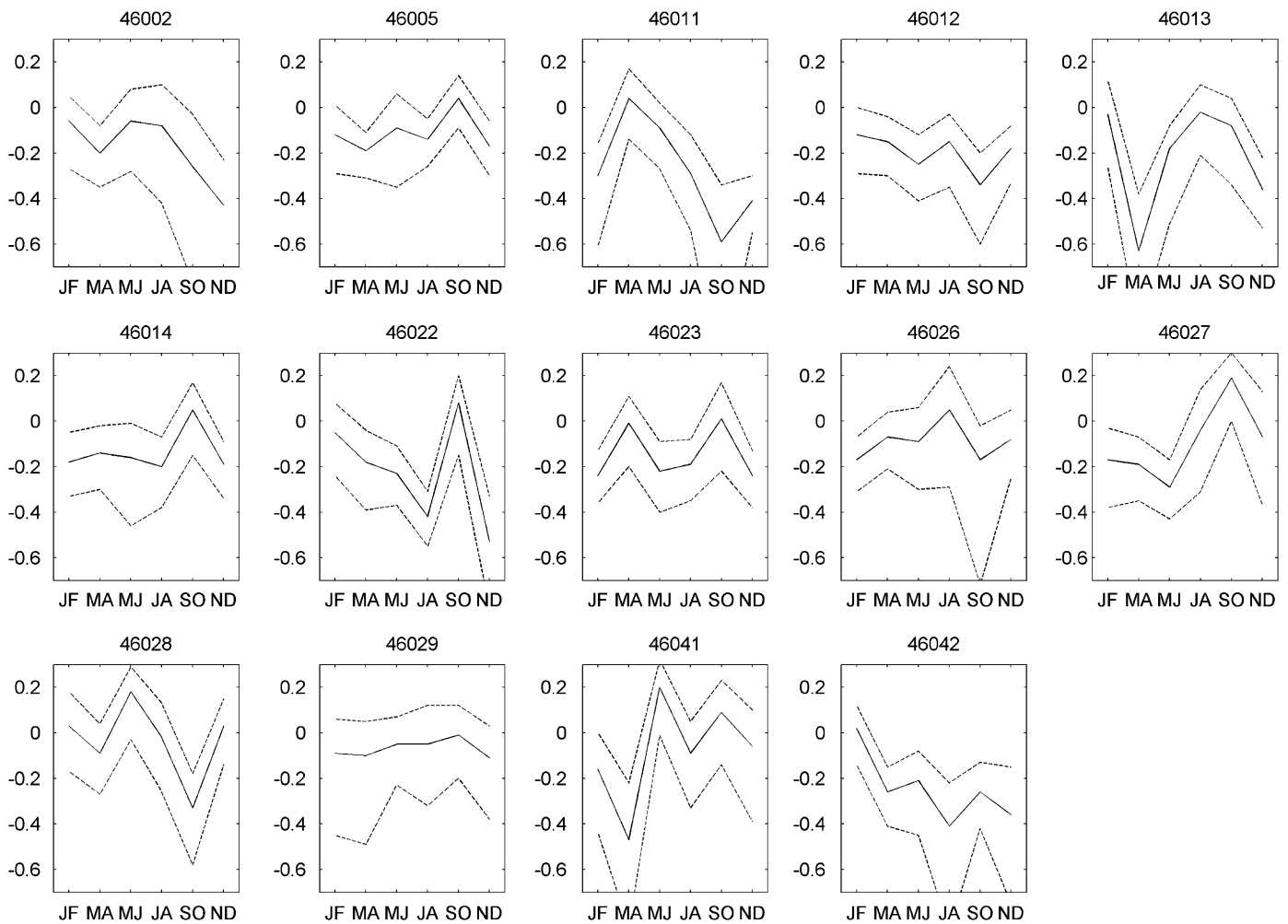


Fig. 6. Estimated shape parameters 95% confidence bounds (dashed lines).

vary roughly in proportion to the threshold with a mean value of  $\hat{\sigma}/u = 0.26$ . There does not appear to be any systematic variation of the shape parameter throughout the year (Fig. 6), the mean and standard deviation of the estimates over all buoys remaining roughly equal in each season. It is possible that the shape parameter may be slightly higher in April–May and September–October, but this may be a result of the rapidly changing distribution causing the compound distribution to appear heavier-tailed. The mean value of the shape parameter over all seasons and all buoys is  $-0.15$ . Fig. 7 shows the mean number of storms exceeding the threshold per month. The values range between 0.5 and 4 depending on the buoy, season and threshold. Again, there was no clear pattern across all the buoys, but for some buoys there are fewer storms in summer than winter. The mean number of storms per month over all seasons and buoys was 2.0.

The 50- and 100-year return values from the seasonal and non-seasonal analyses are shown in Table 2. Annual return values have been calculated from the seasonal parameters using the method described in the Appendix A. The return values from the seasonal analysis are generally close to those calculated from the non-seasonal analysis, but the confidence bounds for the return values from the seasonal analysis were wider in all cases.

For 10 out of the 14 datasets the return values from the seasonal analysis are higher. For buoy 46005 the 50- and 100-year return values from the seasonal analysis are 1.0 and 1.5 m higher than those from the non-seasonal analysis. It appears that this is

due to the fitted distribution in September–October. Fig. 8 shows the distribution of storm peaks throughout the year. It is clear that the distribution cannot be considered stationary through September–October. The result is that the distribution appears heavier tailed, leading to a positive estimate of the shape parameter.

Fig. 9 shows threshold plots for September–October. There is no clear choice of threshold from these plots. Initially a threshold of 4.5 m was chosen, since the scale and shape parameters appear constant after this point, until about 6 m. Also the mean exceedance plot changes at 4.5 m and could be considered linear after this point if the confidence bounds are taken into account. The estimated 50-year return value only shows stability after 7.3 m. However, above 6 m the estimated shape and normalised scale parameters are not stable. From quantile plots (not shown), it is evident that the fit at higher quantiles is slightly better using the threshold of 7.3 m, but at the expense of having far fewer points (18 compared to 117). Using the higher threshold, the annual return values from the seasonal analysis become  $H_{s,50} = 13.0$  m and  $H_{s,100} = 13.4$  m, less than 0.2 m different from those from the non-seasonal analysis. The same problem occurred for buoy 46027 in September–October due to the rapidly changing distribution during this period. A solution to this problem would be to use shorter seasons, or to visually estimate seasonal boundaries. However, using shorter seasons would result in few data and therefore greater uncertainty, making a valid threshold more difficult to identify.

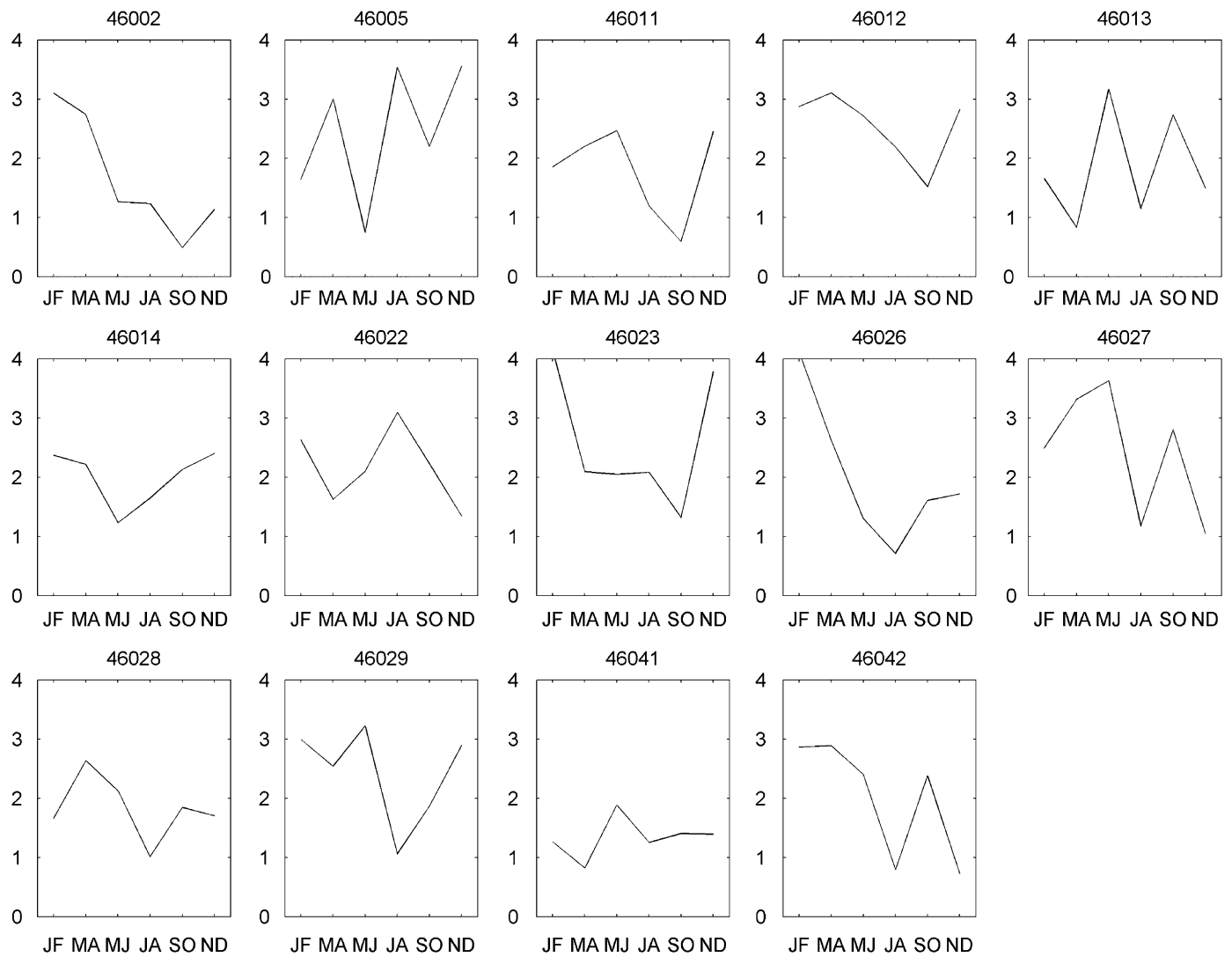


Fig. 7. Mean number of threshold exceedances per month.

Table 2

Maximum observation, 50- and 100-year return values from the seasonal and non-seasonal analyses. 95% confidence bounds shown in brackets.

Buoy	Maximum observation (m)	$H_{s,50}$ (m)		$H_{s,100}$ (m)	
		Non-seasonal	Seasonal	Non-seasonal	Seasonal
46002	14.3	13.8 (12.5, 15.2)	13.5 (12.2, 15.6)	14.3 (12.7, 16.0)	14.3 (12.4, 17.1)
46005	12.5	12.8 (12.1, 13.6)	13.8 (12.8, 15.2)	13.2 (12.4, 14.0)	14.7 (13.4, 16.9)
46011	9.1	8.8 (7.9, 9.5)	8.9 (7.9, 9.7)	9.1 (8.0, 10.0)	9.2 (8.1, 10.4)
46012	8.7	8.9 (8.2, 9.6)	8.9 (8.2, 9.8)	9.2 (8.3, 10.1)	9.2 (8.4, 10.4)
46013	9.6	9.4 (8.5, 10.2)	9.4 (8.6, 10.7)	9.7 (8.7, 10.8)	10.0 (8.7, 11.7)
46014	9.8	10.2 (9.4, 11.0)	10.2 (9.6, 11.1)	10.6 (9.7, 11.5)	10.6 (9.8, 11.9)
46022	11.5	11.2 (10.0, 12.3)	11.3 (9.8, 13.2)	11.6 (10.2, 13.0)	12.0 (10.2, 14.6)
46023	8.0	8.1 (7.7, 8.5)	8.5 (8.0, 9.3)	8.2 (7.7, 8.7)	8.7 (8.2, 10.0)
46026	8.0	8.1 (7.4, 8.7)	8.3 (7.7, 9.1)	8.3 (7.5, 9.1)	8.7 (7.9, 9.7)
46027	9.6	10.3 (9.4, 11.3)	11.0 (9.8, 12.6)	10.7 (9.7, 11.9)	11.9 (10.2, 15.0)
46028	9.0	9.8 (8.9, 10.7)	10.6 (9.5, 11.6)	10.2 (9.2, 11.4)	11.4 (10.1, 12.8)
46029	12.8	12.5 (10.9, 14.0)	12.8 (11.4, 15.3)	13.0 (11.1, 14.8)	13.4 (11.9, 16.8)
46041	11.4	11.2 (9.8, 12.5)	11.9 (10.5, 13.4)	11.6 (10.0, 13.3)	12.8 (11.2, 15.1)
46042	9.4	10.6 (9.4, 11.9)	10.2 (9.0, 11.7)	11.3 (9.8, 12.9)	11.0 (9.2, 12.9)

This example illustrates some of the difficulties in implementing the seasonal model. The difficulty in identifying a valid threshold in one season, seriously affects the results of the analysis. In fact, it is not clear if any threshold can be considered valid for this season. However, in some cases there were similar

difficulties in identifying a valid threshold for the non-seasonal analysis.

For buoys 46028 and 46041 there are fewer data in the winter months than in summer. Fig. 10 shows the amount of data in each season, relative to the season with the most data. The missing

data in the winter months is likely to have introduced a small negative bias to the non-seasonal analyses. The effect of missing data on estimated return values is tested analytically in the next section and shown to have only a small influence.

5. Simulations

In this section the accuracy of discrete seasonal models is compared to that of non-seasonal models using realistic case studies, based on the analysis of the NDBC data presented in the last section. In all four cases the parameters of the distribution are

assumed to vary smoothly throughout the year. In the first two case studies the shape parameter remains constant throughout the year and the scale parameter varies in proportion to threshold. In the third and fourth case studies the shape parameter varies throughout the year to make the distribution more stretched in spring and autumn, as was noticed for some of the buoy records. In the first and third case studies the threshold varies between 2 and 4 m, and in the second and fourth it varies between 2 and 6 m. For simplicity, a year is defined to be 360 days and a month to be 30 days. The times between threshold exceedances are modelled as exponentially distributed with a mean of 15 days, so that the number of storms per month is Poisson distributed with a mean of two storms per month, consistent with the buoy data.

The parameters of the distributions for the four case studies are given by ( $d$  denotes the day of the year)

Case 1

$$u = 3 + \cos(2\pi d/360)\sigma = 0.75 + 0.25 \cos(2\pi d/360)\xi = -0.15$$

Case 2

$$u = 4 + 2 \cos(2\pi d/360)\sigma = 1 + 0.5 \cos(2\pi d/360)\xi = -0.15$$

Case 3

$$\begin{aligned} u &= 3 + \cos(2\pi d/360) \\ \sigma &= 1 + 0.5 \cos(2\pi d/360) - 0.25 \cos(2\pi d/180) \\ \xi &= -0.15 - 0.1 \cos(2\pi d/360) \end{aligned}$$

Case 4

$$\begin{aligned} u &= 4 + 2 \cos(2\pi d/360) \\ \sigma &= 1.5 + 0.75 \cos(2\pi d/360) - 0.3 \cos(2\pi d/180) \\ \xi &= -0.15 - 0.1 \cos(2\pi d/360) \end{aligned}$$

Fig. 11 shows 50-year realisations of the four case studies, together with the theoretical 0.99 and 0.999 quantiles for each day. The solid line indicates the upper limit of the distribution. Return values for the four case studies are shown in Fig. 12. The return values have been calculated using Eq. (A.7) with season lengths of 1 day.

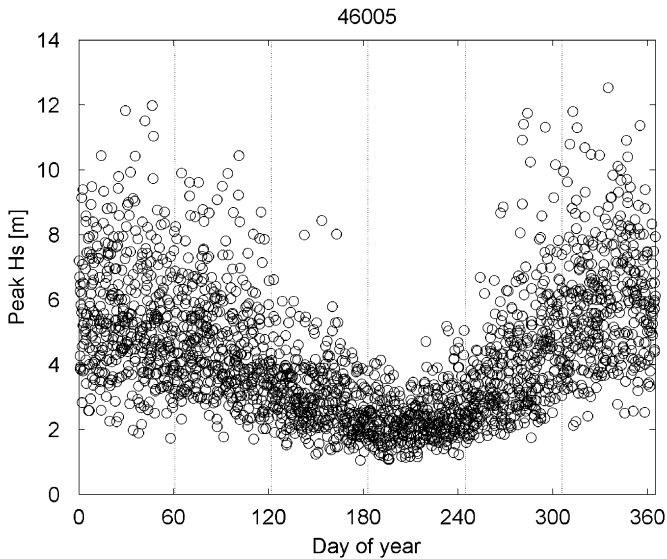


Fig. 8. Storm peak  $H_s$  against day of year for buoy 46005. Dotted lines indicate seasonal boundaries.

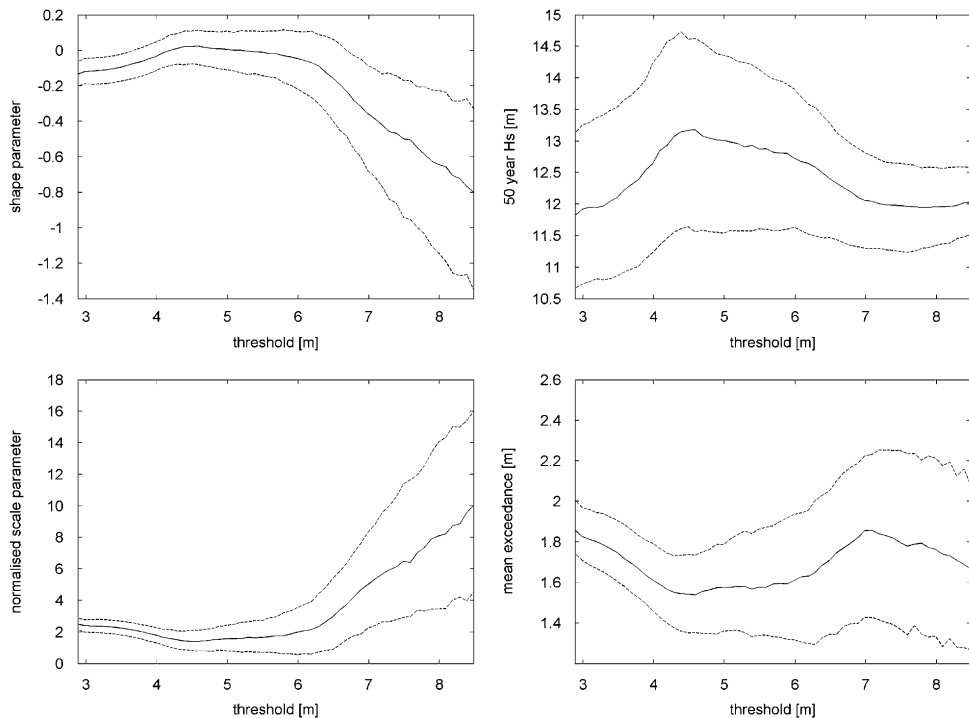


Fig. 9. Threshold plots for data from buoy 46005, September–October. Solid lines: mean bootstrap value; dotted lines: 95% confidence bounds. Dashed line on upper right plot is the maximum observed  $H_s$  between September–October.

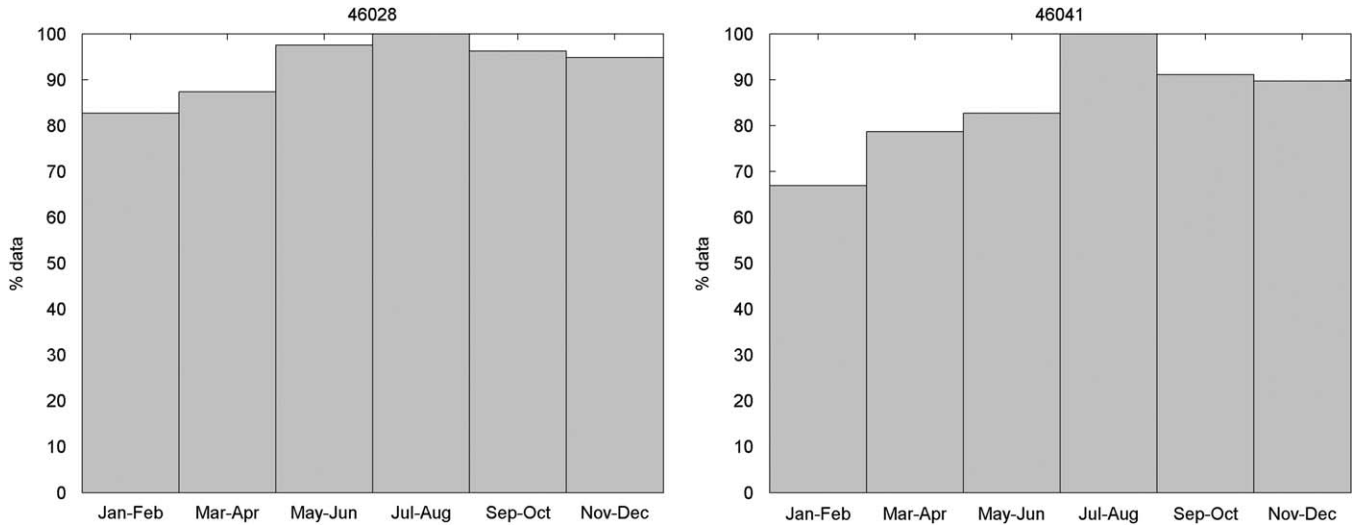


Fig. 10. Percentage of data in each season relative to the season with the most data, for buoys 46028 (left) and 46041 (right).

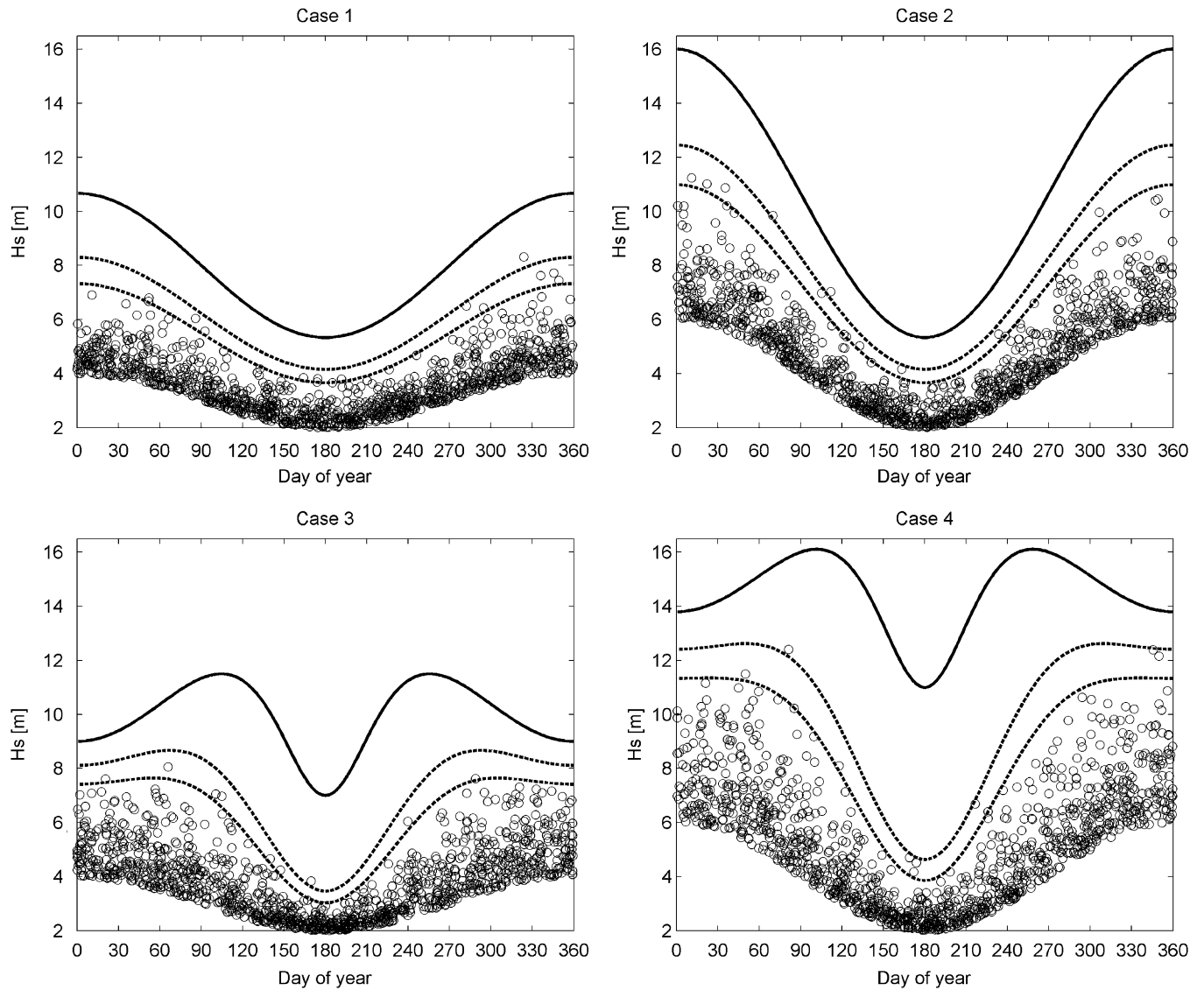


Fig. 11. Distribution of storm peaks in the four case studies. Circles: storm peaks from 50-year realisations. Dashed lines: theoretical 0.99 and 0.999 quantiles for each day. Solid line: upper limit of distribution.

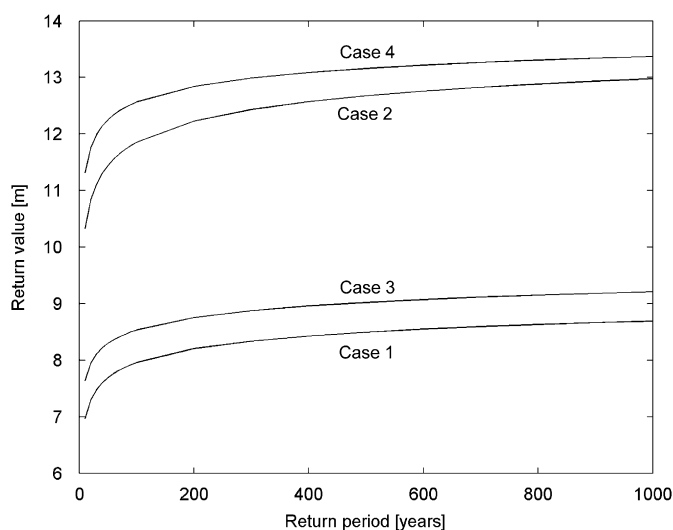


Fig. 12. Return values for the four case studies.

For each case study 1000 simulations have been run, of lengths 10, 20 and 50 years. For each simulation models are fitted using both seasonal and non-seasonal analyses. In the case of the non-seasonal analysis the threshold is chosen as the maximum of  $u$  throughout the year, i.e. 4 m in cases 1 and 3, and 6 m in cases 2 and 4. The seasonal analyses are conducted using seasons of length 1, 2 and 3 months. The midpoint of the first season is defined to be day 0. For each season the threshold is chosen as the maximum of  $u$  in that season.

The results of the simulations are presented in Figs. 13–16, which show the bias and RMS errors of the estimated return values from the non-seasonal and seasonal analyses. The results can be summarised as follows:

- For all the case studies the non-seasonal analysis performs at least as well as the seasonal analyses and in the majority of cases performs significantly better, with lower bias and RMS error.
- Of the seasonal analyses, the 3-month season models has the best performance, followed by the 2-month season models, with the 1-month season models consistently having the highest bias and RMS error.
- For the non-seasonal analyses the bias in estimate of the return value remains roughly constant with the return period, whereas for the seasonal analyses the bias increases with the return period.
- The non-seasonal model has a negative bias in some cases, most notably for case 3, but this is small in comparison to the RMS error.
- The differences in the RMS errors between the seasonal and non-seasonal analyses are small for return periods less than about 50 years.

The finding that the non-seasonal model performs better is contrary to what has been suggested by previous studies on seasonal and other covariate effects. In the examples presented by Jonathan et al. (2008) where storms come from two distinct populations, the compound distribution was not well modelled by the GPD. In the case studies constructed here, where the distribution parameters vary smoothly with season, the compound annual distribution is well modelled by the GPD, even at high return values. These examples are a better representation of the change in the distribution throughout the year, and could equally be interpreted as a change in distribution with direction.

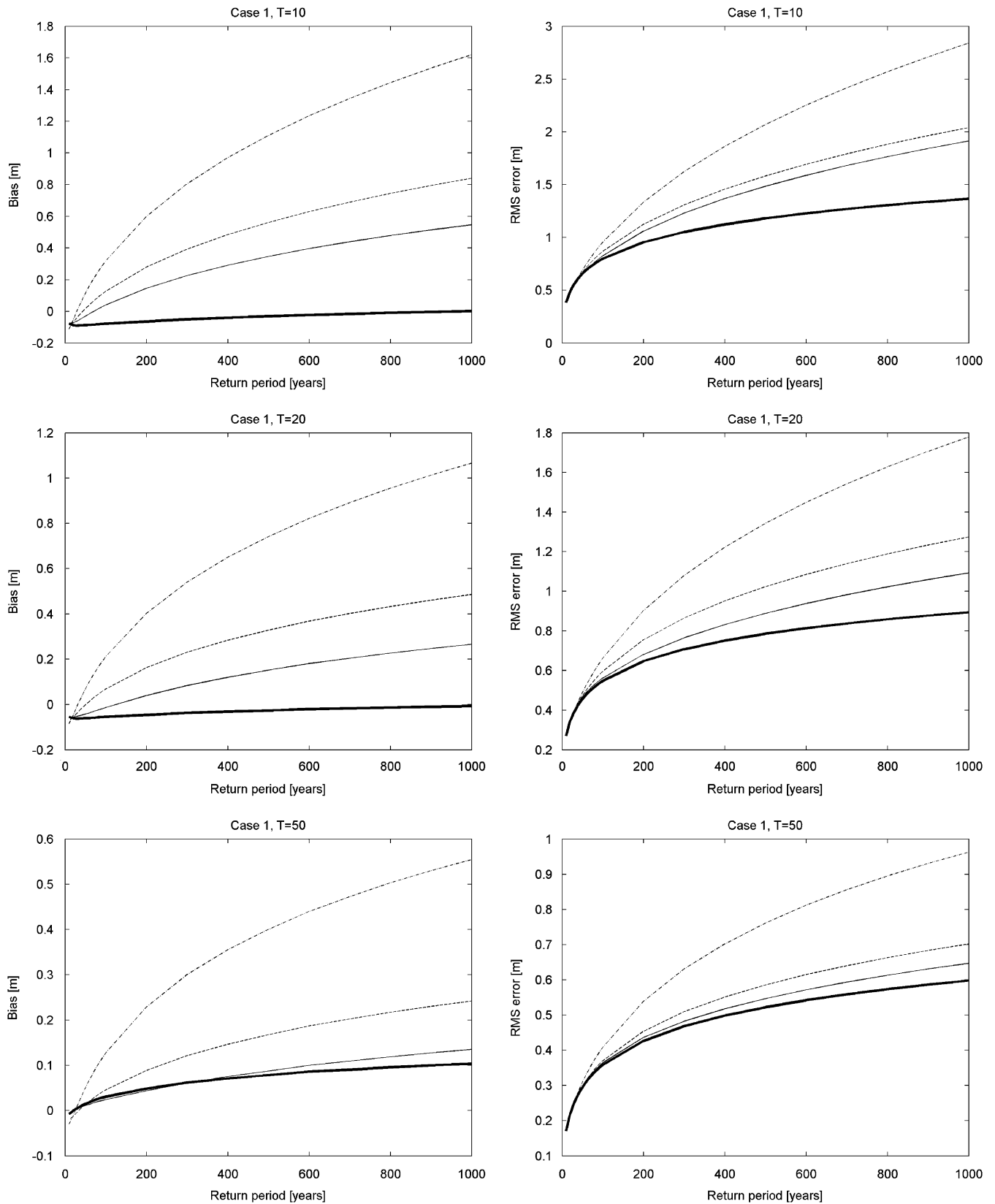
Despite using more data points in total than the non-seasonal models, the seasonal models had a higher RMS error at high return periods. This is caused by the partitioning of the data into smaller sets. For each season, the smaller sample size results in greater uncertainty in the estimates of model parameters. The seasonal models with longer seasons have larger sample sizes and therefore lower RMS errors. It is interesting to note that for cases 3 and 4, where both the scale and shape parameter vary through the year, that the non-seasonal analysis is still able to accurately predict extremes and outperforms the seasonal analyses.

It is perhaps easiest to understand the reason for the positive bias of the seasonal models in terms of maxima. Let  $M$  denote the annual maximum and  $M_i$  denote the maximum in season  $i$ . Denote the distribution function of the annual maximum  $F(x) = \Pr\{M \leq x\}$ , the distribution functions of the seasonal maxima as  $F_i(x) = \Pr\{M_i \leq x\}$  and estimates of these are denoted with a  $\hat{\cdot}$  as usual. Then, assuming the maxima in each season are independent,  $F(x) = \prod F_i(x)$ . Consider a simple example of a two season year, where for some high value  $x$ ,  $F_1(x) = F_2(x) = 0.99$  and  $F(x) = 0.99^2 = 0.98$ . Suppose when the seasonal distributions are estimated from the data that  $\hat{F}_1(x)$  is too low,  $\hat{F}_1(x) = 0.95$  say, but  $\hat{F}_2(x)$  is accurate or too high,  $\hat{F}_2(x) \geq 0.99$ . Then  $\hat{F}(x) \leq 0.95$ . So the effect of an overestimate of the probability of exceeding a high value in one season is not compensated for by underestimates of the probability of exceeding that high value in the other season. Moreover, the sensitivity of the annual distribution to estimates in individual seasons will increase at higher quantiles, as  $F_i(x) \rightarrow 1$ . Therefore the more seasons that the year is divided into, the greater the chance that there will be an overestimate in one season. Since this overestimate is not compensated for by the estimates in other seasons, the annual distribution becomes biased high at high quantiles. An example of this phenomenon was presented in the previous section, where the return values for buoy 46005 were strongly influenced by the estimated distribution in September–October.

There is a small negative bias for the non-seasonal model in the 10-year simulations. This may be a result of non-proportional sampling from each season, as described by Carter and Challenor (1981) (see Section 1). However, the seasonal models generally had higher biases and RMS errors than the non-seasonal models, so their estimates are not considered more accurate. In the simulations the times between threshold exceedances are modelled as exponential with a mean of 15 days. If we denote the length of the simulation in years as  $T$  and the number of seasons that the year is divided into as  $NS$ , then the number of samples in from each season is Poisson distributed with mean  $(T \times 360)/(NS \times 15)$ . For a Poisson distribution the variance is equal to the mean, so as the mean increases, the standard deviation as a fraction of the mean decreases. Therefore, as the length of simulation increases the sampling becomes approximately proportional.

A simulation has also been run to examine the effect of missing data. Case study 2 was chosen since the amplitude of the seasonal cycle is largest in this example. 30% of the data was removed from the period between day 315 and day 45, which corresponds to season 1 in the 3-month season models. The bias and RMS errors for both the seasonal and non-seasonal models are shown in Fig. 17. The non-seasonal estimates are biased low, as would be expected when data is missing from the season with the highest values. The seasonal model shows less bias at low return periods but still has a higher bias and RMS error than the non-seasonal model at high return periods.

Returning to the analysis of the buoy data in the previous section, in light of the simulation studies it is clear that the non-seasonal estimates should be considered as more accurate than



**Fig. 13.** Bias and RMS error in return values from the seasonal and non-seasonal models for case study 1 and simulation lengths of 10, 20 and 50 years. Bold line: Non-seasonal model. Thin line 3-month seasons. Dashed line: 2-month seasons. Dash-dot lines: 1-month seasons.

the seasonal estimates. For 12 of the 14 records, the non-seasonal estimates were less than or equal to the seasonal estimates and in the 2 records where the non-seasonal estimates were higher the

difference was less than 0.4 m, agreeing with the results of the simulation study. In all cases the estimated confidence bounds for the non-seasonal estimates were narrower than those for the

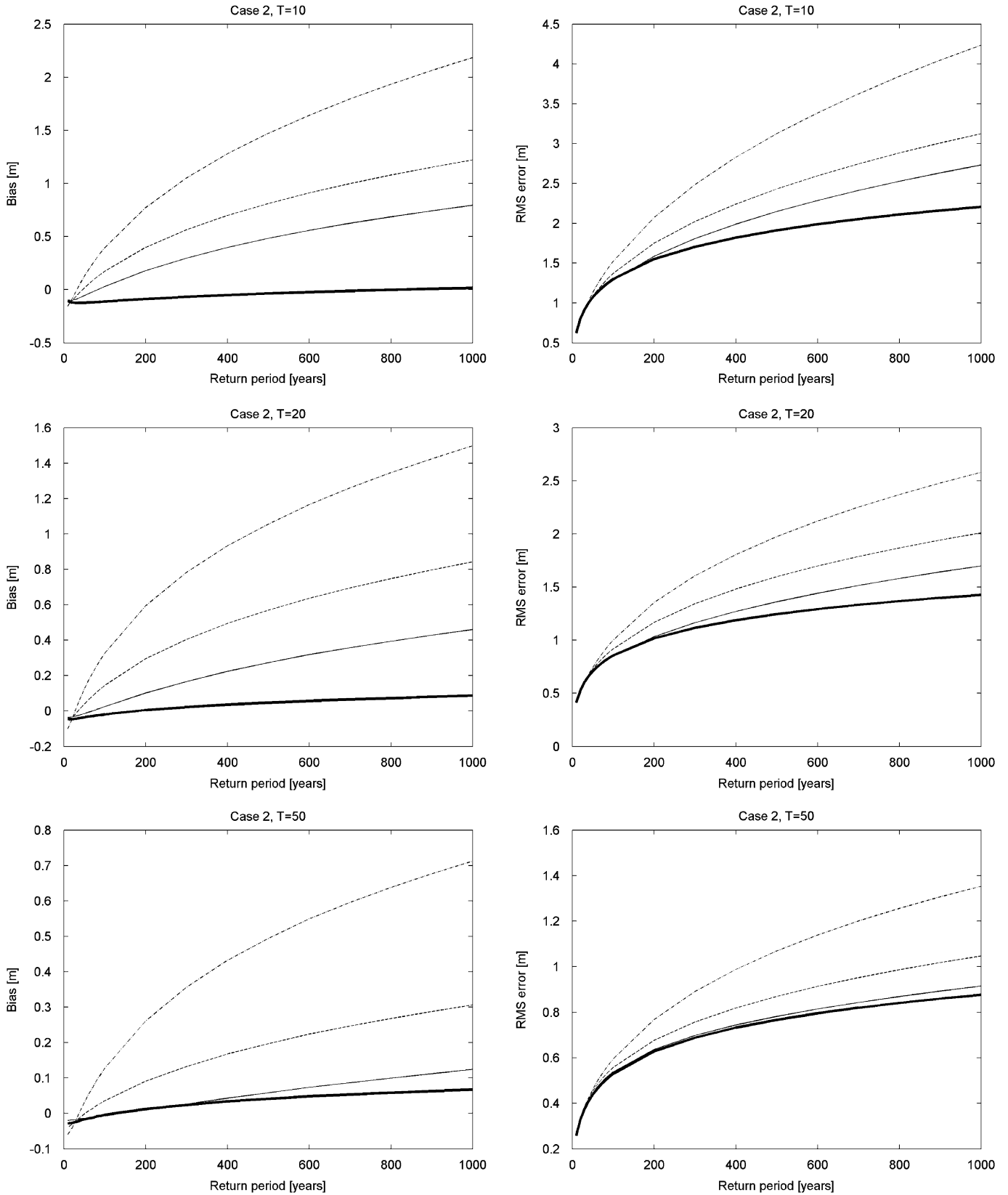


Fig. 14. As previous figure, but for case 2.

seasonal estimates, again agreeing with the results of the simulation study. For buoys 46028 and 46041 where there was 20–30% missing data in the winter, the return values from the seasonal model may be more accurate.

**6. Conclusions**

The use of discrete seasonal models for the estimation of return values of  $H_s$  has been examined using realistic case studies, based on

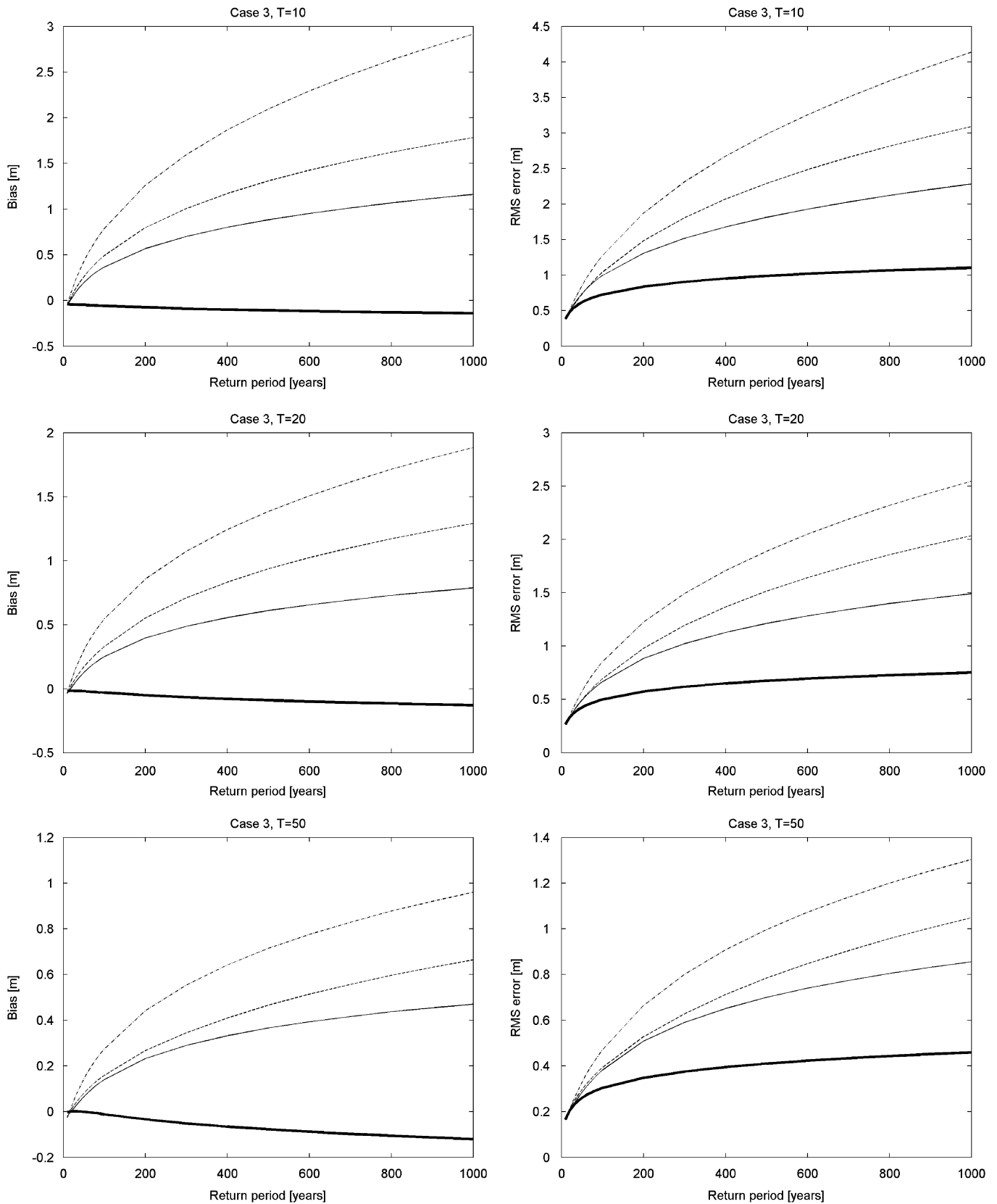


Fig. 15. As previous figure, but for case 3.

examination of buoy data. The results are applicable to discrete directional models as well. Contrary to what would be expected from most literature on the subject, the non-seasonal models performed

better than the seasonal models unless there was a significant amount of missing data. In the case studies considered here, under the assumption that the extremes values follow a GPD with

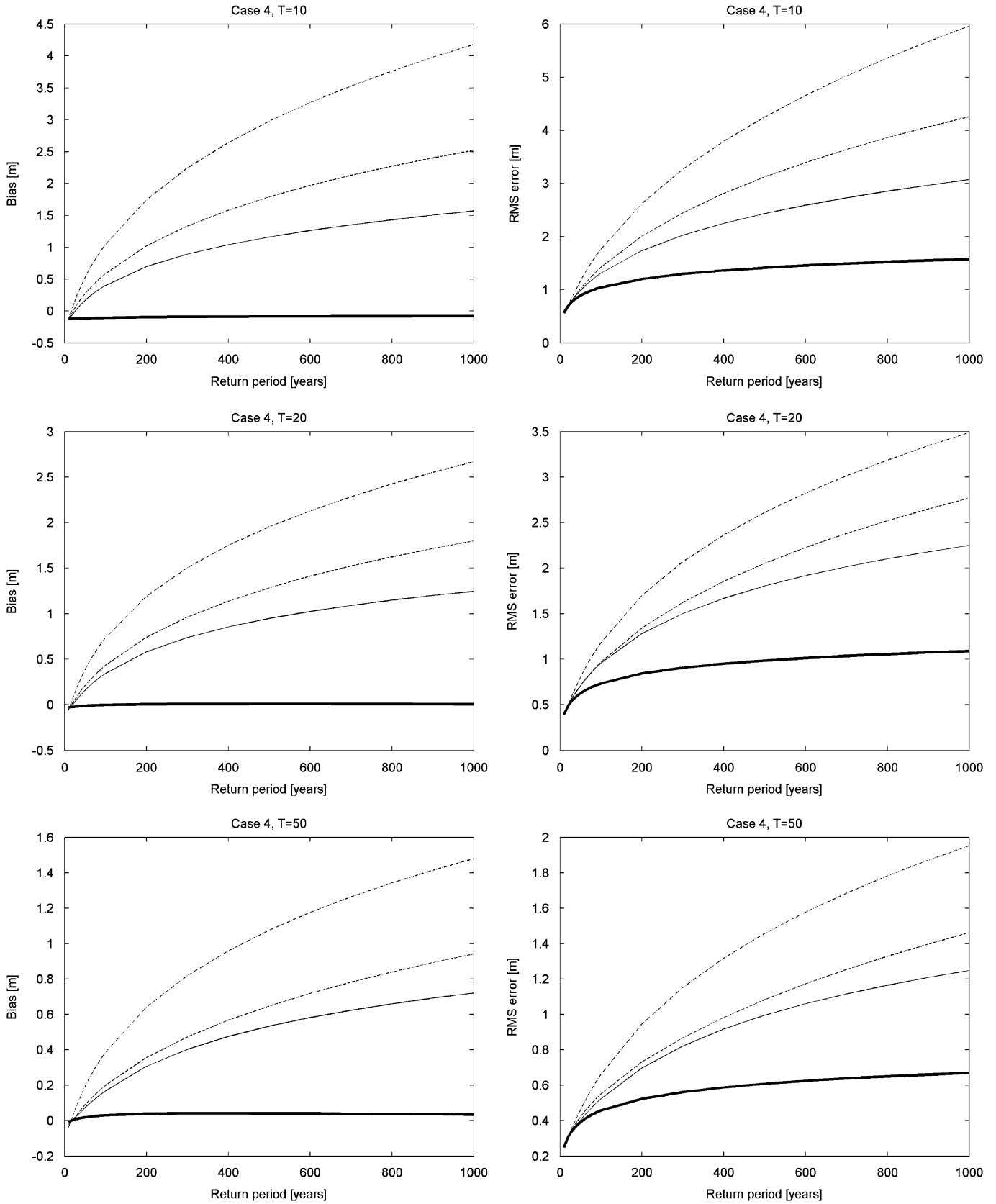
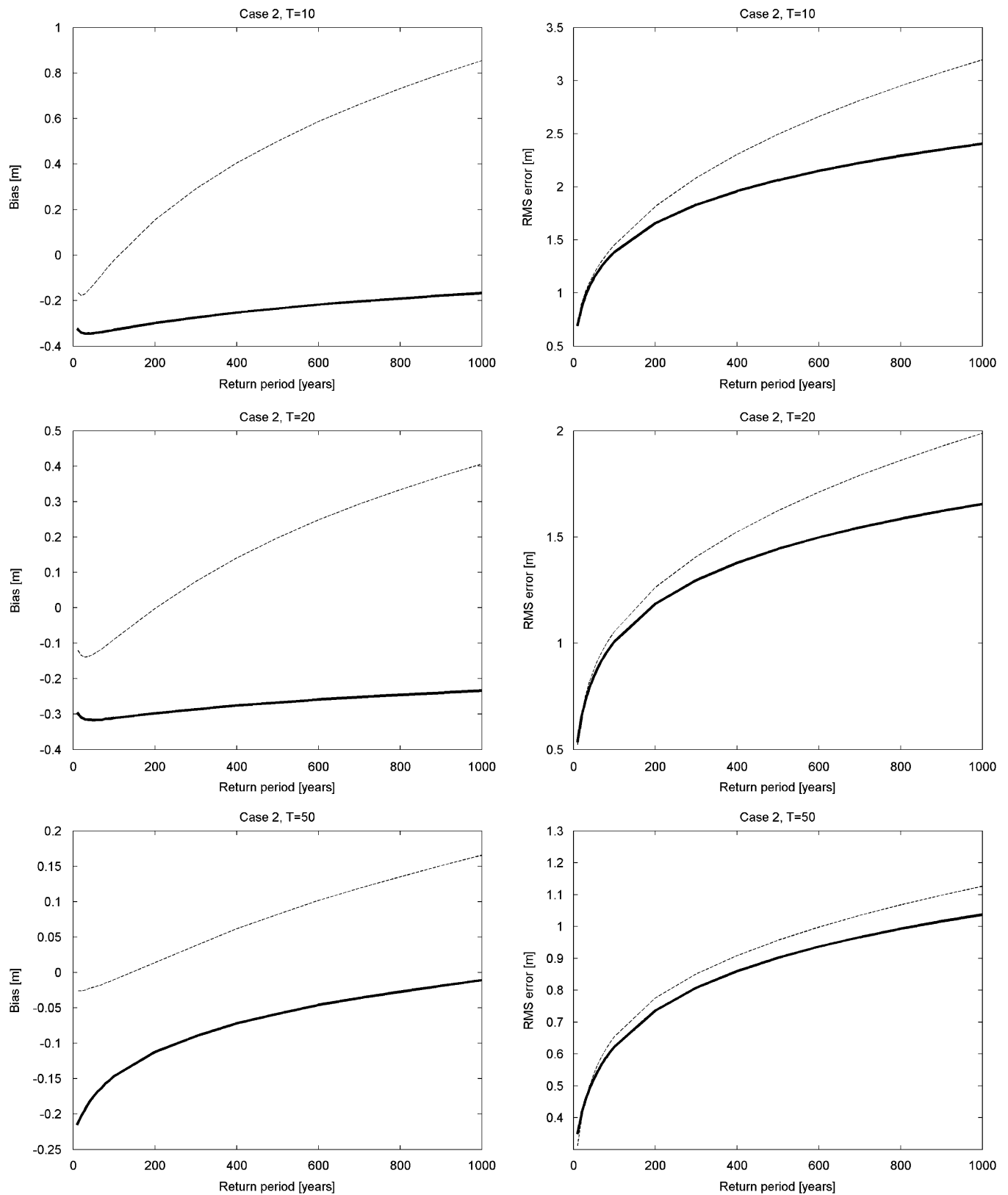


Fig. 16. As previous figure, but for case 4.

parameters varying smoothly throughout the year, the annual distribution is well modelled by another GPD. The seasonal models performed worse because of the greater uncertainty related to

smaller sample sizes. It was shown that this introduces a positive bias into estimates from seasonal models and increases their uncertainty, despite using more data than the non-seasonal models.



**Fig. 17.** Bias and RMS error in return values from the seasonal and non-seasonal models for case study 2 with 30% of data censored from the winter. Bold line: Non-seasonal model. Dashed line: 3-month seasons.

This study has shown that discrete seasonal or directional analyses should not be used to determine annual or omnidirectional return values. An alternative to estimating distribution parameters for discrete directional sectors or seasons is to

estimate them as a smoothly varying function of direction or time (e.g. Anderson et al., 2001; Jonathon and Ewans, 2008). However, any assertions about the accuracy of these models need to be verified using realistic case studies similar to those described here.

## Appendix A. Calculation of annual return values from seasonal parameters

Suppose that the year can be split into  $k$  seasons, not necessarily the same length, over which the distribution of the data can be considered stationary. Let  $X_i$  denote a storm maxima in season in  $i$  and  $X$  denote a storm maxima at any point throughout the year, ignoring season. Let  $F_i(x) = \Pr\{X_i \leq x | X_i > u_i\}$  be the conditional distribution of threshold exceedances in season  $i$ , described by a GPD with parameters  $\xi_i$  and  $\sigma_i$ , with threshold  $u_i$ . For simplicity, it is assumed in the following that  $\xi_i \neq 0$ . In the case that  $\xi_i = 0$  for some  $i$ , then the CDF should be replaced with the appropriate form from Eq. (1).

The derivation of annual return values from the seasonal distributions parallels that given in Section 2 and therefore requires an estimate of the annual distribution function, i.e. the distribution of all storms within a year, ignoring the season. The models for each season are likely to be fitted using different thresholds, so the annual distribution function can only be defined for  $X > u_{\max}$  where  $u_{\max} = \max(u_i; i=1, \dots, k)$ . The distribution function for each season, conditional on exceeding  $u_{\max}$ , is also generalised Pareto, with the same shape parameter, and scale parameter given by Eq. (2):

$$\Pr\{X_i \leq x | X_i > u_{\max}\} = 1 - [1 + \xi_i(x - u_{\max})/\tilde{\sigma}_i]^{-1/\xi_i} \quad (\text{A.1})$$

where  $\tilde{\sigma}_i = \sigma_i + \xi_i(u_{\max} - u_i)$ .

The probability that  $u_{\max}$  is exceeded in each season is given by

$$\Pr\{X_i > u_{\max}\} = \Pr\{X_i > u_i\} \Pr\{X_i > u_{\max} | X_i > u_i\} = \zeta_{u_i} (1 - F_i(u_{\max})) \quad (\text{A.2})$$

where  $\zeta_{u_i} = \Pr\{X_i > u_i\}$ .

To form the annual distribution we must assume that storms in each season are independent. The annual distribution is then given as the weighted average of the distributions from each season, where the weights reflect the probability that a storm within the year, which exceeds  $u_{\max}$ , comes from season  $i$ . The expected number of storms exceeding  $u_{\max}$  in each season is given by

$$r_i = m_i \Pr\{X_i > u_{\max}\} = m_i \zeta_{u_i} (1 - F_i(u_{\max})) \quad (\text{A.3})$$

where  $m_i$  is the total number of samples each season. Weights for each season are then given by the expected number of storms exceeding  $u_{\max}$  from each season divided by the expected number of storms exceeding  $u_{\max}$  each year:  $w_i = r_i/r$ , where  $r = \sum_{i=1}^k r_i$ . The annual CDF, conditional on exceeding  $u_{\max}$ , is then given by the weighted sum of the distribution functions for each season

$$\begin{aligned} \Pr\{X \leq x | X > u_{\max}\} &= \sum_{i=1}^k w_i \Pr\{X_i \leq x | X_i > u_{\max}\} \\ &= 1 - \sum_{i=1}^k w_i (1 + \xi_i(x - u_{\max})/\tilde{\sigma}_i)^{-1/\xi_i} \end{aligned} \quad (\text{A.4})$$

The probability that any storm in the year exceeds  $u_{\max}$  is given by the expected number of storms exceeding  $u_{\max}$  each year, divided by the total number of storms in a year:  $\Pr\{X > u_{\max}\} = r/m$ , where  $m = \sum_{i=1}^k m_i$ . Finally, for  $x > u_{\max}$ , we have that

$$\begin{aligned} \Pr\{X > x\} &= \Pr\{X > u_{\max}\} (1 - \Pr\{X \leq x | X > u_{\max}\}) \\ &= \frac{r}{m} \sum_{i=1}^k w_i (1 + \xi_i(x - u_{\max})/\tilde{\sigma}_i)^{-1/\xi_i} \end{aligned} \quad (\text{A.5})$$

For mixed GPD models the  $N$ -year return value,  $x_N$ , is given by solving

$$\frac{1}{Nm} = \Pr\{X > x_N\} \quad (\text{A.6})$$

or equivalently

$$\frac{1}{Nr} = \sum_{i=1}^k w_i (1 + \xi_i(x_N - u_{\max})/\tilde{\sigma}_i)^{-1/\xi_i} \quad (\text{A.7})$$

In the case  $k=1$  this reduces to Eq. (6), but in all cases this is a monotonic function of  $x_N$  and is easily solved by standard numerical methods.

## References

- Anderson, C.W., Carter, D.J.T., Cotton, P.D., 2001. Wave climate variability and impact on offshore design 4mxtremes. Report for Shell International and the Organization of Oil & Gas Producers, 90 pp.
- Athanassoulis, G.A., Stefanakos, C.N., 1995. A nonstationary stochastic model for long-term time series of significant wave height. *J. Geophys. Res.* 100 (C8), 16149–16162.
- 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 (7), 1032–1048.
- Carter, D.J.T., Challenor, P.G., 1981. Estimating return values of environmental parameters. *Q. J. R. Meteorol. Soc.* 107, 259–266.
- Chaouche, A., Bacro, J.N., 2006. Statistical inference for the generalized Pareto distribution: maximum likelihood revisited. *Commun. Stat. -Theory Methods* 35 (5), 785–802.
- Coles, S., 2001. *An Introduction to the Statistical Modelling of Extreme Values*. Springer-Verlag, London.
- de Zea Bermudez, P., Kotz, S., 2009. Parameter estimation of the generalized Pareto distribution—Parts I & II. *J. Stat. Plan. Inference*, doi:10.1016/j.jspi.2008.11.019, and doi:10.1016/j.jspi.2008.11.020.
- Dupuis, D.J., Tsao, M., 1998. A hybrid estimator for generalized Pareto and extreme-value distributions. *Commun. Stat. -Theory Methods* 27, 925–941.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Ewans, K.C., Jonathan, P., 2008. The effect of directionality on Northern North Sea extreme wave design criteria. *J. Offshore Mech. Arctic Eng.* 130, 041604.
- Ferreira, J.A., Guedes Soares, C., 1998. An application of the peaks over threshold method to predict extremes of significant wave height. *J. Offshore Mech. Arct. Eng.* 120 (3), 165–176.
- Forristall, G.Z., Heideman, J.C., Leggett, I.M., Roskam, B., Vanderschuren, L., 1996. Effect of sampling variability on hindcast and measured wave heights. *J. Waterw. Port Coastal Ocean Eng.* 122 (5), 216–225.
- Grimshaw, S.D., 1993. Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics* 35, 185–191.
- Hardy, T.A., McConochie, J.D., Mason, L.B., 2003. Modeling tropical cyclone wave population of the Great Barrier Reef. *J. Waterw. Port Coastal Ocean Eng.* 129 (3), 104–113.
- Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29, 339–349.
- Jonathan, P., Ewans, K.C., 2006. Uncertainties in extreme wave height estimates for hurricane dominated regions. In: *Proceedings of the 25th International Conference on Offshore Mechanics and Arctic Engineering*, June 4–8, 2006, Hamburg, Germany.
- Jonathan, P., Ewans, K.C., 2007. The effect of directionality on extreme wave design criteria. *Ocean Eng.* 34, 1977–1994.
- Jonathan, P., Ewans, K.C., 2008. Modelling the seasonality of extreme waves in the Gulf of Mexico. In: *Proceedings of the 27th International Conference on Offshore Mechanics and Arctic Engineering*, OMAE2008, June 15–20, 2008, Estoril, Portugal, OMAE2008-57131.
- Jonathan, P., Ewans, K., Forristall, G., 2008. Statistical estimation of extreme ocean environments: the requirement for modelling directionality and other covariate effects. *Ocean Eng.* 35, 1211–1225.
- Leadbetter, M.R., Lindgren, G., Rootzen, H., 1983. *Extremes and Related Properties of Random Sequences and Series*. Springer, New York.
- Mackay, E.B.L., Challenor, P.G., Bahaj, A.S., 2009. A comparison of estimators for the generalised Pareto distribution, in preparation.
- Mathiesen, M., Goda, Y., Hawkes, P., Mansard, E., Martin, M.J., Peltier, E., Thompson, E., van Vledder, G., 1994. Recommended practice for extreme wave analysis. *J. Hydraul. Res.* 32 (6), 803–814.
- Morton, I.D., Bowers, J., Mould, G., 1997. Estimating return period wave heights and wind speeds using a seasonal point process model. *Coastal Eng.* 31, 305–326.
- Moharram, S.H., Gosain, A.K., Kapoor, P.N., 1993. A comparative study for the estimators of the generalized Pareto distribution. *J. Hydrol.* 150, 169–185.
- Naess, A., 1998. Statistical extrapolation of extreme value data based on the peaks over threshold method. *J. Offshore Mech. Arctic Eng.* 120, 91–96.
- Naess, A., Clausen, P.H., 2001. Combination of the peaks-over-threshold and bootstrapping methods for extreme value prediction. *Struct. Saf.* 23 (4), 315–330.
- National Data Buoy Center (NDBC), 1996. Nondirectional and directional wave data analysis procedures. Technical Document 96–01, Stennis Space Center, Miss.

- National Data Buoy Center (NDBC), 2003. Handbook of automated data quality control checks and procedures of the National Data Buoy Center. Technical Document 03– 02, Stennis Space Center, Miss.
- Singh, V.P., Guo, H., 1995. Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME). *Hydrol. Sci. J.* 40, 165–181.
- Stefanakos, C.N., Athanassoulis, G.A., 2006. Extreme value predictions based on nonstationary time series of wave data. *Environmetrics* 17 (1), 25–46.
- Tawn, J.A., 1988. An extreme-value theory model for dependent observations. *J. Hydrol.* 101, 227–250.
- Zhang, J., 2007. Likelihood moment estimation for the generalised Pareto distribution. *Aust. N. Z. J. Stat.* 49 (1), 69–77.