AMERICAN
METEOROLOGICAL
SOCIETY

Journal of Climate

EARLY ONLINE RELEASE

The DOI for this manuscript is doi: 10.1175/2010JCLI3680.1

Global estimates of extreme wind speed and wave height

J. Vinoth and I.R. Young[1]

Faculty of Engineering and Industrial Sciences

Swinburne University of Technology, Melbourne, Australia

[1] *Corresponding author:* I.R. Young, Vice-Chancellor, Swinburne University of Technology, P.O. Box 218, Hawthorn, Vic 3122, Australia. Email: iyoung@swin.edu.au

Abstract

A long term data set of satellite altimeter measurements spanning 23 years, of significant wave height and wind speed is analysed to determine extreme values corresponding to a 100-year return period. The analysis considers the suitability of both the Initial Distribution Method (IDM) and Peaks-over-threshold (POT) approaches and concludes that for wave height both IDM and POT methods can yield reliable results. For the first time, the global POT results for wave height show spatial consistency, a feature afforded by the larger dataset. The analyses also show that the POT approach is sensitive to spatial resolution. Since wind speed has greater spatial and temporal variability than wave height, the POT approach yields unreliable results for wind speed as a result of under-sampling of peak events. The IDM approach does, however, generate extreme wind speed values in reasonable agreement with buoy estimates. The results show that the altimeter data base can estimate 100-year return period significant wave height to within 5% of buoy measurements and the 100-year wind speed to within 10% of buoy measurements when using the IDM approach. Due to the long data set and global coverage, global estimates of extreme values can be developed on a $1^{\circ} \times 1^{\circ}$ grid when using the IDM and a coarser $2^{\circ} \times 2^{\circ}$ for the POT approach. The high resolution $1^{\circ} \times 1^{\circ}$ grid, together with the long duration of the dataset means that fine scale features not previously identified using altimeter data are clearly apparent in the IDM results. Goodness of fit tests show that the observed data conform to a Fisher-Tippett Type 1 (FT-1) distribution. Even in regions such as the Gulf of Mexico, where extreme forcing is produced by small scale hurricanes, the altimeter results are consistent with buoy data.

1. Introduction

The determination of extreme values of wind speed and wave height is a common requirement in many offshore applications, such as offshore structural design and operation. The usual approach is to estimate the value of the 50 or 100 year return period wind speed or wave height based on a measured time series of limited duration. Here, for example, the 100 year return period wave height, is the wave height which would be exceeded on average once in 100 years. As measured records are not available for 50 or 100 years, the normal approach is to fit a probability distribution to recorded data and extrapolate this to the required probability level (return period). Obviously, the accuracy of such an approach is limited by factors such as the length of the recorded time series and the ability of the chosen probability distribution to represent the extreme tail (low probability of occurrence) of the distribution. From a practical point of view, a major challenge is that long duration time series are seldom available at a particular location of interest. In such situations, it is common to use numerical model hindcast results in place of recorded data. Naturally, model data is only as reliable as the model physics and forcing wind fields. Hence, such an approach introduces an additional element of potential error. This is particularly the case when considering hindcasts of events many years in the past, where wind fields may be of low quality.

In recent years, the advent of satellite altimeter observations has provided the prospect of global coverage of measurements of wind speed and wave height. A number of studies have applied this data to the estimation of extreme wave and, to a lesser extent, wind conditions. Such studies have highlighted a number of unique challenges associated with the use of such data for extreme value analysis. Although altimeter data exists for approximately 25 years, previous studies have used only relatively short records, with a maximum duration of approximately 11 years. Use of longer duration time series would involve combining data from a variety of different platforms and hence it would be necessary to ensure a consistent calibration of these multi-platform measurements over this extended period. The sampling pattern of the altimeter is also quite different to buoy data. In comparison to buoy data, altimeters sacrifice temporal coverage for high spatial density. That is, the satellite track revisits a particular location infrequently (typically of order 10 days) but provides excellent spatial coverage (typically observations every 5 to 7 km along the ground track). As a result, previous studies have averaged satellite data in a region surrounding the point of interest. Hence, point buoy measurements and spatial altimeter data measure slightly different quantities.

The present analysis extends these previous studies by using a far longer satellite altimeter data set, consisting of measurements from 7 separate altimeter missions dating back to 1985. This combined data set has been consistently calibrated and validated by Zieger at al (2009). The extended data set provides a much more suitable basis for the assessment of the use of such data for extreme value analysis and to develop appropriate techniques for the estimation of extreme wind speed and wave height over the world's oceans. In addition to developing such techniques, the present analysis also investigates the global distribution of extreme (100 year return period) wind and wave conditions.

The arrangement of the paper is as follows. Section 2 provides a brief review of extreme value theory as applied to ocean wind and waves and an overview of previous applications of these approaches to altimeter data. Section 3 provides a summary of previous studies of extreme wind and wave climate using altimeter data. The altimeter data set to be used in this analysis is described in Section 4, followed by validation and testing against buoy data of the application of extreme value analysis to this data in Section 5. Section 6 examines the global distribution of extreme wind speed and wave height. Finally, conclusions and discussion of the results are outlined in Section 7.

2. Introduction to Extreme Value Analysis

2.1 Theoretical Distributions

As outlined by Gumbel (1958), the aim is to determine the probability distribution from a sample of measured data. To form a valid distribution, the observations should be independent and identically distributed. The requirement for independence means that successive observations should not be correlated with one another. As typical wind generation systems will have durations of a number of hours, this implies that wind/wave observations may need to be separated by many hours to satisfy this condition. The requirement of identically distributed observations means that should an area be exposed to quite different meteorological forcing events (e.g. trade winds and tropical cyclones), these systems should be considered separately.

Noting these basic requirements, there are three approaches which have been applied: the Initial Distribution Method, the Annual Maximum Method and the Peaks-over-Threshold Method.

The Initial Distribution Method (IDM) uses all recorded data and fits a Cumulative Distribution Function (CDF) to this data. As there is no theoretical means to determine the most appropriate function, a number of CDFs are typically used, the one achieving the best fit to the data generally being accepted. Typical CDFs used with the IDM include (see Tucker, 1991; Goda, 1988):

The Fisher-Tippett Type 1 (FT-1) or Gumbel distribution

$$F(x) = \exp\left[-\exp\left(-\frac{x-A}{B}\right)\right] \quad (1)$$

where $F(x)$ is the CDF of the variable x and A and B are parameters determined by the fitting process.

The Weibull two-parameter (W2P) distribution

$$F(x) = 1-\exp\left[-\left(\frac{x}{B}\right)^k\right] \quad (2)$$

where k is also a fitting parameter.

The Weibull three-parameter (W3P) distribution

$$F(x) = 1-\exp\left[-\left(\frac{x-A}{B}\right)^k\right] \quad (3)$$

The parameters in these distributions are generally determined by fitting the model CDF to the empirical cumulative distribution, either by least squares, the method of moments or the maximum likelihood method (see Holthuijsen, 2007). Again, there is no theoretical way to choose between these fitting methods. Historically, (1) has been fitted using both the method of moments (in which case it is called the FT-1 distribution) and maximum likelihood (where it is called the Gumbel distribution) (Evans et al, 2000). In the remainder of the paper, the terminology FT-1 will be used when the distribution has been fitted with the method of moments and FT-1G when the distribution has been fitted using the maximum likelihood method. The Weibull distribution, together with the Annual Maximum and Peaks-over-threshold methods considered below, are most commonly fitted using maximum likelihood (Evans et al, 2000).

The IDM suffers from two obvious deficiencies. Firstly, as buoy observations are typically made at either hourly or three-hourly intervals, the data almost certainly violate the requirement that they are independent. Secondly, interest is in the tail of the distribution (extreme values), although the vast majority of the data used to fit the distribution comes from moderate conditions, that is, the body of the distribution. Thus, extrapolation to extreme events may result in significant error.

Despite these shortcomings, the IDM is commonly used (see for example Goda, 1992, 1988; Ochi, 1992, Tucker, 1991).

One means to overcome the issue of independence of the data is to use only the maximum value in a 12 month period, the Annual Maximum Method (AMM). Extreme-value theory indicates that the cumulative distribution of maxima will follow a generalized extreme-value (GEV) distribution (Castillo, 1988)

$$F(x) = \exp\left\{ -\left[1 + k\left(\frac{x-A}{B} \right) \right]^{-1/k} \right\} \tag{4}$$

In addition to addressing the issue of independence, the AMM also provides a theoretical rationale for the choice of the CDF (4) and uses only extreme values associated with the tail of the distribution to determine the fitting parameters. However, as only one value per year is used, the available data is typically very small. As such, it is generally not practical to use the AMM approach with ocean wind and wave data.

A compromise between these two approaches is the Peaks-over-threshold (POT) method (Goda, 1992; Van Gelder and Vrijling, 1999; Ferreira and Soares, 1998).

Under this approach, a threshold value is defined and a peak value is associated with each rise and fall of the measured data above the threshold. The rationale for this approach is that, if the threshold is selected correctly, the peak value associated with each storm (an extreme event) will be recorded. From extreme-value theory (Castello, 1988; Coles, 2001), the distribution of the maxima in such a sequence of values above a threshold follows the Generalised Pareto Distribution (GPD)

$$F(x) = 1 - \left[1 + k\left(\frac{x - A}{B}\right)\right]^{-1/k} \tag{5}$$

Although not having the same theoretical basis as (5), the Weibull distribution (3) has also been used to fit POT data (Soares and Henriques, 1996).

Although the theoretical basis for the POT is compelling, the major issue is in selecting the value of the threshold. As pointed out by Coles (2001), there is no theoretical basis for the selection of such a value. For prediction at a single location, examination of storm data may enable a threshold value to be objectively determined. Alves and Young (2003) attempted this on a global scale using a database of storm conditions. A more straight forward approach has been proposed by Anderson et al (2001), Caires and Sterl (2005) and Challenor et al. (2005), who propose simply to consider all data above a percentile value. Both the 90th and 93rd percentile have been suggested as reasonable limits.

2.2 Extrapolation to Extreme Values

Once the appropriate CDF is adopted, the aim is to use this distribution to determine the extreme wind speed or wave height. Therefore, it is necessary to determine the probability level which is associated with the 100-year event (exceeded once in 100 years), $P\left(x < x^{100}\right)$, where x can be either H_s or U_{10}. Following Goda (1988) and Mathiesen et al (1994), for the POT approach, P can be determined as

$$P\left(x < x^{100}\right) = 1 - N_Y / \left(100 N_{POT}\right) \tag{6}$$

where N_{POT} is the number of data points in the POT analysis and N_Y is the number of years covered by the analysis.

For the IDM analysis, P is commonly determined as

$$P\left(x < x^{100}\right) = 1 - D / T_{100} \tag{7}$$

where D is a decorrelation time scale in hours for observations of x and T_{100} is the number of hours in 100 years. Consistent with the decorrelation time scale used in previous studies (Tucker, 1991; Cooper and Forristall, 1997; Teng, 1998), a value of $D = 3$ hours was used. Values of D ranging from 1 hour to 24 hours were tested. The resulting values of 100-year return period significant wave height, H_s^{100} and 100-year return period wind speed, U_{10}^{100} did not vary significantly with the chosen value of D.

In the subsequent analysis, both the IDM and POT will be used to determine these extreme values. In order to clarify the method used to estimate the extreme value, the terminology H_s^{100} *(IDM)* or H_s^{100} *(POT)* (and similarly for wind speed) will be adopted. If no parentheses are included, then reference is to the generic extreme value, independent of the method of calculation.

2.3 Goodness of Fit

In the application of any of the proposed extreme value approaches, the key issue is how well the recorded data fits the selected extreme value CDF. The "goodness of the fit" to the data can be determined from the statistics of the residual between the empirical cumulative distribution function [EDF, $F_n(x)$] (i.e. the distribution obtained from the recorded data) and the theoretical extreme value CDF model, $F(x)$ (see Stephens, 1986). There are two general techniques which have been proposed to measure this discrepancy.

Supremum statistics consider the largest positive and largest negative differences between the EDF and the CDF, where D^+ is the largest positive difference and D^- is the largest negative difference

$$D^+ = \sup_x \left[F_n(x) - F(x)\right] \tag{8}$$

$$D^- = \sup_x \left[F(x) - F_n(x)\right] \tag{9}$$

$$D = \sup_x \left|F_n(x) - F(x)\right| = \max\left(D^+, D^-\right) \tag{10}$$

The statistic D is the basis of the Kolmogorov–Smirnov test.

Quadratic statistics measure the discrepancy as

$$Q = n\int_{-\infty}^{\infty}\left[F_n\left(x\right)-F\left(x\right)\right]^2 \psi(x)\,\mathrm{dF}\left(x\right) \qquad (11)$$

When $\psi = 1$, (11) defines the Cramér-von Mises statistic, W^2 and when $\psi(x) = \left[F(x)\right]\left[1-F(x)\right]^{-1}$, (11) defines the Anderson and Darling (1952) statistic, A^2.

The values of D, W^2 and A^2 can then be tested at the desired percentage point level to determine whether the null hypothesis (that the EDF is well approximated by the chosen CDF) is true or false.

Goda and Kobune (1990) have proposed an alternative to the above approaches, by simply looking at the value of the correlation coefficient between the data x and its reduced variate (e.g. The FT-1 distribution (1) can be expressed in linear form as $y = \left(x - A\right)/B$, where $y = -\ln\left\{-\ln\left[F\left(x\right)\right]\right\}$ is the reduced variate. In this case, the Goda and Kobune (1990) approach would evaluate the correlation coefficient between y and x.) A value of the correlation coefficient near one, indicates a good fit.

As interest is concentrated on how well the shape of the extreme value tail of the distribution is represented, rather than the body of the distribution, a variation of the above approaches is to "censor" the data by only applying the goodness-of-fit test to data above an upper percentile (Stephens, 1986). For instance, the tests could be applied only to the upper 20% of data.

Tabulated values against which calculated values of D, W^2 and A^2 can be tested have been present for the FT-1, FT-1G and Weibull (2 and 3 parameter) distributions by Stephens (1970, 1977, 1986) and Goda and Kobune (1990). Embrechts et al (1997) and Koltz and Nadarajah (2000) have developed approaches to determine confidence limits for the GPD. Similar tests or confidence limits for the GEV do not appear to be available in the literature.

3. Altimeter observations of extreme wind and wave climate

A number of studies have demonstrated that altimeter data can be used to determine global wind and wave climatology, including monthly mean values and exceedence probabilities up to the 90th percentile (Challenor et al, 1990; Tournadre and Ezraty, 1990; Young, 1994, 1999; Young and Holland, 1996). The first attempt to determine extreme wave heights from altimeter data was conducted by Carter (1993). He considered the 3 years of GEOSAT data for the North Atlantic. The data was "gridded" in $2^o \times 2^o$ bins and the IDM was used with a FT-1 distribution. Despite the very short record, the approach yielded values of 50-year return period significant wave height, H_s^{50} *(IDM)* consistent with buoy and climate atlas results for the area.

Panchang et al (1999) carried out a similar study, comparing H_s^{50} *(IDM)* from the altimeter with buoys along the Atlantic coast of North America. They concluded that short duration altimeter records can be used to estimate extreme values, but noted the "rule of thumb" suggestions of Hogben (1988) that the data duration should be at least a third of the return period (implying 16 years for H_s^{50}).

Alves and Young (2003) used a 10-year data set from GEOSAT, ERS-1 and Topex/Poseidon. In contrast to previous studies they applied both the IDM (FT-1 distribution) and POT (W3P distribution) to estimate global values of H_s^{100} . For POT calculations, a spatially variable threshold was used, based on a global storm data base. They noted that the POT results were spatially highly variable, indicating the sensitivity of the choice of threshold and the magnitude of the errors associated with extrapolating the short duration time series to H_s^{100} *(POT)* values. As a result, they concluded that the IDM was superior for application to short duration altimeter records.

Challenor et al (2005) and Wimmer et al (2006) considered approximately 11 years of data from ERS1, ERS2 and Topex/Poseidon. In a similar fashion to Alves and Young (2003) they considered both the IDM (FT-1 distribution) and the POT with a

$2^o \times 2^o$ grid, but applied the GPD distribution. Also, they set the threshold for the POT as the 90th percentile value. They found that this approach gave good agreement when compared to a single North Atlantic buoy (NODC 44004). The spatial diagrams for the POT approach provided for the North Atlantic did, however, show significant spatial variability, as also noted by Alves and Young (2003).

Chen et al (2004) calculated global values of both U_{10}^{100} and H_s^{100} based on the IDM and a FT-1G distribution from 8 years of Topex data. In contrast to other studies they used a $1^o \times 1^o$ grid. They concluded that the differences between buoy and altimeter derived extreme values were approximately 10% for wind speed and 5% for significant wave height.

4. Altimeter data base

As noted above, in order to obtain reliable estimates of extreme values of wind speed and wave height, it is desirable to have as long a data set as possible. Previous studies of the use of altimeter data for extreme value analysis have used data sets less than 11 year long. In order to assess 50 or 100 year return period statistics, it is desirable to have a data set at least spanning two decades. In addition, it is essential that the calibration of the measurement instruments has been consistent over this period of time, so as to minimize the effects of changes in measurement platform over the period. For this purpose, the altimeter database of Zieger et al (2009) has been adopted. In the 23 year period 1985-2008, a total of 7 separate satellite altimeter missions were operational (see Figure 1). This data set provides almost continuous coverage over this period, with a break in the period 1990-1991.

Zieger et al (2009) calibrated each of the satellite altimeters for wind speed and wave height against deep water insitu buoys. These calibrated results were then cross-validated between satellites operating at the same time, by considering co-located measurements at cross-over points between the satellites. In this manner, any discontinuities due to changes of hardware or software or instrument drift were eliminated. The resulting dataset is believed to provide a consistent, high quality, global view of oceanic wind speed and wave height throughout this extended period.

As such, it represents a unique tool to investigate extreme values of wind speed and wave height on a global scale.

5. Buoy – Altimeter Extreme Value Comparisons

Development of the appropriate techniques for the extreme value analysis of altimeter data and the assessment of the performance of these techniques was undertaken by comparison with a number of deep water NODC (US National Ocean Data Center) data buoys (Evans et al, 2003). The buoys used in this analysis are the same as those adopted by Alves and Young (2003) and a subset of those used by Zieger et al (2009) for the calibration of the data base. These buoys were operational throughout the full 23 year duration of the data set and generally recorded each of wind speed and significant wave height at a 1 hour interval (note that some buoys used a 3 hour interval in the early years of the data set). The buoy locations were chosen such that they were a minimum of 200km offshore and in a water depth of at least 300m. These criteria were chosen to ensure: that altimeter data would not be influenced by the transition from land to water and an averaging area of $2^{o} \times 2^{o}$ centred on the buoy would not include significant regions of finite depth. Full details of the data processing are given by Zieger et al (2009). As shown in Figure 2, the buoy locations cover a variety of geographical regions including: north Pacific, north Atlantic, Gulf of Mexico and the Pacific trade-wind belt (Hawaii).

5.1 Averaging area

The significant difference between buoy and altimeter data is that buoys provide high temporal density (1 hour measurements) but poor spatial density (single or small number of points), whereas the altimeter has poor temporal density (ground tracks are repeated on the order of once every 10 day) but excellent spatial coverage (data point every 5-7km along the track). The approach which has been traditionally used is to define an area and combine all altimeter data for this area, the assumption being that the wave climate in the area is homogeneous. As the area becomes larger, the validity of the assumption of homogeneity clearly becomes questionable. However, the larger amount of data will yield an extreme value estimate with smaller confidence limits (i.e. apparently more stable).

A variety of averaging areas have been used in previous studies. Carter (1993), Alves and Young (2003), Challenor et al (2005) and Wimmer et al (2006) all adopted a $2^{o} \times 2^{o}$ region, whereas Chen et al (2004) concluded that a smaller $1^{o} \times 1^{o}$ region yielded acceptable results. Panchang et al (1999) compared a number of different sized regions and finally adopted an even smaller 50 km radius. In contrast, based on numerical experiments, Cooper and Forristall (1997) recommended a larger region of 200-300km. Because of the significantly larger data set available in the present study, the influence of averaging area can be directly investigated.

A number of different sized regions were considered, including: $1^{o} \times 1^{o}$, $2^{o} \times 2^{o}$ and a 50km radius. In each case, these regions were centred on each of the buoys detailed in Figure 2. Each pass of the satellite across the averaging region will generate a number of observations of wind speed and wave height, these observations clearly not being independent. Wimmer et al (2006) refer to this process as clustering of the data. In this analysis a number of different approaches to de-clustering were investigated, including: using all the data, replacing the pass with the mean, the median or the maximum. All of these approaches yielded very similar results. Ultimately, the median value was adopted as this provided a stable measure, is statistically valid (i.e. independence of observations) and is consistent with previous studies (eg. Wimmer et al, 2006).

The relative error of the altimeter estimate of the extreme wave height can be defined as $\Delta r = \left(H_{s_{alt}}^{100} - H_{s_{buoy}}^{100} \right) / H_{s_{buoy}}^{100}$, where the subscripts refer to the altimeter and buoy values respectively. Values of Δr for each of the buoys and each size of averaging area are shown in Table 1 for the IDM and FT-1 distribution and Table 2 for the POT and W3P distribution. In addition, the values summed over all n buoys of $r_1 = 1/n \sum |\Delta r|$ and $r_2 = 1/n \sum \Delta r$ are shown. The IDM results for r_1 and r_2 are very similar for all of the averaging areas and the values of Δr across all individual buoys are generally less than $\pm 6\%$. The values of r_1 are all less than 3.5% and values of r_2 are all less than $\pm$ 1.0%.

The results of Table 1 clearly show that the IDM approach is insensitive to the averaging area. Adopting the larger $2^{\circ}\times 2^{\circ}$ averaging area increases the amount of data available to construct the altimeter CDF, but this does not result in better agreement between buoy and altimeter.

In addition, the 50km radius data was resampled such that only cases where the buoy and altimeter measurements were sampled within 30 minutes of each other were considered. Although this resulted in comparable values of sampled wave height, it significantly reduced the actual number of observations in the fitting process for the extreme value distribution. As a result, r_1 actually increased slightly to 3.07%. In situations where there are strong spatial gradients of wave height/wind speed increasing the size of the averaging area may compromise accuracy. Even for the Gulf of Mexico buoys (42001 and 42002), which are subject to the small scale forcing of hurricanes, the largest area tested ($2^{\circ}\times 2^{\circ}$) does not increase the relative error.

It should be noted that the excellent agreement between buoy and altimeter derived extreme values obtained using the IDM approach, does not validate this extreme value method for altimeter data. Both the buoy and altimeter approaches using the IDM are estimates of the extreme value. It is possible that neither are a good estimates of the actual extreme value (the same argument holds for all approaches). Rather, the good agreement indicates that the altimeter data gives a very similar estimate of the CDF to the buoy and that this estimate is insensitive to the averaging area. The potential error in estimating the extreme value is introduced in extrapolating the limited duration data set to the extreme value. Here, the longer duration data set available in this study will reduce the extent of this extrapolation and hence the potential for error (compared to previous altimeter studies).

As a result of the comparisons in Table 1, subsequent analysis of the IDM approach has generally concentrated on the $1^{\circ}\times 1^{\circ}$ averaging area, as the finer resolution has the potential to identify features in areas where there may be significant spatial gradients of wave height.

The results for the POT and W3P distribution comparisons between buoy and altimeter in Table 2 are quite different to the IDM results of Table 1. The values Δr, r_1 and r_2 are significantly larger for the POT analysis. Also, the averaging area has a clear impact on the results for the POT analysis. For the $1^{\circ}\times 1^{\circ}$ analysis the values of Δr at every location are negative, indicating that the altimeter approach underestimates compared to the buoy. This underprediction decreases when the averaging area is increased to $2^{\circ}\times 2^{\circ}$ (i.e. r_2 reduces in magnitude from -20.20% to -11.73%). The underprediction disappears completely when the data is resampled, such that the buoy and altimeter data used are collocated in space and time ($r_2 = 3.34\%$). This result clearly indicates that, given comparable data, the POT approach will yield similar results for buoy and altimeter data. However, as the available altimeter data is limited, undersampling impacts on the POT estimated extreme values. Increasing the size of the averaging area reduces this effect, but does not completely eliminate it.

This result is not surprising. The IDM approach defines the body of the CDF and then attempts to extrapolate to extreme events. Such an approach is insensitive to undersampling, as the body of the CDF can be well defined with limited observations (the error comes in the extrapolation). In contrast, the POT approach attempts to define the extreme value tail of the CDF from observations of "peak" events. If undersampling means that insufficient "peaks" are measured, the result will directly impact on the resulting tail of the distribution (the observed distribution).

As the POT approach appears to be sensitive to the averaging area, in subsequent analysis, both $1^{\circ}\times 1^{\circ}$ and $2^{\circ}\times 2^{\circ}$ averaging regions are considered.

5.2 Initial Distribution Method

The IDM was applied to $1^{\circ}\times 1^{\circ}$ regions centred on each of the buoys for the FT-1 and FT-1G distributions. The resulting 100 year return period values for both buoy and altimeter are shown in Table 3 for H_s^{100} *(IDM)* and Table 4 for U_{10}^{100} *(IDM)* (note that the $2^{\circ}\times 2^{\circ}$ result is also shown, but not discussed here). Buoy and altimeter results are in excellent agreement for H_s^{100} *(IDM)* for both distributions, as noted in Table 1.

The distributions yield values of $r_1 = 2.11\%$ for the FT-1 distribution and $r_1 = 2.92\%$ for the FT-1G distribution. These values of r_1 are approximately half those reported by both Alves and Young (2003) and Chen et al (2004) using IDM FT-1 and IDM FT-1G respectively. The much larger data sets and longer duration of the time series significantly reduces the potential error in extrapolating the extreme value distributions. As a result, the buoy and altimeter results for the extreme values are in better agreement than these previous studies.

Figure 3a shows the values of H_s^{100} *(IDM)* (FT-1) at each location calculated for both the altimeter and buoy, together with 95% confidence limits calculated using the residual of the correlation coefficient (REC approach) proposed by Goda (2000). The excellent agreement at all buoy locations between buoy and altimeter H_s^{100} *(IDM)* is clear. As shown in the figure and Table 3, the large data sets mean that the confidence limits for H_s^{100} *(IDM)* are very small. Goodness-of-fit tests were performed for each of the Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling and Goda tests, applied to the top 20% of data. As the number of data points in the sample increases these tests become progressively more demanding to satisfy. Hence, for the present data set, they are extremely demanding. The number of the tests that each of the distributions satisfied is shown in Table 3. The majority of the altimeter FT-1 results satisfied 2 or 3 tests. The larger buoy data sets result in very small confidence intervals and all locations satisfying only 1 test. It is, however, clear that the results for H_s^{100} *(IDM)* are very similar for both buoy and altimeter, and both conform to the assumed FT-1 cumulative distribution.

Although both FT-1 and FT-1G results agree well between buoy and altimeter, the FT-1G results are generally smaller than the FT-1 results (for both buoy and altimeter). The FT-1G results for buoy data are approximately 8% lower than the corresponding FT-1 buoy results, when averaged over all the buoys. This clearly demonstrates that the IDM is sensitive to how the extrapolation of the CDF is performed. The FT-1G altimeter results also satisfy less goodness-of-fit tests, with many locations satisfying none of the tests.

Table 4 shows the comparable IDM results for wind speed, U_{10}^{100} *(IDM)*. The values of r_1 for each of the distributions are larger than the comparable values for H_s^{100} *(IDM)* - for FT-1, $r_1 = 9.91\%$ and for FT-1G, $r_1 = 6.74\%$. These larger differences between buoy and altimeter reflect the greater spatial variability of wind speed compared to wave height and are consistent with the conclusion by Chen et al (2004) that differences between buoy and altimeter extreme values are larger for wind speed than wave height. Figure 3c shows a comparison at each buoy, together with the respective 95% confidence limits. It is clear that the agreement is not as good as for wave height (Figure 3a) and the buoy values often fall outside the confidence limits for the altimeter. The goodness-of-fit tests (see Table 4) show that the FT-1 distribution is slightly superior to the FT-1G distribution.

As the buoys used for the present comparisons are the same as adopted by Alves and Young (2003), a direct comparison can be made. The mean error between the two sets of altimeter estimates of H_s^{100} *(IDM)* (as measured by Δr between the two sets of altimeter estimates) was less than 5% at each buoy location, indicating comparable results.

5.3 Peaks-over-Threshold

The POT approach was also applied at each buoy location to the altimeter, and buoy data within both $1^o \times 1^o$ and $2^o \times 2^o$ regions around each buoy. Both the GPD and W3P distributions were tested. Before the POT can be applied, it is necessary to define the threshold to be used. As noted above, there is no theoretical basis for selection of the threshold. The only practical approaches for application on a global scale are a fixed threshold (or range of fixed thresholds, which has little physical rationale), or the fixed percentile approach proposed by Anderson et al (2001), Caires and Sterl (2005) and Challenor et al. (2005). Selection of such a percentile approach was explored in some detail. Figure 4 shows values of H_s^{100} *(POT)* calculated at buoy 46002 for different values of the percentile threshold. Results are shown for both the W3P and GPD distributions. It is clear from this figure that there is no clear choice for the selection of the threshold. As the threshold increases, the value of the resulting H_s^{100} *(POT)* increases smoothly for the W3P approach. The GPD is less sensitive to the choice of the threshold. As pointed out by previous researchers (Alves and Young, 2003; Challenor et al, 2005; Wimmer et al, 2006) selection of the threshold is a serious shortcoming of the POT approach. Consistent with Challenor et al (2005) a value of 90% has been selected for further analysis, although it should be pointed out that the selection of an alternative level would impact the calculated values of extreme wind speed and wave height.

Table 3 provides values of H_s^{100} *(POT)* using the POT with W3P and GPD distributions at each buoy location and Table 4 provides the corresponding values of U_{10}^{100} *(POT)*. The mean error values for H_s^{100} *(POT)* are: GPD - $r_1 = 17.31\%$ and W3P - $r_1 = 12.58\%$ ($2^o \times 2^o$ region); U_{10}^{100} are: GPD - $r_1 = 40.62\%$ and W3P - $r_1 = 11.26\%$ ($2^o \times 2^o$ region). These values are significantly larger than values for the IDM, indicating significantly poorer agreement between buoy and altimeter for the POT. This poorer agreement between buoy and altimeter for the POT approach is also clear in Figure 3b (for wave height) and Figure 3d (for wind speed). This figure also shows the impact of the sampling region, with the $2^o \times 2^o$ region reducing the level of underprediction. The much wider confidence interval associated with the POT estimates can also be seen in these figures and Tables 3 and 4. Despite the larger confidence limits and the much less demanding criteria for goodness-of-fit, the POT W3P results satisfy a similar number of tests to the IDM results.

The POT and IDM results for H_s^{100} are in reasonable agreement. Both GPD and W3P POT H_s^{100} *(POT)* buoy values are approximately 3% lower than the corresponding buoy IDM FT-1 results when averaged across all buoys. In contrast, the wind speed results show much greater variability. The IDM FT-1G U_{10}^{100} *(IDM)* buoy values are approximately 11% higher than the corresponding IDM FT-1 buoy values. The differences between the POT and IDM U_{10}^{100} estimates are even greater. The POT GPD and POT W3P buoy estimates of U_{10}^{100} *(POT)* are 33% and 26% lower, respectively than the IDM FT-1 buoy estimates, when averaged across all buoys. Wind speed has much greater spatial and temporal variability than wave height and this is clearly reflected in the broader range of values of U_{10}^{100} resulting from these different estimation techniques.

The results of the buoy – altimeter comparisons of extreme value estimates can be summarised as follows.

For wave height H_s^{100}

- The IDM and POT approaches produce similar results when applied to the same buoy data, irrespective of the CDF assumed (i.e. FT-1, FT-1G, GPD, W3P)
- Altimeter IDM results are in good agreement with the buoy estimates and are not sensitive to the averaging area used (i.e. $1^\circ \times 1^\circ$ verses $2^\circ \times 2^\circ$)
- The POT altimeter results show greater variability than the IDM results when compared to corresponding buoy estimates. In addition, the POT results are sensitive to the averaging area, with a $1^\circ \times 1^\circ$ averaging region significantly underestimating the extremes and a $2^\circ \times 2^\circ$ region moderately underestimating the values.

For wind speed U_{10}^{100}

- There are significantly greater differences between buoy derived values of extreme wind speed obtained by IDM and POT approaches than is the case for wave height
- The differences between buoy and altimeter extreme value estimates are smaller for the IDM than the POT, although larger than for wave height
- Whereas wave height POT GPD and W3P gave similar error results (comparing buoy to altimeter), for wind speed, the POT GPD produces significantly larger differences than the corresponding POT W3P.

As a result of these comparisons, the global distributions to be considered below have adopted the following approaches:

- IDM FT-1 and $1^\circ \times 1^\circ$ averaging area
- POT W3P and both $1^\circ \times 1^\circ$ and $2^\circ \times 2^\circ$ averaging areas

5.4 Altimeter calibration at extreme values

Although the H_s^{100} *(IDM)* comparisons above indicate quite good agreement between buoy and altimeter extreme values, it should be remembered that the calibration results of Zieger et al (2009) involved few extreme wind/wave conditions. Therefore, some questions exist over the ability of the altimeter to measure extreme values of wind speed and wave height. As our focus is on the accurate determination of the shape of the extreme tail of the CDF, such data points are likely to significantly influence the accuracy of the results. To investigate the extreme value performance of the altimeter, a $2^o \times 2^o$ region was considered centred on each of the buoys. The full data sets for the respective buoy and altimeter pairs were then considered and percentile-percentile plots (also called Q-Q plots) were developed for each of the buoy-altimeter pairs for both H_s and U_{10}. Figure 5 shows a typical result for Buoy 46002. The altimeter and buoy are in excellent agreement for all values, including the most extreme value plotted, the 99th percentile.

6. Global distribution of extreme values

Based on the results in Section 5, both $1^o \times 1^o$ and $2^o \times 2^o$ grids were considered for the world and all data binned as appropriate. Each pass through a bin was represented by the median value and the resulting observations analysed using the various extreme value methods. For the IDM approach, the FT-1, FT-1G and W2P were all applied and for the POT approach the W3P and GPD were applied. Thresholds set at both 90% and 93% were used for each of these POT distributions. These approaches were used to calculate both U_{10}^{100} and H_s^{100} for every bin. In addition, the Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling and Goda goodness-of-fit tests were applied for each case tested. This full set of combinations is very large and it is not practical to present all results here. Figures 6 to 10 show a sub-set of this full set of cases.

Figure 6 shows colour contour plots of H_s^{100} for each of IDM FT-1 (Figure 6a), POT W3P (Figure 6b) and POT GPD (Figure 6c) for a $1^o \times 1^o$ grid. The corresponding

$2^o \times 2^o$ results are shown in Figure 7. The IDM FT-1G and IDM W2P distributions are not shown. The spatial distributions of these CDFs are both similar to the FT-1. The FT-1G distribution is also similar in magnitude to the FT-1 distribution, but the W2P is approximately 30% to 40% lower than the FT-1. Figure 8 shows the corresponding set of plots for U_{10}^{100} with a $1^o \times 1^o$ grid and Figure 9 with a $2^o \times 2^o$ grid. Figure 10 shows the global distribution of Cramér-von Mises goodness-of-fit results, censored at 20% for both IDM FT-1 H_s^{100} *(IDM)* and U_{10}^{100} *(IDM)* ($1^o \times 1^o$ grid).

The three panels in Figure 6 clearly highlight the differences between the three approaches/distribution functions. The IDM FT-1 (Figure 6a) is spatially consistent with the results of both Alves and Young (2003) and Chen et al (2004). The extreme values at the high latitudes of both hemispheres and the relatively calm equatorial regions are the defining characteristics. As one would intuitively expect, the distributions of H_s^{100} *(IDM)* are spatially relatively smooth. The POT results (Figure 6b, POT W3P; Figure 6c, POT GPD) are, in contrast, more spatially variable, this small scale variability being larger for the GPD results. Clearly, this variability is the result of undersampling in the much smaller data set of peaks above the threshold. The larger data sets available in this study and the method of selecting the threshold (ie. a fixed 90% value, rather than a variable level) do, however, mean that the results still reproduce the major spatial features (large wave heights at high latitude and calm equatorial regions). This is in contrast to the POT results of Alves and Young (2003) and Wimmer et al (2006) which showed much greater spatial variability than reported here. It is believed that the larger data set in the present analysis results in enhanced POT results. As shown by the buoy comparisons in Section 5, the $1^o \times 1^o$ POT results also underestimate the extreme values. This is clear in Figure 6, with the POT results producing lower values than the IDM, particularly in the high latitudes.

In addition to the large scale features outlined above, Figure 6a shows many local features. The high waves, $H_s^{100}\left(IDM\right) > 14$m, associated with the Somali Jet (which is linked to the Asian Monsoon) in the Arabian Sea is a clear local phenomenon. Also, a local maximum is clear within the Bay of Bengal, probably associated with tropical cyclone activity. Within the Mediterranean Sea, a local region of intensification is clear between France and North Africa, consistent with the strong Mistral winds of this region. The effects of land masses are also clear, with areas of reduced wave extremes east of South America and New Zealand, where the strong westerly winds will have reduced fetch to generate high wave conditions. Even the isolated Kerguelen Islands in the southern Indian Ocean cast a significant "wave shadow" in the lee of the islands.

The $2^{o} \times 2^{o}$ results for H_s^{100} (Figure 7) are qualitatively similar to the $1^{o} \times 1^{o}$ results, although as noted above, the POT results are slightly higher due to the undersampling issue. As expected, the results are spatial smoother than the $1^{o} \times 1^{o}$ results. As noted above, these results represent the first published global POT results which are spatially consistent. Whereas the quality of the IDM results will see little improvement with further extending the duration of the altimeter data set, the POT results are clearly gaining in reliability with the longer data sets available.

Figure 10a shows the global distributions of the IDM FT-1 H_s^{100} *(IDM)* and the points (in black) where the Cramér-von Mises goodness-of-fit test (applied to the top 20% of the data) is accepted. With the exception of the Goda test, the other goodness-of-fit tests give very similar results. The Goda test is significantly less demanding and the vast majority of the points across the globe satisfy this criteria. As seen in Figure 10a, regions where the wave climate tends to form a single population satisfy the goodness-of-fit criteria. Areas such as the high latitudes and trade-wind belts tend to have similar meteorological conditions year round and hence result in CDFs which conform well to the (in this case) FT-1 distribution. In contrast, the tropics will have wave climates composed of tropical cyclones and larger scale meteorological events, which are clearly separate populations. Similarly, the Arabian Sea is subject to intense local winds at the time of the Somali jet but otherwise experience locally light winds and Southern Ocean swell. These are quite distinct populations and hence, the goodness-of-fit test fails in this area. A more detailed analysis which separated these populations on a regional basis would be feasible, but beyond the scope of this analysis.

Figure 8 shows the corresponding plots (to Figure 6) of U_{10}^{100} ($1^{\mathrm{o}} \times 1^{\mathrm{o}}$). As with H_s^{100}, the IDM FT-1 result shows a relatively smooth distribution of wind speed with clear extreme wind belts at high latitudes of both hemispheres. The POT results (both W3P and GPD) are very noisy with a high degree of short scale spatial variability. This trend is stronger for the POT GPD results than the POT W3P. In both cases the results are clearly degraded by under-sampling of extreme events. It is clear that variations in sampling density caused by satellite tracks can be seen. The U_{10}^{100} *(POT)* results are poorer than the comparable H_s^{100} *(POT)* values. This occurs because wind systems are generally spatially smaller than the waves they generate. Once generated, waves tend to propagate away from the generation source. Hence, it is necessary to have greater sampling density for wind speed than for wave height.

The $2^{\mathrm{o}} \times 2^{\mathrm{o}}$ results in Figure 9 show some improvement in the spatial consistency of the POT results. Satellite track patterns are, however, still visible and even with this larger averaging region undersampling clearly degrades the results. Whereas the $2^{\mathrm{o}} \times 2^{\mathrm{o}}$ data for H_s^{100} *(POT)* produced quite acceptable results, the corresponding U_{10}^{100} *(POT)* results clearly indicate that a longer duration record and possibly also greater spatial density of observations is required to obtain reliable results for wind speed.

As for wave height, the many small scale features are again evident with the IDM FT-1 approach. One notable difference between the wind speed and wave height results is the extent of “shadowing” down wind of land in the southern hemisphere. This was quite evident in the wave height results but is much less evident for wind speed. This indicates that the effect is caused by the reduction in wave fetch down wind of the land, rather than a reduction in the strength of extreme winds.

The Cramér-von Mises goodness-of-fit test acceptance results for U_{10}^{100} *(IDM)* FT-1 is shown in Figure 10b. The number of points which satisfy the criteria are much reduced compared to wave height. This is consistent with the greater level of spatial and temporal variability of wind speed data. As a result, the data does not conform to the chosen CDF as well as for wave height. This is particularly the case in the equatorial and tropical regions of all oceanic basins. Of course, it must be remembered that this is a particularly demanding test and the Goda test was satisfied at most points. Again, resampling the data on a seasonal basis is likely to improve the goodness-of-fit.

As noted above, the spatial distributions of 100-year return period values for both wind speed and wave height developed in this study with the IDM FT-1 are similar to those of Alves and Young (2003) and Chen et al (2004). The magnitudes of H_s^{100} *(IDM)* values also appear to be very similar to those of Alves and Young (2003). A quantitative comparison with the results of Chen et al (2004) is not possible. However, a visual comparison of global extreme value charts for both wind speed and wave height indicate the Chen et al (2004) estimates are 5% to 10% lower than the present results. This presumably reflects the shorter data set (less than half as long) and the different fitting method for the adopted FT-1 CDF used by Chen et al (2004). One can, however, conclude that the present results are consistent with these previous studies, but enhance the spatial resolution.

7. Discussion and Conclusions

This analysis presents results from a much longer data set of global altimeter observations than previously reported, to determine extreme wind speed and wave height. The analysis shows that the IDM approach with an FT-1 distribution produces results with a mean error less than 5% for wave height and less than 10% for wind speed when compared to the same analysis conducted for buoy data (see r_1 values in Tables 3 and 4). The IDM FT-1 distribution also generally satisfies demanding goodness-of-fit tests on a global basis. The goodness-of-fit tests for the IDM FT-1 are also more often satisfied for wave height than wind speed. This indicates that the FT-1 distribution is a better approximation to the altimeter wave height data than the wind speed data. Such results reflect the nature of the present analysis, where no attempt has been made to separate wind and wave populations in geographical areas where multiple populations might exist. Because of the long duration of the present records, such an approach may be a possible extension of the present work.

Despite the better theoretical under-pinning of the peaks-over-threshold analysis, this approach does not perform as well as the IDM. The POT analysis is much more sensitive to undersampling than the IDM analysis. As a result, the averaging area over which data is binned is an important consideration for the POT analysis. Even with the extended data set, a bin size of $1^{o}\times1^{o}$ consistently underestimates extreme values compared to buoy data. The larger $2^{o}\times2^{o}$ bin size improves results, but there is still evidence that the altimeter POT results are lower than the comparable analysis applied to buoy data. Previous studies which have applied the POT method to altimeter data have shown significant spatial variability. The present $2^{o}\times2^{o}$ POT analysis, which uses a significantly longer altimeter data set, produces the first global plots of extreme significant wave height, using this technique, with an acceptable level of spatial variability. The POT wave height results do capture many of the global features evident in the IDM analysis. However, when applied to wind speed, the POT approach yields unacceptable results. These poor results are attributed to the greater spatial variability of wind speed, compared to wave height. As a result, the data set is still too small to capture sufficient "peaks" to adequately define the extreme tail of the CDF.

The present analysis does, however, show that when applied to buoy data the IDM and POT analyses yield comparable results. The analysis also shows that, given longer altimeter data sets, the POT analysis will produce acceptable extreme value altimeter estimates.

The theoretical limitations of the IDM centre on the ability of the approach to accurately model the tail of the CDF and the validity of using data which may not be independent. As there is no absolute measure of the extreme values for comparative purposes, the ability of the IDM to model the tail cannot be absolutely determined. However, the approach gives results which are in excellent agreement with buoy data which has greater temporal sampling density and hence more extreme data points in the analysis. Although comparison with buoy data does not provide an authoritative answer, it provides some level of confidence in the altimeter results. When applied to altimeter data, the issue of independence of the data is not considered a major issue. Altimeter passes through the $1^{o}\times1^{o}$ bins occurred only once every few days, even when there were multiple satellites operational in the latter years of the data set. As a result, the data will generally satisfy the requirement for independence.

The present analysis clearly shows that with approximately 23 years of data, the satellite altimeter can provide high quality estimates of extreme wind speed and wave height conditions on a global basis. The volume of the data also means that small scale features can be resolved, opening up the possibility for regional studies of extreme wave climate. With the duration and sampling density of the present data set, either the IDM or POT approached can be used to obtain extreme value wave height estimates comparable with buoys. As the duration of the data set continues to grow, it is likely that the POT approach, with its sounder theoretical underpinning, will replace the IDM approach for extreme wave height estimates. The greater spatial variability of wind speed, however, means that the altimeter data is still too sparse to produce POT extreme value results comparable to buoy estimates. For wind speed, only the IDM approach produces acceptable results, a situation which is likely to remain unless future satellite missions can provide greater spatial and temporal sampling density.

Acknowledgements: This work was financially supported by the Australian Research Council (ARC) through Linkage Grant LP0882422. The support of the ARC and our industry partner RPS MetOcean is gratefully acknowledg

8. References

Alves, J.H.G.M. and I.R. Young, 2003, On estimating extreme wave heights using combined Geosat, Topex/Poseidon and ERS-1 altimeter data, *Applied Ocean Research*, **25**, 167-186, doi:10.1016/j.apor.2004.01.002.

Anderson, C.W., D.J.T. Carter and P.D. Cotton, 2001, Wave Climate variability and Impact on Offshore Structure Design Extremes, *Report for Shell International and the Organization of Oil & Gas Producers*, 88pp.

Anderson, T.W. and D.A. Darling, 1952, Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes, *Ann. Math. Statist.*,**23**,193-212, doi:10.1214/aoms/1177729437.

Caires, S. and A. Sterl, 2005, 100-year return value estimates for wind speed and significant wave height from the ERA-40 data, *J. Clim*., **18**: 1032-1048, doi:10.1175/JCLI-3312.1.

Carter, D.J.T., 1993, Estimating extreme wave heights in the NE Atlantic from GEOSAT data, Department of Energy – Offshore Technology Report, OTH 1993, **396**, 28pp.

Castillo, E., 1988, *Extreme value theory in engineering*, Academic Press Inc., Boston, MA, 389pp.

Chen, G., S. –W. Bi and R. Ezraty, 2004, Global structure of extreme wind and wave climate derived from TOPEX altimeter data, *Int. J. Remote Sens*., **25**, 1005-1018, doi:10.1080/01431160310001598980.

Challenor, P.G., S. Foale and D.J. Webb, 1990, Seasonal changes in the global wave climate measured by the GEOSAT altimeter, *Int. J. Remote Sens.*, **11**, 2205-2213, doi:10.1080/01431169008955170.

Challenor, P.G., W. Wimmer and I. Ashton, 2005, Climate change and extreme wave heights in the north Atlantic, *Proc. 2004 Envisat & ERS Symposium,* ESA SP-572.

Cooper, C.K. and G.Z. Forristall, 1997, The use of satellite data to estimate extreme wave climate, *J. Atmos. Oceangr. Tech.,* **14**, 254-266, doi: 10.1175/1520-0426(1997)014<0254:TUOSAD>2.0.CO;2.

Coles, S., 2001, *An introduction to statistical modelling of extreme values*, Springer-Verlag, London, 208pp.

Embrechts, P., C. Klüppelberg, and T. Mikosch, 1997, *Modelling Extremal Events for Insurance and Finance*, New York: Springer.

Evans, M., N. Hastings and B. Peacock, 2000, *Statistical distributions*. 3rd ed., John Wiley & Sons. 248pp.

Evans, D., C.L. Conrad and F.M. Paul, 2003, *Handbook of Automated Data Quality Control Checks and Procedures of the National Data Buoy Center*. NOAA National Data Buoy Center Tech. Document 03-02, 44pp.

Ferreira, J.A. and C.G. Soares, 1998, An application of the peaks over threshold method to predict extremes of significant wave height, *J. Offshore Mech. & Arctic Eng.,* **120**, 165-176, doi:10.1115/1.2829537 **.**

Goda, Y., 1988, On the methodology of selecting design wave height, *Proc. 21st Int. Conf. on Coastal Eng.*, Malaga, Spain, 899-913.

Goda, Y., 1992, Uncertainty in design parameter from the viewpoint of statistical variability, *J. Offshore Mech. and Arctic Eng.,* **114**, 76-82, doi:10.1115/1.2919962 **.**

Goda, Y., 2000, *Random seas and design of maritime structures*, World Scientific, 433pp.

Goda, Y. and K. Kobune, 1990, Distribution function fitting for storm wave data, *Proc. 22nd Int. Conf. on Coastal Eng.*, Delft, The Netherlands, 18-31.

Gumbel, E.J., 1958, *Statistics of extremes*, Columbia Univ. Press, New York, 375pp.

Hogben, N., 1988, Experience from compilation of global wave statistics, *Ocean Eng.*, **15**, 1-31.

Holthuijsen, L.H., 2007, *Waves in oceanic and coastal waters*, Cambridge Univ. Press, 387pp.

Kotlz, S., and S. Nadarajah, 2000, *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press.

Mathiesen, M. et al., 1994, Recommended practice for extreme wave height, *J. Hydraulic Res,* 32, 803-814.

Ochi, M.K., 1992, New approach in estimating the severest sea state from statistical data, *Proc. Coastal Eng. Conf.*, ASCE, New York, 512-523.

Panchang, V., L. Zhao and Z. Demirbilek, 1999, Estimation of extreme wave heights using GEOSAT measurements, *Ocean Eng.*, **26**, 205-225, doi:10.1016/j.apor.2004.01.002.

Soares, C.G. and A.C Henriques, 1996, Statistical uncertainty in long-term distributions of significant wave height, *J. Offshore Mech. and Arctic Eng.*, **118**, 284-291, doi:10.1115/1.2833917.

Stephens, M.A., 1970, Use of the Kolmogorov-Smirnov, Cramer-Von Mises and related statistics without extensive tables, *J. Roy. Stats. Soc.*, Ser B., **32**, 115-122.

Stephens, M.A., 1977, Goodness of fit for the extreme value distribution, *Biometrika*, **64**, 583-588, doi:10.1093/biomet/64.3.583.

Stephens, M.A., 1986, *Tests based on EDF statistics*, Chap 4 in: D'Agostino, R.B. and M.A. Stephens (eds.), 1986, "Goodness-of-fit techniques", Marcel Decker, New York, 533pp.

Teng, C.C., 1998, Long-term and extreme waves in the Gulf of Mexico, *Proc. Of Conf. on Ocean Wave Kinematics and Loads on Structures,* Houston, USA, 342-349.

Tournadre, J. and R. Ezraty, 1990, Local climatology of wind and sea state by means of satellite radar altimeter measurements, *J. Geophys. Res.*,**95**, 18,255-18,268, doi: 10.1029/JC095iC10p18255.

Tucker, M.J., 1991, *Waves in ocean engineering*, Ellis Horwood, 431pp.

Van Gelder, P.H.A.J.M. and J.K. Vrijling, 1999, On the distribution function of the maximum wave height in front of reflecting structures, *Proc. of Coastal Structures*, Santander, June 7-10.

Wimmer, W., P. Challenor and C. Retzler, 2006, Extreme wave heights in the North Atlantic from altimeter data, *Renewable Energy*, **31**, 241-248, doi:10.1016/j.renene.2005.08.019.

Young, I.R., 1994, Global Ocean Wave Statistics Obtained from Satellite Observations, *App. Ocean Res.*, **16**, 235-248, doi:10.1016/0141-1187(94)90023-X.

Young, I.R., 1999, Seasonal variability of the global ocean wind and wave climate. *Int. J. Climatology*, **19**, 931-950, doi:10.1002/(SICI)1097-0088(199907)19:9<931::AID-JOC412>3.0.CO;2-O.

Young, I.R. and G.J. Holland, 1996, *Atlas of the Oceans: Wind and Wave Climate*, Pergamon Press, ISBN 0-08-042519-4, 241pp.

Zieger, S., J. Vinoth and I.R. Young, 2009, Joint calibration of multi-platform altimeter measurements of wind speed and wave height over the past 20 years. *J. Atmos. & Ocean. Tech*., **26**, 2549-2564, doi: 10.1175/2009JTECHA1303.1.

Geographic Region	Station		$1^o \times 1^o$ region		$2^o \times 2^o$ region		50 km radius		Resampled	
US East Coast	41002		3.35%		1.67%		5.81%		-4.15%	
	44004		3.12%		1.12%		2.46%		-3.98%	
Gulf of Mexico	42001		-3.86%		2.91%		-6.12%		0.49%	
	42002		-0.40%		-0.96%		-0.23%		-1.34%	
US West Coast	46001		-2.47%		-2.02%		-0.69%		0.63%	
	46002		1.81%		2.57%		0.37%		4.24%	
	46003		-0.45%		0.35%		-1.06%		-4.98%	
	46005		1.08%		1.35%		1.40%		-2.55%	
	46006		-1.78%		-1.69%		-1.42%		-1.78%	
Hawaii	51001		-5.56%		-4.80%		-5.41%		-1.52%	
	$r_1 = 1/n\sum\lvert\Delta r\rvert$	$r_2 = 1/n\sum\Delta r$	2.11%	-0.03%	2.47%	0.58%	2.01%	0.00%	3.07%	-0.99%

Table 1: Values of the relative error, Δr between buoy and altimeter for different averaging areas for the altimeter. The results are shown for H_s^{100} *(IDM)* using the IDM approach with a FT-1 distribution. Also shown are the mean absolute error, r_1 and the mean error r_2. Each column shows a different averaging area with the buoy locations shown for each row.

<table>
<tr><th>Geographic Region</th><th colspan="2">Station</th><th colspan="2">$1^{o} \times 1^{o}$ region</th><th colspan="2">$2^{o} \times 2^{o}$ region</th><th colspan="2">50 km radius</th><th colspan="2">Resampled</th></tr>
<tr><td rowspan="2">US East Coast</td><td colspan="2">41002</td><td colspan="2">-18.37%</td><td colspan="2">-13.91%</td><td colspan="2">-20.21%</td><td colspan="2">-11.86%</td></tr>
<tr><td colspan="2">44004</td><td colspan="2">-17.62%</td><td colspan="2">-13.80%</td><td colspan="2">-39.16%</td><td colspan="2">-2.61%</td></tr>
<tr><td rowspan="2">Gulf of Mexico</td><td colspan="2">42001</td><td colspan="2">-27.46%</td><td colspan="2">3.56%</td><td colspan="2">-41.61%</td><td colspan="2">3.25%</td></tr>
<tr><td colspan="2">42002</td><td colspan="2">-22.46%</td><td colspan="2">-23.30%</td><td colspan="2">-27.73%</td><td colspan="2">-7.52%</td></tr>
<tr><td rowspan="5">US West Coast</td><td colspan="2">46001</td><td colspan="2">-16.01%</td><td colspan="2">-14.12%</td><td colspan="2">-14.12%</td><td colspan="2">9.99%</td></tr>
<tr><td colspan="2">46002</td><td colspan="2">-25.11%</td><td colspan="2">-9.64%</td><td colspan="2">-13.56%</td><td colspan="2">5.59%</td></tr>
<tr><td colspan="2">46003</td><td colspan="2">-17.38%</td><td colspan="2">-15.58%</td><td colspan="2">-27.66%</td><td colspan="2">13.32%</td></tr>
<tr><td colspan="2">46005</td><td colspan="2">-7.49%</td><td colspan="2">0.72%</td><td colspan="2">-7.23%</td><td colspan="2">7.46%</td></tr>
<tr><td colspan="2">46006</td><td colspan="2">-24.88%</td><td colspan="2">-14.63%</td><td colspan="2">-30.07%</td><td colspan="2">4.87%</td></tr>
<tr><td>Hawaii</td><td colspan="2">51001</td><td colspan="2">-25.16%</td><td colspan="2">-16.54%</td><td colspan="2">-24.24%</td><td colspan="2">10.93%</td></tr>
<tr><td></td><td>$r_1 = 1/n \sum |\Delta r|$</td><td>$r_2 = 1/n \sum \Delta r$</td><td>20.20%</td><td>-20.20%</td><td>12.58%</td><td>-11.73%</td><td>24.56%</td><td>-24.56%</td><td>7.74%</td><td>3.34%</td></tr>
</table>

Table 2: Values of the relative error, Δr between buoy and altimeter for different averaging areas for the altimeter. The results are shown for U_{10}^{100} *(POT)* using the POT approach with a W3P distribution. Also shown are the mean absolute error, r_1 and the mean error r_2. Each column shows a different averaging area with the buoy locations shown for each row.

	IDM						POT (90%)					
	FT-1 [H_s^{100} *(IDM)*] (m)			FT-1G [H_s^{100} *(IDM)*] (m)			GPD [H_s^{100} *(POT)*] (m)			W3P [H_s^{100} *(POT)*] (m)		
Station	Buoy	Altimeter $1^{\circ} \times 1^{\circ}$	Altimeter $2^{\circ} \times 2^{\circ}$	Buoy	Altimeter $1^{\circ} \times 1^{\circ}$	Altimeter $2^{\circ} \times 2^{\circ}$	Buoy	Altimeter $1^{\circ} \times 1^{\circ}$	Altimeter $2^{\circ} \times 2^{\circ}$	Buoy	Altimeter $1^{\circ} \times 1^{\circ}$	Altimeter $2^{\circ} \times 2^{\circ}$
41002	11.12 ± 0.05^{1}	11.50 ± 0.35^{1}	11.31 ± 0.24^{1}	9.58 ± 0.05^{1}	9.92 ± 0.35^{1}	9.87 ± 0.23^{1}	13.60 ± 0.26	11.05 ± 1.90	11.00 ± 1.14	12.35 ± 0.29^{1}	10.08 ± 1.87^{4}	10.64 ± 1.22^{4}
44004	13.51 ± 0.07^{1}	13.93 ± 0.39^{2}	13.66 ± 0.21^{1}	11.77 ± 0.07^{1}	12.35 ± 0.38^{0}	12.15 ± 0.21^{1}	12.86 ± 0.29	9.35 ± 2.20	10.03 ± 1.04	13.55 ± 0.36^{1}	11.17 ± 2.01^{2}	11.68 ± 1.29^{3}
42001	7.98 ± 0.04^{1}	8.06 ± 0.30^{2}	8.63 ± 0.21^{3}	7.03 ± 0.04^{1}	7.36 ± 0.29^{1}	7.35 ± 0.21^{0}	11.23 ± 0.16	6.71 ± 1.90	13.05 ± 1.04	8.84 ± 0.20^{1}	6.41 ± 1.53^{4}	9.15 ± 2.22^{4}
42002	8.07 ± 0.04^{1}	8.03 ± 0.22^{3}	7.99 ± 0.25^{1}	7.42 ± 0.04^{1}	7.55 ± 0.21^{0}	7.47 ± 0.24^{1}	9.09 ± 0.17	5.81 ± 1.46	5.88 ± 1.19	8.44 ± 0.21^{1}	6.54 ± 1.14^{2}	6.47 ± 0.85^{3}
46001	16.14 ± 0.07^{1}	15.74 ± 0.37^{3}	15.81 ± 0.28^{0}	15.54 ± 0.07^{1}	15.21 ± 0.37^{0}	15.28 ± 0.27^{0}	12.52 ± 0.29	11.43 ± 1.70	11.55 ± 1.17	14.13 ± 0.40^{1}	11.86 ± 1.94^{2}	12.13 ± 1.44^{3}
46002	15.10 ± 0.07^{1}	15.38 ± 0.50^{3}	15.49 ± 0.27^{3}	13.86 ± 0.07^{1}	14.43 ± 0.49^{0}	14.47 ± 0.26^{0}	14.10 ± 0.31	10.30 ± 2.56	12.55 ± 1.32	14.76 ± 0.39^{1}	11.05 ± 2.59^{2}	13.33 ± 1.36^{3}
46003	17.31 ± 0.10^{1}	17.23 ± 0.44^{3}	17.37 ± 0.31^{3}	16.63 ± 0.10^{1}	16.98 ± 0.43^{2}	17.04 ± 0.30^{0}	16.24 ± 0.39	11.85 ± 2.25	12.23 ± 1.36	15.69 ± 0.56^{1}	12.96 ± 2.59^{3}	13.25 ± 1.61^{3}
46005	16.28 ± 0.08^{1}	16.46 ± 0.42^{3}	16.50 ± 0.32^{3}	15.16 ± 0.07^{1}	15.49 ± 0.41^{0}	15.53 ± 0.31^{0}	14.04 ± 0.32	12.51 ± 2.62	13.40 ± 1.76	14.91 ± 0.42^{1}	13.79 ± 2.15^{3}	15.02 ± 1.65^{3}
46006	16.95 ± 0.09^{1}	16.65 ± 0.53^{2}	16.66 ± 0.33^{0}	15.65 ± 0.09^{1}	15.73 ± 0.51^{0}	15.88 ± 0.32^{0}	15.31 ± 0.34	12.19 ± 2.13	12.31 ± 1.34	15.67 ± 0.46^{1}	11.77 ± 2.73^{4}	13.37 ± 1.71^{3}
51001	10.64 ± 0.05^{1}	10.05 ± 0.26^{2}	10.13 ± 0.19^{2}	9.79 ± 0.05^{1}	9.47 ± 0.26^{1}	9.62 ± 0.19^{0}	10.47 ± 0.22	7.89 ± 1.48	9.12 ± 0.96	10.99 ± 0.24^{1}	8.23 ± 1.36^{3}	9.17 ± 0.98^{3}
r_1		2.11%	2.47%		2.92%	2.57%		25.83%	17.31%		20.20%	12.58%

Table 3: Buoy and altimeter values of H_s^{100} with 95% confidence intervals at different buoy locations for different analysis methods and CDFs. Both $1^{\circ} \times 1^{\circ}$ and $2^{\circ} \times 2^{\circ}$ averaging areas are shown. The integer superscripts represent the number of goodness-of-fit tests satisfied. The mean absolute error, r_1 between buoy and altimeter is also shown, as in Table 1.

	IDM						POT (90%)					
	FT-1 [U_{10}^{100} *(IDM)*] (m/s)			FT-1G [U_{10}^{100} *(IDM)*] (m/s)			GPD [U_{10}^{100} *(POT)*] (m/s)			W3P [U_{10}^{100} *(POT)*] (m/s)		
Station	Buoy	Altimeter $1^{o} \times 1^{o}$	Altimeter $2^{o} \times 2^{o}$	Buoy	Altimeter $1^{o} \times 1^{o}$	Altimeter $2^{o} \times 2^{o}$	Buoy	Altimeter $1^{o} \times 1^{o}$	Altimeter $2^{o} \times 2^{o}$	Buoy	Altimeter $1^{o} \times 1^{o}$	Altimeter $2^{o} \times 2^{o}$
41002	36.82 ± 0.18^{1}	43.55 ± 1.32^{1}	$42.350.83^{2}$	40.13 ± 0.18^{1}	42.99 ± 1.29^{1}	42.90 ± 0.81^{1}	26.00 ± 0.51	41.66 ± 6.63	42.80 ± 3.34	28.21 ± 0.94^{1}	35.56 ± 6.83^{3}	36.54 ± 4.28^{4}
44004	40.92 ± 0.20^{1}	47.37 ± 1.29^{3}	47.30 ± 0.74^{2}	44.10 ± 0.19^{1}	48.92 ± 1.26^{2}	48.82 ± 0.72^{1}	26.42 ± 0.44	36.87 ± 3.65	41.06 ± 2.75	30.50 ± 1.02^{1}	32.04 ± 6.65^{3}	34.71 ± 4.24^{1}
42001	33.73 ± 0.15^{1}	31.58 ± 1.07^{4}	33.17 ± 0.63^{2}	36.43 ± 0.15^{1}	34.44 ± 1.04^{1}	34.83 ± 0.61^{1}	30.65 ± 0.38	17.00 ± 4.96	32.76 ± 2.13	27.06 ± 0.79^{1}	19.84 ± 2.95^{1}	27.95 ± 3.83^{4}
42002	34.43 ± 0.15^{1}	32.89 ± 0.84^{4}	33.72 ± 0.82^{1}	38.14 ± 0.15^{1}	36.14 ± 0.82^{1}	36.25 ± 0.80^{1}	24.11 ± 0.41	27.18 ± 2.94	33.81 ± 2.21	26.48 ± 0.80^{1}	24.19 ± 4.34^{4}	28.18 ± 3.23^{4}
46001	40.01 ± 0.19^{1}	43.38 ± 1.01^{3}	43.89 ± 0.73^{1}	44.47 ± 0.18^{1}	47.50 ± 0.98^{1}	47.80 ± 0.71^{1}	22.60 ± 0.51	24.08 ± 2.77	25.65 ± 2.02	28.49 ± 096^{1}	25.99 ± 2.69^{1}	28.24 ± 3.78^{3}
46002	36.09 ± 0.17^{1}	38.44 ± 1.19^{2}	39.34 ± 0.64^{2}	40.65 ± 0.17^{1}	43.05 ± 1.15^{0}	43.37 ± 0.62^{1}	25.23 ± 0.50	21.82 ± 3.83	24.22 ± 2.17	27.99 ± 0.90^{1}	22.70 ± 6.10^{3}	27.47 ± 3.28^{3}
46003	40.71 ± 0.24^{1}	45.44 ± 1.11^{1}	45.50 ± 0.76^{1}	45.86 ± 0.24^{1}	50.75 ± 1.08^{1}	50.66 ± 0.74^{1}	23.96 ± 0.60	23.61 ± 3.40	28.87 ± 1.80	27.63 ± 1.28^{1}	26.75 ± 5.68^{3}	28.99 ± 3.91^{4}
46005	37.61 ± 0.18^{1}	42.03 ± 1.02^{1}	42.47 ± 0.76^{1}	41.52 ± 0.17^{1}	45.57 ± 0.99^{1}	45.98 ± 0.73^{1}	24.56 ± 0.51	26.82 ± 3.00	29.96 ± 2.03	29.21 ± 0.92^{1}	27.30 ± 5.25^{4}	29.02 ± 3.89^{4}
46006	40.13 ± 0.20^{1}	43.67 ± 1.32^{4}	43.04 ± 0.79^{1}	43.95 ± 0.20^{1}	46.21 ± 1.29^{1}	46.13 ± 0.77^{1}	26.48 ± 0.46	35.82 ± 4.92	33.65 ± 2.36	28.63 ± 1.07^{1}	31.90 ± 6.82^{3}	30.92 ± 4.08^{4}
51001	30.41 ± 0.13^{1}	32.56 ± 0.85^{1}	33.46 ± 0.60^{1}	37.68 ± 0.13^{1}	37.83 ± 0.82^{1}	38.33 ± 0.59^{1}	19.49 ± 0.30	35.23 ± 2.93	49.10 ± 1.93	21.38 ± 0.71^{1}	26.18 ± 8.75^{1}	30.62 ± 6.40^{2}
r_1		9.91%	9.50%		6.74%	6.91%		30.39%	40.62%		13.77%	11.26%

Table 4: Buoy and altimeter values of U_{10}^{100} with 95% confidence intervals at different buoy locations for different analysis methods and CDFs. Both $1^{o} \times 1^{o}$ and $2^{o} \times 2^{o}$ averaging areas are shown. The integer superscripts represent the number of goodness-of-fit tests satisfied. The mean absolute error, r_1 between buoy and altimeter is also shown, as in Table 2.

List of Figures

Figure 1: Periods of operation of each of the satellite altimeter missions which make up the Zieger et al (2009) data base.

Figure 2: Locations of NODC buoys used for validation of the altimeter extreme value analysis.

Figure 3: 100-year return period estimates at various buoy locations for (a) H_s^{100} with IDM and FT-1 distribution, (b) H_s^{100} with POT and W3P distribution, (c) U_{10}^{100} with IDM and FT-1 distribution, and (d) U_{10}^{100} with POT and W3P distribution. The $1^o \times 1^o$ altimeter values are shown by the open circles and the $2^o \times 2^o$ altimeter values by the solid dots (only shown for POT). The 95% confidence limits for the $1^o \times 1^o$ IDM and the $2^o \times 2^o$ POT altimeter values are shown by the grey shading. The buoy values are shown with a cross and the buoy $\pm 95\,\%$ confidence interval by a vertical line (not visible for the IDM as it is very small).

Figure 4: Altimeter and buoy H_s^{100} *(POT)* at buoy location 46002 for different threshold values. Panel (a) shows the W3P distribution and panel (b) the GPD distribution.

Figure 5: Percentile-percentile (or Q-Q) plot between altimeter and buoy derived values of H_s (left panel) and U_{10} (right panel). The most extreme value shown in the figures is the 99th percentile.

Figure 6: Contour plots of the global distribution of H_s^{100} (m) on a $1^o \times 1^o$ grid. (a) IDM method and FT-1 distribution, (b) POT method and W3P distribution with 90% threshold, (c) POT method and GPD distribution with 90% threshold.

Figure 7: Contour plots of the global distribution of H_s^{100} (m) on a $2^o \times 2^o$ grid. (a) IDM method and FT-1 distribution, (b) POT method and W3P distribution with 90% threshold, (c) POT method and GPD distribution with 90% threshold.

Figure 8: Contour plots of the global distribution of U_{10}^{100} (ms^{-1}) on a $1^o \times 1^o$ grid. (a) IDM method and FT-1 distribution, (b) POT method and W3P distribution with 90% threshold , (c) POT method and GPD distribution with 90% threshold.

Figure 9: Contour plots of the global distribution of U_{10}^{100} (ms^{-1}) on a $2^o \times 2^o$ grid. (a) IDM method and FT-1 distribution, (b) POT method and W3P distribution with 90% threshold , (c) POT method and GPD distribution with 90% threshold.

Figure 10: Contour plots of H_s^{100} *(IDM)* (panel a) and U_{10}^{100} *(IDM)* (panel b), as in Figures 6 and 7, respectively. Points which satisfy the Cramér-von Mises goodness-of-fit test censored at 20% for IDM FT-1 are shown by a black dot.

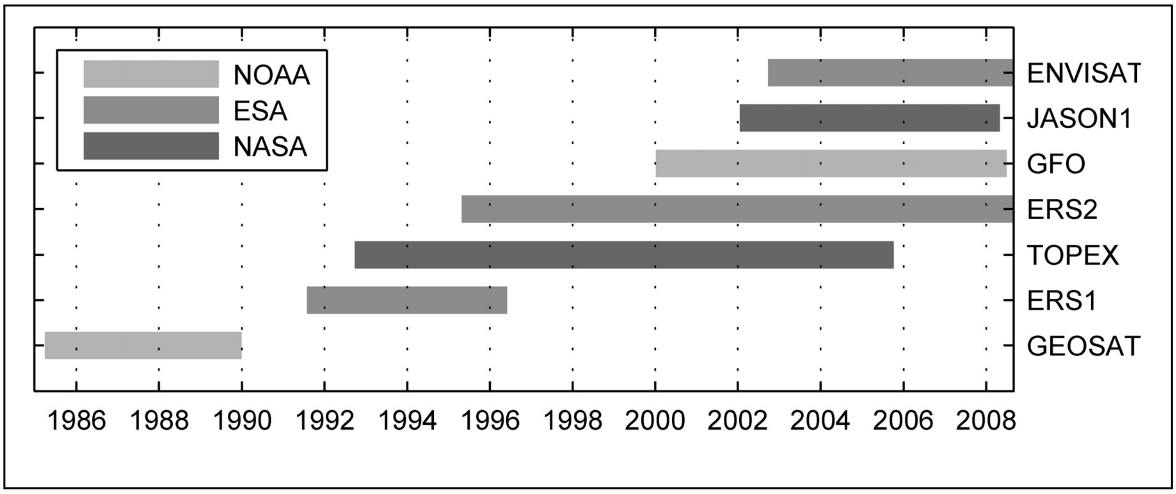

Figure 1: Periods of operation of each of the satellite altimeter missions which make up the Zieger et al (2009) data base.

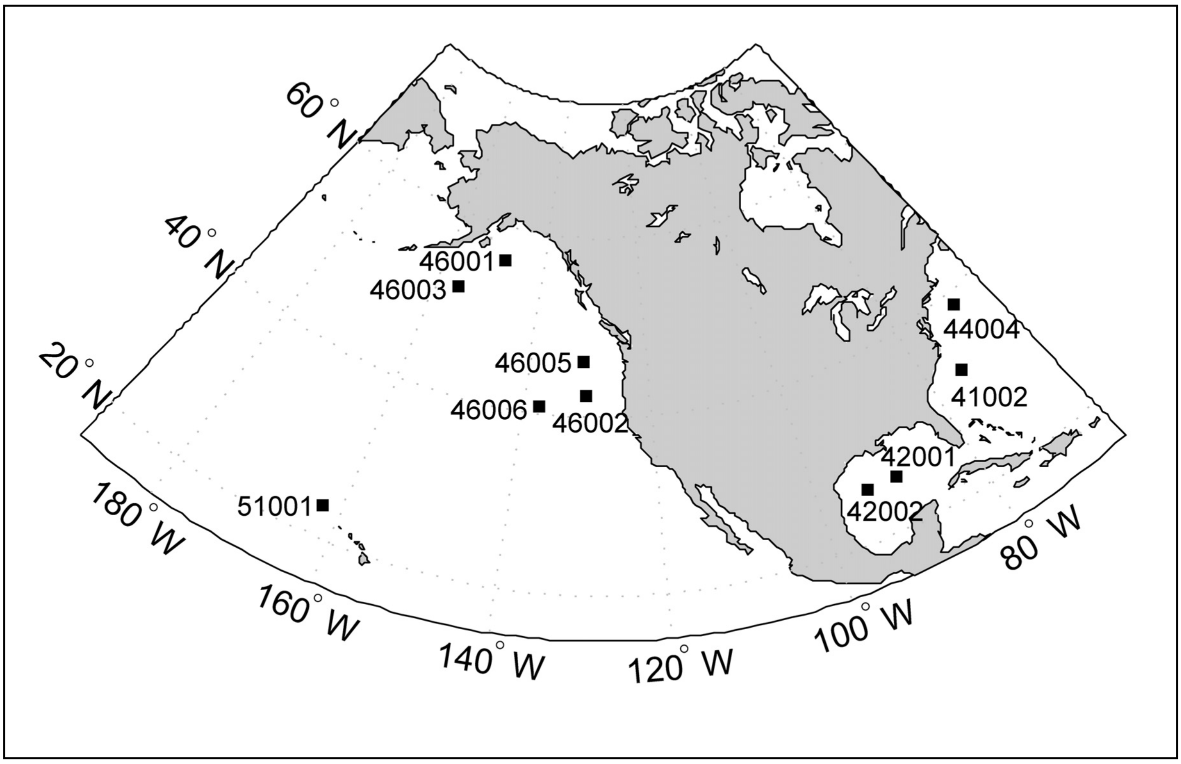

Figure 2: Locations of NODC buoys used for validation of the altimeter extreme value analysis.

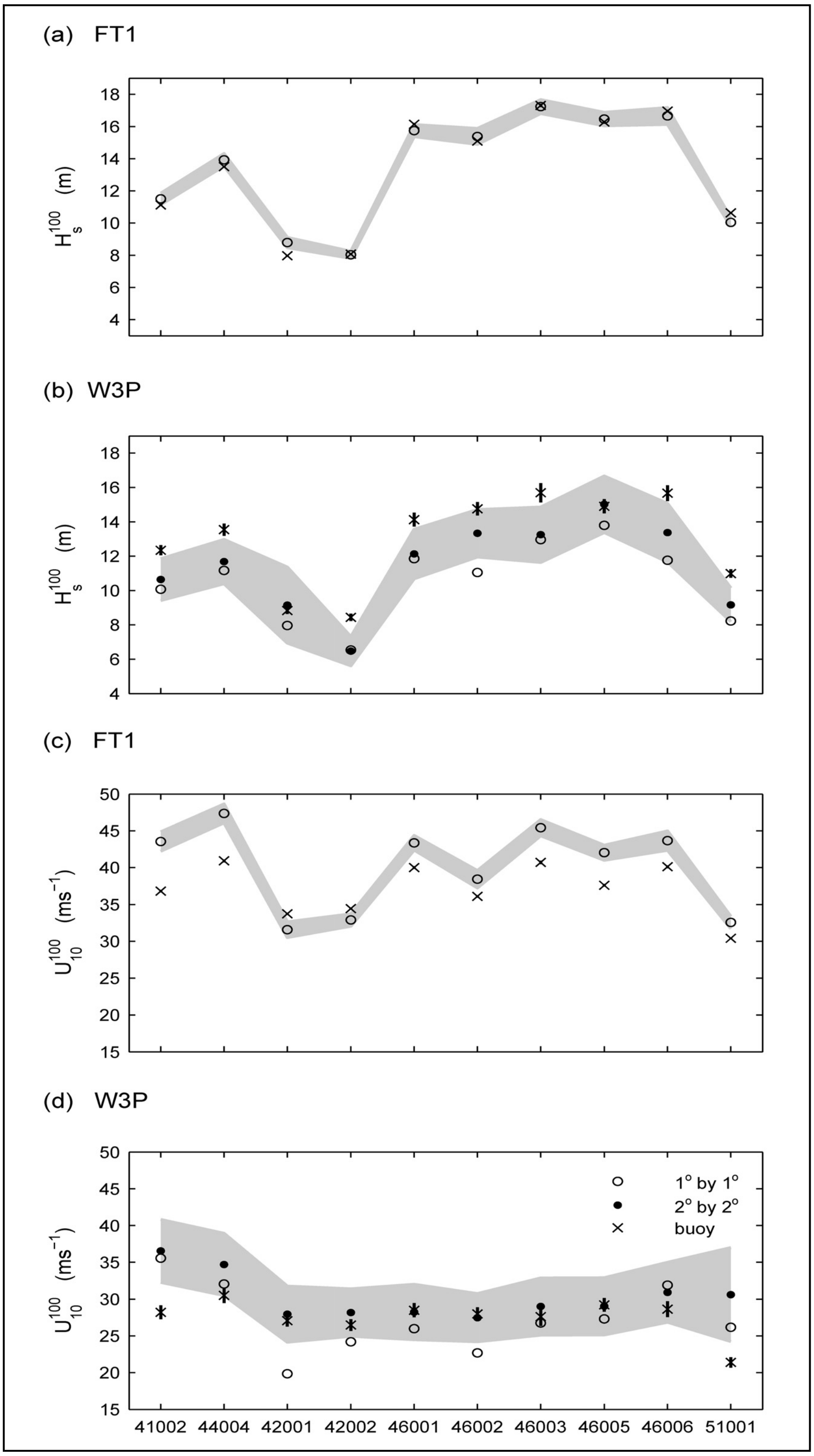

Figure 3: 100-year return period estimates at various buoy locations for (a) H_s^{100} with IDM and FT-1 distribution, (b) H_s^{100} with POT and W3P distribution, (c) U_{10}^{100} with IDM and FT-1 distribution, and (d) U_{10}^{100} with POT and W3P distribution. The $1^o \times 1^o$ altimeter values are shown by the open circles and the $2^o \times 2^o$ altimeter values by the solid dots (only shown for POT). The 95% confidence limits for the $1^o \times 1^o$ IDM and the $2^o \times 2^o$ POT altimeter values are shown by the grey shading. The buoy values are shown with a cross and the buoy $\pm 95\,\%$ confidence interval by a vertical line (not visible for the IDM as it is very small).

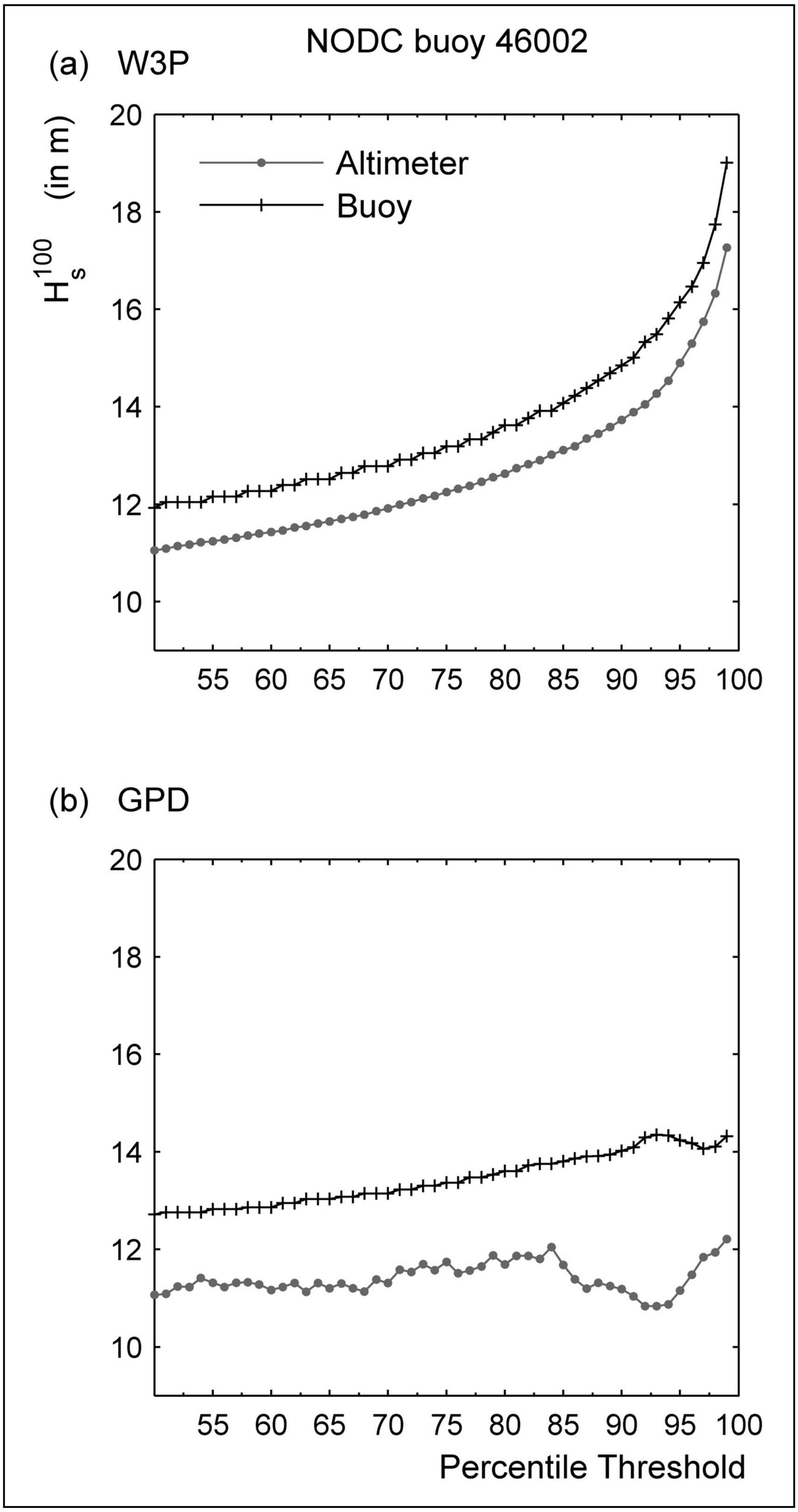

Figure 4: Altimeter and buoy H_s^{100} *(POT)* at buoy location 46002 for different threshold values. Panel (a) shows the W3P distribution and panel (b) the GPD distribution.

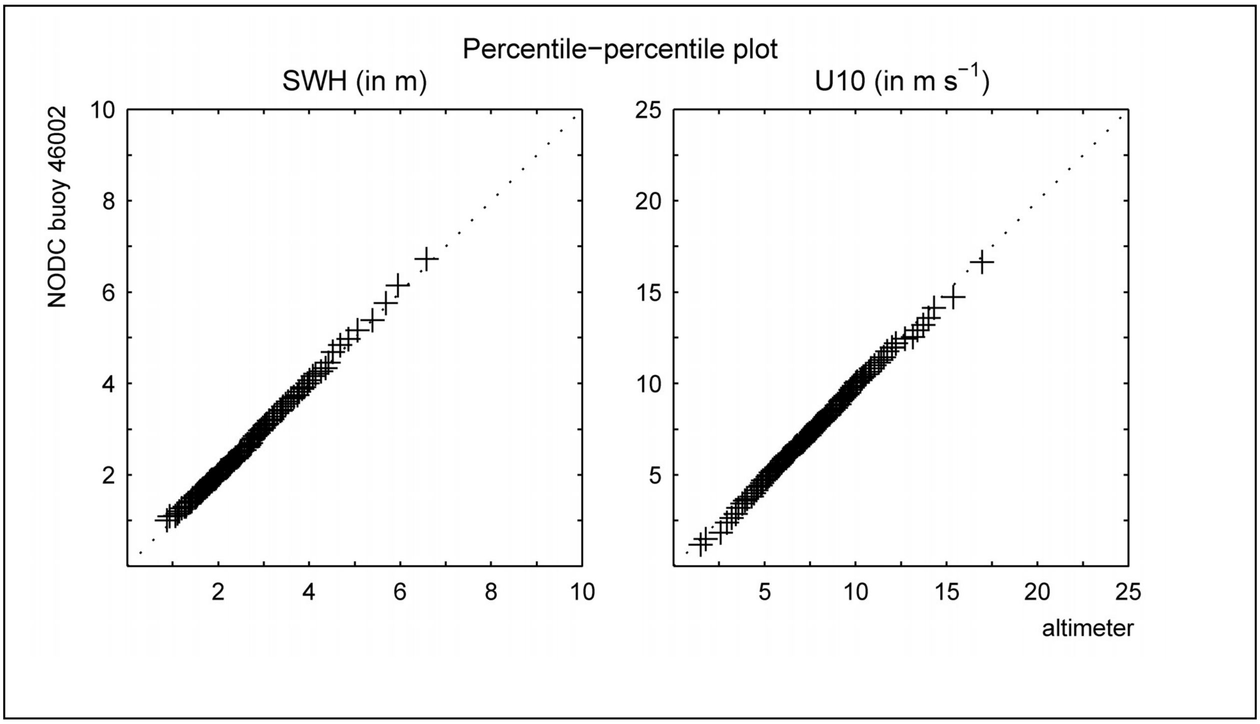

Figure 5: Percentile-percentile (or Q-Q) plot between altimeter and buoy derived values of H_s (left panel) and U_{10} (right panel). The most extreme value shown in the figures is the 99th percentile.

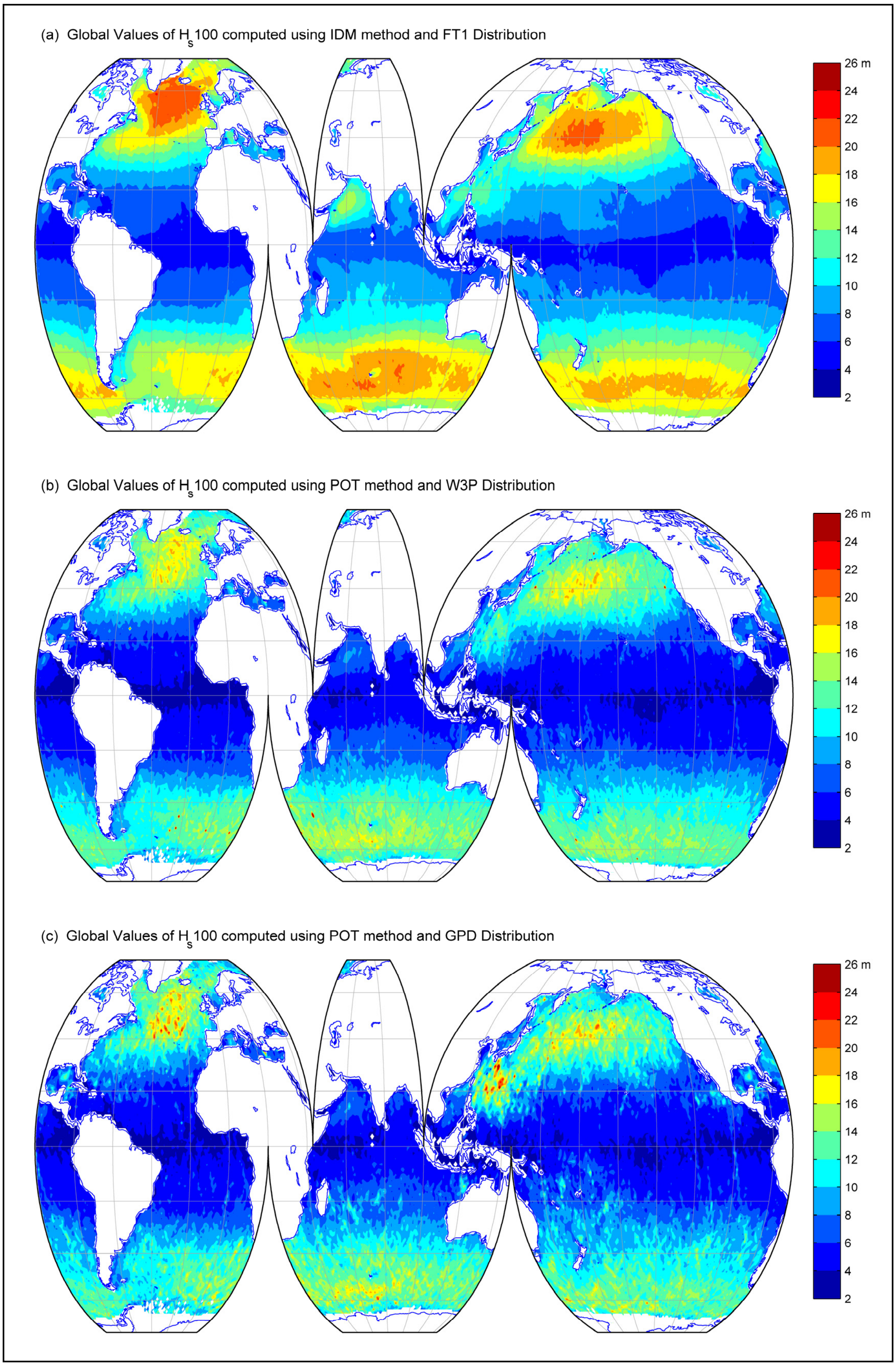

Figure 6: Contour plots of the global distribution of H_s^{100} (m) on a $1^{o} \times 1^{o}$ grid. (a) IDM method and FT-1 distribution, (b) POT method and W3P distribution with 90% threshold, (c) POT method and GPD distribution with 90% threshold.

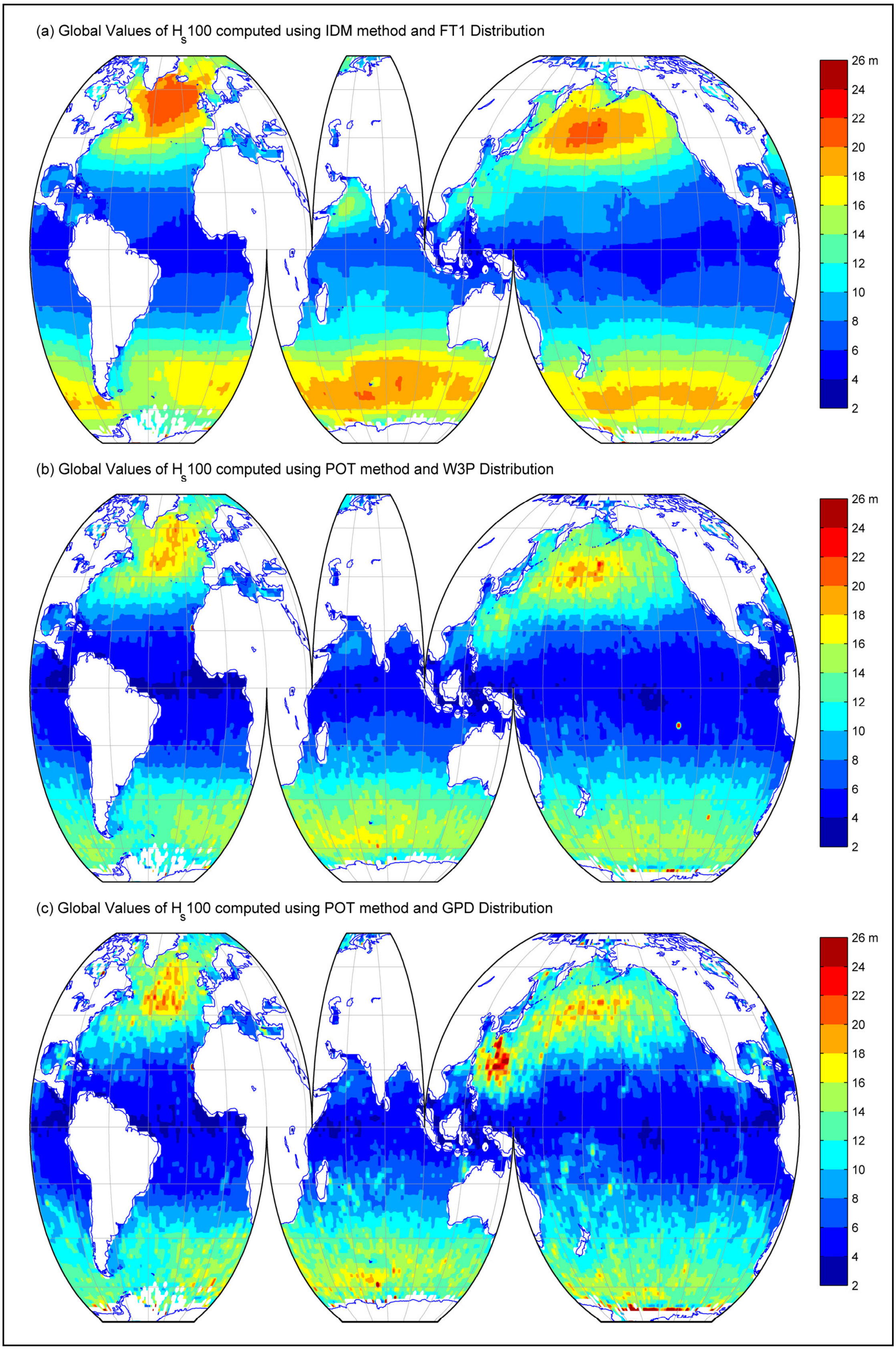

Figure 7: Contour plots of the global distribution of H_s^{100} (m) on a $2^o \times 2^o$ grid. (a) IDM method and FT-1 distribution, (b) POT method and W3P distribution with 90% threshold, (c) POT method and GPD distribution with 90% threshold.

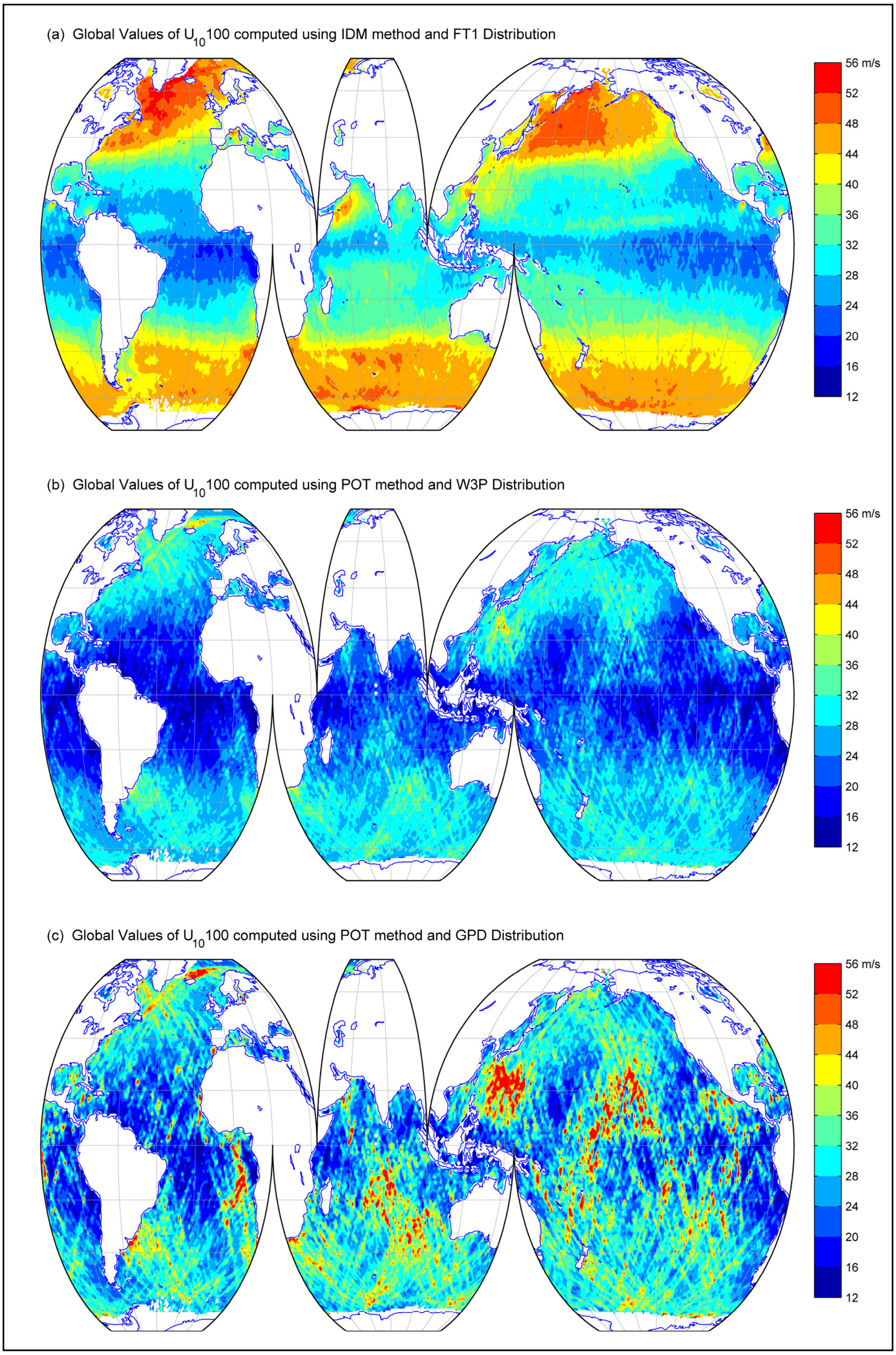

Figure 8: Contour plots of the global distribution of U_{10}^{100} (ms^{-1}) on a $1^{\circ} \times 1^{\circ}$ grid. (a) IDM method and FT-1 distribution, (b) POT method and W3P distribution with 90% threshold , (c) POT method and GPD distribution with 90% threshold.

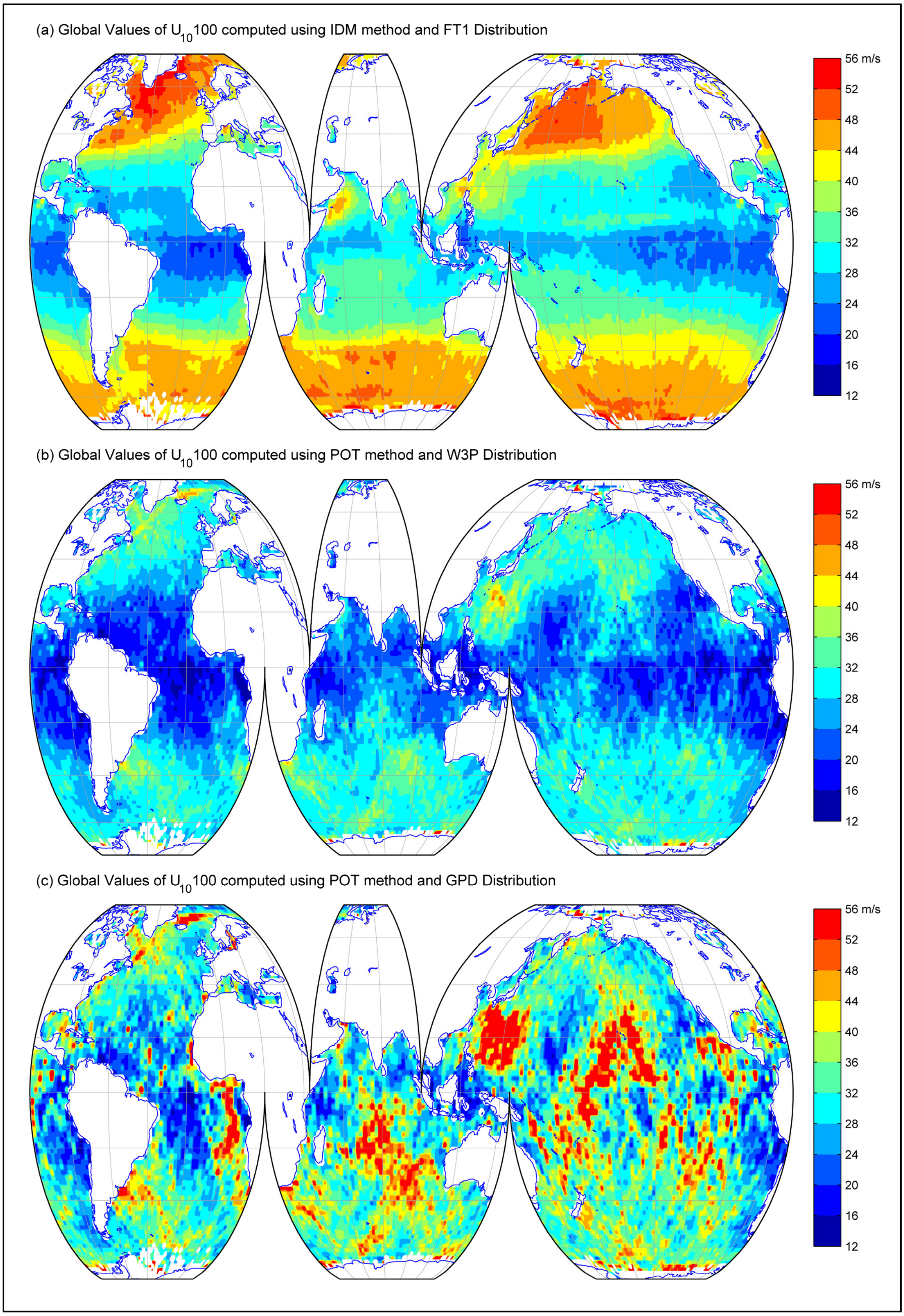

Figure 9: Contour plots of the global distribution of U_{10}^{100} (ms^{-1}) on a $2^{\circ} \times 2^{\circ}$ grid. (a) IDM method and FT-1 distribution, (b) POT method and W3P distribution with 90% threshold , (c) POT method and GPD distribution with 90% threshold.

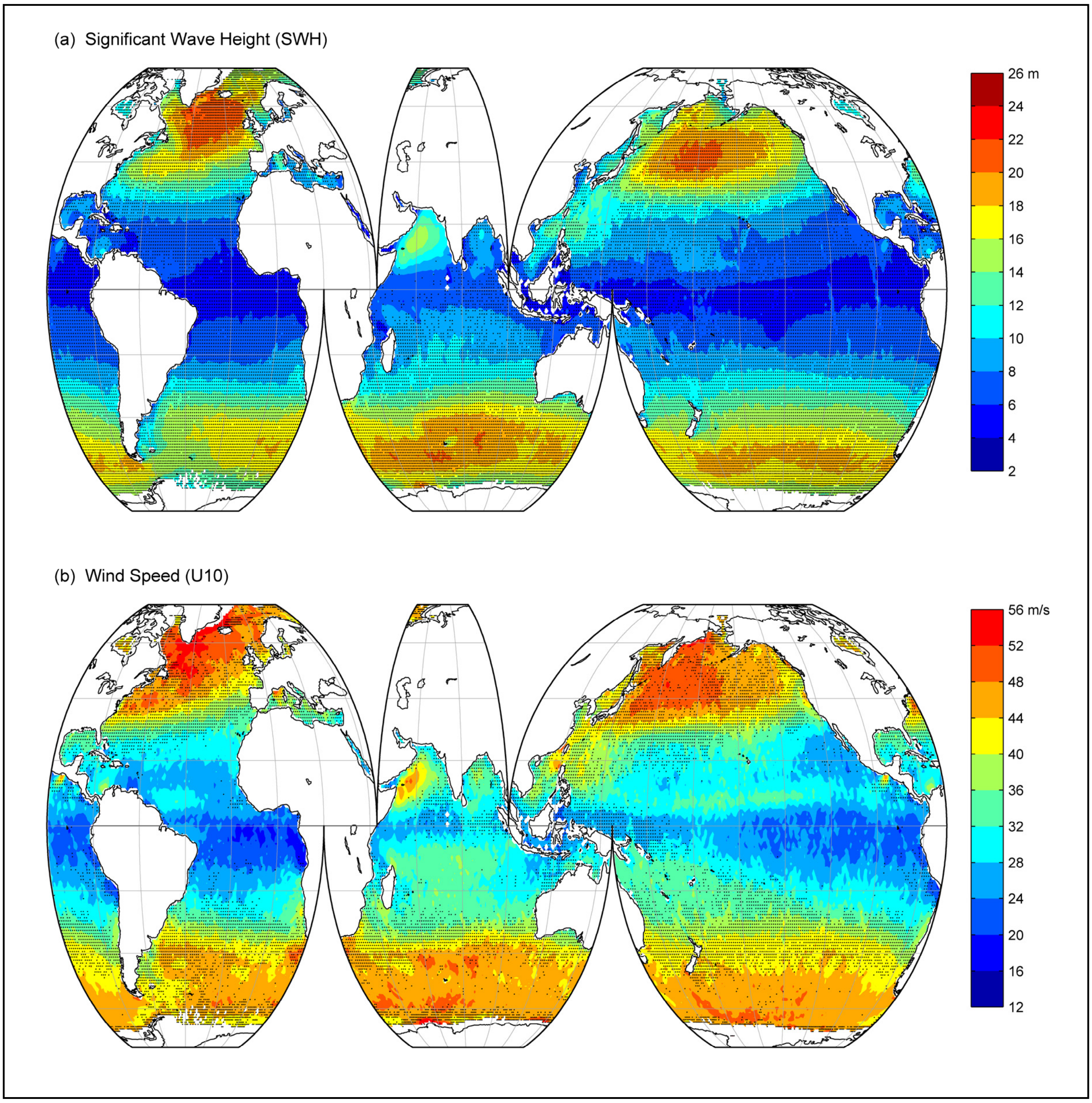

Figure 10: Contour plots of H_s^{100} *(IDM)* (panel a) and U_{10}^{100} *(IDM)* (panel b), as in Figures 6 and 7, respectively. Points which satisfy the Cramér-von Mises goodness-of-fit test censored at 20% for IDM FT-1 are shown by a black dot.