

## Uncertainties of the 50-year wave height estimation using generalized extreme value and generalized Pareto distributions in the Indian Shelf seas

**T. Muhammed Naseef<sup>1</sup>, V. Sanil Kumar<sup>1\*</sup>, Jossia Joseph<sup>2</sup>, B.K.Jena<sup>2</sup>**

<sup>1</sup>Ocean Engineering Division, CSIR-National Institute of Oceanography  
(Council of Scientific & Industrial Research), Dona Paula 403 004, India

<sup>2</sup>Coastal and Environmental Engineering Division, National Institute of Ocean Technology,  
Pallikaranai, Chennai 600 100, India

\*corresponding author: V.S.Kumar email:sanil@nio.org Tel: 0091 832 2450 327

### Abstract

Information about waves with specific return period in a region is essential for the safe design of marine facilities. In this study, significant wave height for 50-year return period is estimated using Generalized Extreme Value (GEV) distribution and Generalized Pareto Distribution (GPD) based on the 15-year wave hindcast data. In order to realize the dependency of nature of the time series data on return value estimation, three types of data series; Daily Maxima (DM), Monthly Maxima (MM) and Annual Maxima (AM) are considered for GEV, whereas for GPD, threshold values are estimated from the parent dataset at 6-h and the DM series. The GEV distribution shows that AM predicts higher significant wave height followed by MM and then DM. The large number (~ 50%) of smaller wave height value (< 1 m) in the DM leads to smaller estimate in wave height for 50-year return period for DM series compared to other data series. Among the locations studied, the maximum value of the significant wave height with 50-year return period by GEV with AM data series is 7.15 m in the western shelf seas and is 7.36 m for the eastern shelf seas, whereas the values based on GPD with Peak Over Threshold are 6.94 m and 7.42 m respectively. Case studies are also done to know the influence of tropical cyclone on the estimated 50-year return value.

**Key words:** Extreme value distribution, surface waves, design wave height, return period, Indian Ocean.

## 1. Introduction

Information on wave height having a certain return period is required to plan marine operation, navigation and design of marine facilities. The success of coastal structures largely depends on the reliable assessment of the environmental loads against which the structures have to withstand (Goda 1991). Long-term statistical estimation of both design waves and operational waves have to be done before planning a marine structure. Such estimates not only contribute to the level of protection but also the scale of investment. Hence, extreme wave analysis is the prime step in coastal and offshore structure design. One of the prominent methods of such analysis is a statistical analysis of the sea state with available wave data. Thus, the accurate information about long-term extreme sea state of interest helps scientifically to the research community and economically to the government and industry (Teena et al. 2012).

In the middle of the 20<sup>th</sup> century itself, scientists like (Gumbel 1959) has developed a statistical method used for forecasting natural random variables like wind speed. Previous authors, like (Goda 1992; Goda et al. 1993; and Mendez et al. 2006) have discussed in detail about the extreme value analysis of design waves. The studies of Ferreira and Guedes Soares (2000) indicates that the forecast of extreme values of significant wave height (SWH) must rely on the methods based on extreme value theory in which it makes use of the largest observation in the sample and whose models are vindicated by general theories. Coles (2001) discussed in detail about the statistics of extreme value prediction considering block maxima and Peak Over Threshold (POT) data points. Bazargan et al. (2005) proposed a method in which it uses non-stationary data for calculating return values for a region. Uncertainty in extreme values of design parameters was quantified by Ying li et al. (2008) with re-sampling techniques. The seasonality and duration in extreme value distributions of SWH using a time-dependent version of POT approach were studied by Mendez et al. (2008).

The present study is conducted to map the SWH for 50-year return period along peninsular India following the results of Teena et al. (2012) who have compared the SWH estimates at a coastal location using different types of distributions on GEV and GPD series. Similar studies have been conducted by Moritz et al. (2004) who analysed the extreme wave height variability along Oregon coast, USA and Neelamani et al. (2007) who estimated the extreme waves for Kuwaiti territorial waters. Panchang et al. (2013) study show that if an engineer wishes to emphasize the upper tail where extremes are likely to occur, then the GEV estimates are to be preferred. Tropical Cyclones (TC) are the main natural hazards which make a catastrophic effect on tropical coastal regions and nearshore platform facilities. Previous studies (Singh et al. 2000) indicate that the frequency and intensity of tropical cyclone in the Indian Ocean increased remarkably during May and November

which account for the higher frequency of intense cyclones in the region. Both the Arabian Sea and the Bay of Bengal are relatively cyclone active regions. Case studies are done to know the influence of tropical cyclone on the estimated 50-year return period significant wave height. The present analysis aims at identifying the uncertainties in the estimation of the design wave height estimate based on different methods utilizing the wave hindcast data for the period 1998-2012. Since the hindcast data is for a limited period of 15 years, we have estimated only SWH for a return period of 50-years.

## **2. Materials and methods**

### **2.1 Data**

For the study, nineteen locations in the shelf seas of India are selected (Table 1). The locations are spaced approximately at  $1.5^\circ$  in the latitude since earlier studies (Kumar et al. 2012; Kumar et al. 2014) has shown that the daily average wave height does not vary significantly within ~200 km distance. Among the selected locations, 11 locations are in the Arabian Sea (W01 to W11) and 8 locations (E01 to E08) are in the Bay of Bengal (Figure 1) and these locations are in water depth greater than 50 m (Kumar and Anoop, 2015). The simulated wave data using the third-generation spectral wave model MIKE21 (Sorensen et al. 2004) generated for the preparation of the wave atlas of the Indian coast (Sivakholundu et al. 2014) is used in the study. The MIKE21 wave model is based on the wave action balance equation where the wave field is represented by the wave action density spectrum and the numerical integration of wave action balance equation is formulated in Cartesian co-ordinates (Komen et al. 1994). The model simulates the wave growth, decay and wave transformation in deep and shallow waters. The model includes wave growth by action of wind, non-linear wave-wave interaction, wave refraction and shoaling. The dissipation due to bottom-friction, white-capping and depth-induced breaking are also included. The directional discretization is done for  $360^\circ$  divided in 16 directions. A logarithmic spectral discretization is employed with a minimum frequency of 0.055 Hz, 25 frequencies, and a frequency factor of 1.1. The details of the model equations and methods of solution are explained in the MIKE-21 user manual (DHI, 2011). The bathymetry was prepared using ETOPO1 (Earth Topography-1 minute) (Amante and Eakins 2009) and MIKE-CMAP data. The model wind forcing is provided by ECMWF operational archive wind data (Trenberth, 1992) at every six hours interval (<http://browse.ceda.ac.uk/browse/badc/ecmwf-op>). The time step of the simulation is 300 s.

The wave hindcast data are validated with the measured data collected using a moored Seatex buoy under the National Data Buoy Programme (Premkumar et al. 2000) in the Arabian Sea (AD02

15.00° N; 69.01° E) during February-December 2012 and in the Bay of Bengal (DS3 12.15° N; 90.75° E) during January-December 2003. Comparison of hindcast SWH data with the measured data indicates that hindcast data is reasonably in good agreement with the observation (Figure 2). The scatter plots of wave height at AD02 are reasonably tight, with Pearson correlation coefficient of 0.97 with bias value of 0.13 m and scatter index of 0.16. The difference between the maximum SWH value is 0.55 m and the difference in the 90 percentile SWH is 0.34 m. In the Bay of Bengal at DS3, the difference between the measured and hindcast SWH data is 0.36 m with the correlation coefficient of 0.92, bias value of 0.3 m and scatter index of 0.23.

In the eastern shelf seas of India, at the southern locations (E01 and E02), the maximum SWH was in December 2011 during the TC Thane. During the TC Thane, maximum SWH of 6 m is measured off Puducherry, eastern shelf seas of India (Kumar et al. 2013) and the maximum SWH estimated at location E02, which is in the TC Thane track is 5.86 m. The model estimated value is close (< 3% error) to the measured SWH value during the TC.

## 2.2 Methods

The estimation of SWH for 50-year return period is carried out by using two extreme value distributions; 1) Generalized Pareto Distribution (GPD) and 2) Generalized Extreme Value (GEV) distribution. A brief description of these two distributions is given in Annexure 1. The present analysis is carried out by using the wave data covering 1998-2012. The parameter determination and fitting of statistical distributions, both GEV and GPD, were carried out separately for each data series by using the routines in WAFO (2000). WAFO is a routine for statistical analysis of wave data. It is important to consider the reliability and appropriateness of the chosen model to deal with the complex wave process. Graphical methods like CDF plots and Quantile-Quantile (Q-Q) plots are used in the model validation process. A typical Q-Q plots which correspond to location W06 are shown in Figure 3. Here, quantiles of an empirical distribution are plotted against the quantiles of fitted GEV/GPD in Q-Q plots. These plots show that both GEV and GPD methods fit very well with the data.

### 2.2.1 GEV

Daily Maxima (DM), Monthly Maxima (MM) and Annual Maxima (AM) data series are extracted from the 6 hourly data and used in the GEV distribution. Parameter estimation is carried out by probability weighted moments (PWM) for the three data series (DM, MM, AM) since PWM methods hold good fit for the distributions having small length (Hosking et al. 1985, Katz et al. 2002).

### 2.2.2 GPD

Apart from DM data series, all the data at 6 h interval are also considered in the GPD. The selection of threshold is an important practical strategy in GPD. In the present study, the threshold selection is carried out by following two-step threshold selection framework proposed by Bernardara et al. (2014) for over threshold modelling and simple mean excess (SME) plots. The former method consists of two steps for selecting a threshold value for GPD estimations; i) A physical threshold aiming with de-clustering of extreme and independent events from the whole data, ii) A statistical threshold through optimisation procedure in which setting a threshold for the convergence of each data points toward the GPD. A second threshold selected from SME plot by choosing lowest value above which the graph is an approximately straight line. The final threshold selected by comparing the both statistical threshold and SME threshold (Figure 4).

For POT analysis, it is necessary that the data possess independent and identical distribution and to ensure this, first the data is extracted by assigning a tentative threshold of mean SWH of the data series. After that, a physical threshold is set in order to obtain 10 values per year in average in such a way that minimum separation time of 48 hours for corresponding peaks with a minimal duration of 8 hours. Finally, the statistical threshold SWH are selected by plotting parameter stability plot and it is compared with the threshold obtained from SME plot by selecting the lowest value above which the line is approximately straight line (Figure 4).

### 2.2.3 Case study to examine the influence of cyclones

We selected three locations (W02, E03 and E08) among the study regions and seven tropical cyclonic disturbances in order to study the uncertainty of SWH return value estimation due to the extreme events. Three different cases are considered in this analysis and results are compared with actual estimation using three different data series. In the first case, the SWH corresponding to the greatest cyclone in the study region is replaced with the average SWH of the 15-year data series. In the second case, minimum SWH is replaced with SWH corresponding to the greatest cyclone in the study region which result to an additional cyclone event in the data series. The third case deals with data series without considering the SWH related to the greatest cyclone.

## 3. Results and Discussion

15 years of wave hindcast data indicate that in the western shelf seas of India, during 45.9 to 58.6% of the time in a year, the SWH is less than 1 m except at location W11 (Table 1). At W11, only during 21.4% of the time, SWH is less than 1 m. In the eastern shelf seas, during 49.8 to 64.8% of time in a year the SWH is less than 1 m. Wave climate in the study region has different features

during the monsoon and non-monsoon seasons. During the pre-monsoon season (FMAM), the region is relatively calm with SWH ranging 0.5-1.5 m, maximum average SWH occurs during the southwest monsoon season (JJAS) when SWH reaches its peaks around 3 m (Anoop et al. 2015). Average SWH during the post-monsoon season (ONDJ) is higher than that in the pre-monsoon season. During the post-monsoon season, average SWH observed in the Bay of Bengal is higher (~2-2.5 m) than that in the Arabian Sea (1.5-2 m) (Anoop et al. 2015). During 1998-2012, the maximum SWH in the Arabian Sea is 7.36 m at location W02 and is 5.99 m at location E08 in the Bay of Bengal. The monthly mean SWH is high in July at all locations studied except at locations E01 to E03. During July, the average SWH along the western shelf seas of India ranged from 2.3 to 3.2 m.

### 3.1 Western shelf seas of India

Figure 5 and Figure 6 shows the estimated parameters for GEV and GPD using PWM methods for different locations in the western shelf seas of India. It is seen that for both AM and MM, majority of GEV shape parameters are negative, indicating the type III distribution (Weibull) which is the most favored distributions in coastal and ocean engineering (Goda et al. 2010; Teena et al. 2012). The rest of the parameters corresponds to type II (Frechet) distribution, and there is no parameterisation corresponds to type-I (Gumbel) distribution for these data series. Shape parameter for DM series indicates type-II distribution except for location W01 and W04 where it is type-III. The GPD shape parameters for DM series are correspond to Frechet distribution whereas shape parameter for 6-h series are related to Weibull distribution except W11 where it is Frechet distribution.

Figure 7 shows the 50-year return period wave height values estimated by GEV using AM, MM and DM data series and GPD using 6-h and DM along the western shelf seas of India. Red dashed lines represent wave height values using MM data, black and green solid lines correspond to values using DM data and black and blue solid thick lines correspond to AM data for GEV and 6-h data for GPD respectively. MM and DM show a clear seasonal cycle with maximum peaks corresponding to the period June-July. Both GEV AM and GPD 6-h distributions project comparable return values. In the case of GEV, both AM and MM estimations are very close to each other with a correlation of 0.99 and root mean square difference (RMSD) of 0.14 m for the area. However, in the case of GPD, the comparability is reduced with a correlation of 0.90 and RMSD of 0.49 m while comparing between DM and 6-h series (Table 2). The peaks corresponding to summer monsoon periods are estimated during May-October having dominance during June. The estimated SWH value for 50-year return period is 7.15 m at W04 by AM and followed by 7.09 m at the same location by MM (Table 1). GPD using 6-h series gives 6.94 m corresponds to 50-year return period. Apart from

dominant peak corresponding to summer monsoon, there are minor peaks during November for central locations. On considering the seasonal cycle of estimated SWH for 50-year return values by each method, DM estimates shows comparable projections among GEV and GPD with less variability within the seasons, whereas MM shows higher variability during summer monsoon period. This may be due to the clustering effect of peaks within the data series. Because MM data set consist of fewer data points having higher SWH compared to DM. The nature of the data series influences the parameter estimation which ultimately affects the SWH estimation. It is seen in shape and GPD scale parameter estimates (Figure 6), in which DM data is more sensitive compared to 6-h data.

Using a data with interval of 6 h have the potential to miss the peaks. The 1-h data set used in Teena et al. (2012) is examined to quantify the underestimation of the peak in 6-h data. Two years are selected; i) 1996, a cyclone year and ii) 2008, a non-cyclone year. The annual maximum SWH for the year 1996 based on 1-h interval data is 5.53 m and that based on 6-h interval data is 5.35 m indicating an underestimation of 3% during the cyclone period. The annual mean SWH is same (1.27 m) for both 1-h and 6-h data, whereas the 90 percentile SWH is 2.48 and 2.47 m for the 1-h and 6-h data. For the year 2008, the annual maximum SWH is 3.76 m for both the 1-h and 6-h data. Similarly the annual mean and 90 percentile SWH has similar values of 1.19 m and 2.38 m respectively.

The maximum value of estimated SWH for 50-year return period in the western shelf seas is at W04 (7.15 m by GEV AM and 6.94 m by GPD 6-h), whereas that in the eastern shelf seas is at E03 (7.36 m by GEV AM and 7.42 m by GPD 6-h). The maximum SWH based on 15 years of hindcast data is less than the estimated 50-year return period at all locations, except at W02. At W02, the 50-year return value (6.91 m by GEV AM and 6.25 m by GPD 6-h) is less than the value (7.36 m) happened during the 15 year period. The SWH for 100-year return period is 8.16 m by GEV using AM data series. The high value of 7.36 m is caused by a cyclonic storm and the study indicates that the return period of the cyclone is more than 50 years and hence a smaller SWH is obtained for 50-year return period. Further, a detailed analysis is carried out to examine the influence of the tropical cyclone and is presented below.

#### Tropical Cyclone influence - location W02

A deep depression developed over the west central Bay of Bengal on 21 June 2007 moved westward and landed on 22 June 2007 at the Andhra Pradesh coast. It further moved across southern India reaching the Arabian Sea as a low-pressure system. After sea fall, the deep depression further

intensified into cyclonic storm 'Yemyin' on 25 June 2007 centered around 22.9° N and 67.3° E (IMD 2007). The wave hindcast model estimated a maximum SWH of 7.36 m at the location W02 during this cyclone. Excluding this event, the maximum SWH at W02 from 1998 to 2012 is 4.50 m. SWH due to the cyclonic storm 'Yemyin' creates anomaly with rest of the data points when considering DM series. In this case study, we altered the SWH during the cyclone period with different values and results are given below.

Figure 8 shows the 50-year return value estimations obtained in different case studies. Table 3 shows the percentage difference among the case studies when different data series are used. Obviously, Case 2 shows percentage increases for all data series. Among the data series, estimation by GEV using MM series accounts maximum percentage difference followed by AM series. Case 1 and Case 3 results of GEV DM series are exactly same. Because DM series still holds the cyclonic effect within the data due the presence of adjacent data points even after changing the highest value corresponds to the cyclone. Both AM and MM data series includes only 15 data points so that the percentage difference corresponds to these data series is very high. It indicates that the number of data points in different series also leads to the uncertainty in return value estimation. The maximum increase is occurred with monthly series of about 22.4% for GEV calculation during Case 2 and maximum decrease is 36.5% account with AM data series using GEV in Case 3.

### 3.2 Eastern shelf seas of India

Figure 9 shows the 50-year return period wave height estimated by GEV and GPD distribution using AM, MM and DM data series along the eastern shelf seas of India. Unlike western shelf seas, estimation using different data series makes higher variability among the return value periods, however GEV/GPD methods for a particular data series estimates are comparable results (Table 2). For example, in the case of GEV, the correlation between DM and MM estimation reduced to 0.68 and correspondingly RMSD increased to 0.50 m, whereas the GEV and GPD estimates using DM data series obtained with higher correlation ~ 0.99 and 0.08 RMSD. In order to test the robustness of the analysis, return values for the 15 years period also compared with the AM data which results in 0.87 correlation for GEV and GPD. AM data series estimates higher return values compared to both MM and DM data series for all locations except E06 and E07 in the region (Table 2). Maximum SWH estimated is 7.42 m by GPD method at E03 and 7.36 m by GEV method at E03. Table 1 and Figure 9 also shows that there is dominance in SWH value during the northeast monsoon over the southwest monsoon period along the southern locations and on reaching northern parts, it tends to get reversed. MM estimates show higher variability in SWH values during different seasons. Maximum wave height obtained during November-December period in the southern



locations (E01-E04) and during June-August period in the northernmost locations (E05-E08). Along the coast, the maximum variability in the GEV estimates among DM and MM estimation recorded at E03 followed by E02 with a difference in SWH of about 3.77 m and 3.69 m respectively

Parameter estimation using PWM method shows that GEV estimation using both AM and MM series are Weibull distribution along the eastern coast except for the location E07 where it is Frechet distribution. GEV shape parameter ranging in between -0.40 m to -0.18 m while GPD shows higher scale of -0.43 m to 0.46 m (Figure 6). GEV location parameter for AM series are almost double of the DM series up to E04 and after that, the contrast further increases as AM values are higher.

#### Tropical Cyclone influence - locations E03 and E08

The Bay of Bengal is remarkably prone to tropical cyclone activity compared to the Arabian sea. Therefore different case studies in Bay of Bengal make significant difference with the features we have seen in Arabian Sea. During the study period, tropical cyclones/Depressions mainly concentrated along northern and southern locations. Therefore, in the eastern shelf sea, we considered two locations (E03 and E08) for the case studies. At E03, three tropical cyclones are considered which have significant influence on the return value estimates. While in the case of E08, we considered four depressions/deep depressions. The characteristic features of the tropical cyclone and depressions are presented in Table 4. The seasonal cycle of 50-year return value estimation corresponding to the different cases for both locations are shown in Figure 10 and Figure 11 respectively. In the case of location E03, the tropical cyclones considered for the location happened during the months, May, November and December and hence, its influence can be seen in different cases (Figure 10). Table 3 present the percentage difference occurred during different cases. Unlike in the case of W02, three cases result in different return value estimations and GEV/GPD estimations corresponding to DM series makes higher percentage difference. The addition of extra cyclone effect (case 2) made a significant difference and the deletion of cyclone data points (case 3) made higher reduction to the result with AM series. GEV estimation corresponds to DM series is least affected except in the case 2 whereas the GEV with MM and AM data series is most affected during different cases for the location E03 (Figure 10). The maximum increase of 90.4% with GEV with DM series during case 2 and the maximum decrease of 34.2% occurred with GEV with AM series during case 3 (Table 3).

Depressions/deep depressions considered for cases with location E08 happened two events each during June and September. Different return values obtained by different cases are shown in

Figure 11 and Table 3. Unlike the other two locations, W02 and E03, here we can notice that the different results of GPD estimation with DM series are less affected compared to the actual results except in case 2. The case studies show that the GEV estimation makes maximum increase of 60.3% with DM series during case 2 followed by GPD with 55.5% since two events occurred in September. The maximum decrease of 26.3% occurred with GEV estimation with AM data during case 3. The study shows that occurrence of another cyclone with intensity similar to the strongest cyclone occurred in the region increase the 50-year return value by 60.3%. Whereas if the strongest cyclone has not occurred, the 50-year return value reduces by 26.3%.

#### **4. Summary and Conclusion**

The extreme wave analysis is carried out by using 6 hourly simulated data for the period of 1998 to 2012 using numerical model MIKE21. The SWH having 50-year return period is obtained by GEV and GPD distributions adopting PWM method for parameter estimation. The parameter estimation for both the eastern and western shelf seas indicates that Weibull distribution for GEV method predominantly Whereas in the case of GPD, the estimated parameters indicates Frechet distributions for DM along West coast locations and 6-h along East coast locations In order to understand the behavior of the estimates on the nature of time series data, we extracted three types of data series, viz. DM, MM and AM and fitted the GEV distributions separately. For GPD, 6-h data and DM are used. The analysis shows that SWH return values largely depend on the nature of the data series. That is, MM data consist of comparatively higher values of distinct extreme events with less number of data points. Whereas in the case of DM, the data points are more clustered, so it leads to underestimation of SWH. Maximum SWH estimation is obtained by using AM data series along both the western and eastern shelf seas of India. In the western shelf seas, maximum SWH corresponding to 50-year return period estimated is 7.15 m at the location W04 using AM data and 7.09 m using MM data. Along the western shelf seas of India, 50-year return period SWH values are less in northern and again in the southern regions compared to central regions. In the eastern shelf seas, maximum SWH estimated is 7.42 m by GEV at E03 followed by 7.36 m at same location by GPD using 6-h. Replacing the maximum value in DM data series with an average value of the DM series results a maximum difference of 19.4%, while removing the maximum data results in difference up to 18.6%. Addition of another maximum value in the DM series makes a significant increase (90.4%) in the SWH with 50-year return period. Values presented in the paper is to be used with caution for design purpose since cyclone wind conditions are not resolved properly in the coarse grid of the model.

## Annexure 1. Distributions used in the study

### Generalized Extreme Value (GEV)

Jenkinson (1955) proposed the classical theory based on extreme value distribution advanced by Fisher and Tippet (1928), in which the three kinds of typical extreme value distributions were coupled into a three parameter GEV distribution as a new model (Gumbel,1959). This has the cumulative distribution function (CDF) as:

$$F(X) = \begin{cases} \exp \left\{ - \left( 1 - k \left( \frac{X-\beta}{\alpha} \right)^{1/k} \right) \right\}, & k \neq 0 \\ \exp \left\{ - \exp \left( - \frac{X-\beta}{\alpha} \right) \right\}, & k = 0 \end{cases} \quad (A1)$$

Where the parameters  $\alpha$ ,  $\beta$ ,  $k$  are called the scale (dispersion), location (mode of distribution), shape parameters (type of distribution) and satisfy  $-\infty < k < \infty$ ,  $\alpha > 0$  and  $-\infty < \beta < \infty$ . For,  $k = 0$ , the GEV is Fisher-Tippet type I or Gumbel distribution, For  $k > 0$  it is the Fisher-Tippet type II or Frechet distribution, and for  $k < 0$ , it is the Fisher-Tippet type III or Weibull distribution. The tail of the Frechet distribution is ‘heavier’ (that is, decrease more slowly) than that of the Gumbel distribution and for Weibull distribution it is ‘lighter’ (decreases more quickly and reaches zero) than that of the Gumbel distribution (Palutikof et al. 1999). The estimation of N-year return value, the value that will be exceeded, on average, once in N years, for GEV is given as (Palutikof et al. 1999):

$$X_T = \begin{cases} \beta + \frac{\alpha}{k} \left\{ 1 - \left[ - \ln \left( 1 - \frac{1}{T} \right) \right]^k \right\}, & k \neq 0 \\ \beta - \alpha \ln \left[ - \ln \left( 1 - \frac{1}{T} \right) \right], & k = 0 \end{cases} \quad (A2)$$

### Generalized Pareto Distribution (GPD)

One of the major drawbacks of GEV is that it only considers a single peak (maximum) value within each reference period of the observation. Thereby it ignores the other extreme values that have significant importance in the extreme event analysis. Another drawback is that GEV sampling scheme requires a long period of observations (large sample size) for good results. Alternative approaches to the GEV sampling, known as Peak Over Threshold (POT), which considers all values greater than a significantly high threshold. GPD is capable of extracting the maxima in each epoch above a given threshold following POT sampling scheme. So that GPD can overcome the above mentioned inconveniences. The distribution of  $x$  over  $u$  for a threshold  $u$ , is as:

$$F_u(y) = \Pr\{x - u \leq y \mid x > u\} = \frac{F(x) - F(u)}{1 - F(u)} \quad (A3)$$

which is the probability that the value of  $x$  exceeds  $u$  by at most a quantity  $y$ , where  $y = x - u$ .

For high threshold,  $u$ , the distribution of the excess,  $F_u(y)$ , converges to GPD and the CDF is (Balkema and de Haan 1974; Pickands 1975):

$$G(x, k, \alpha, \beta) = \begin{cases} 1 - \left(1 - k \frac{x-\beta}{\alpha}\right)^{1/k}, & k \neq 0 \\ 1 - e^{-(x-\beta)/\alpha} & k = 0 \end{cases} \quad (\text{A4})$$

when  $k = 0$ , the GPD is an exponential distribution (type I tail or medium-tailed) with mean  $\alpha$ . if  $k > 0$ , it is a Pareto II type distribution (short-tailed or type II tail); and when  $k < 0$ , it is in the form of ordinary Pareto distribution (type III tail or long-tailed) (Caires 2011).

The quantile estimation for GPD for threshold value  $u$  for different types of parameter estimation methods is (Abild et al. 1992):

$$X_T = \begin{cases} k + \frac{\alpha}{\beta} [1 - (\lambda T)^{-k}] & k \neq 0 \\ k + \alpha \ln(\lambda T) & k = 0 \end{cases} \quad (\text{A5})$$

Here,  $\lambda$  is, the crossing rate of threshold  $u$ , given as  $N_u/N$ . Where  $N_u$  is the total number of exceedances over the selected threshold  $u$  and  $N$  are the total numbers of years of record; and  $\alpha$  and  $k$  are the fitted GPD parameters to  $x-u$ , where  $x > u$ .

## References

- Abild J, Andersen EY, Rosbjerg D (1992) The climate of extreme winds at the Great Belt, Denmark J Wind Eng and Indus Aerodynamics 41(1):521-532. doi:10.1016/0167-6105(92)90458-M,.
- Anoop, TR, Kumar, VS, Shanas, PR, Johnson, G (2015) Surface wave climatology and its variability in the North Indian Ocean based on ERA-Interim reanalysis. J Atmo and Oce Tech. doi: <http://dx.doi.org/10.1175/JTECH-D-14-00212.1>.
- Balkema, AA, De Haan, L (1974) Residual life time at great age. Ann prob 2: 792-804.
- Bernardara, P, Mazas F, Kergadallan X, Hamm L (2014) A two-step framework for over-threshold modelling of environmental extremes. Nat Hazard Earth Sys Science 14(3): 635-647. doi:10.5194/nhess-14-635-2014.
- Caires, S, Sterl, A (2005) 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data. J Climate 18(7): 1032-1048. doi: <http://dx.doi.org/10.1175/JCLI-3312.1>.
- Coles S (2001) An introduction to statistical modeling of extreme values (Vol. 208) Springer London p209 ISBN: 978-1-4471-3675-0.
- Fisher RA, Tippett, LHC (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. Math Proc Cambridge (Vol. 24, No. 02) Cambridge University Press 180-190.

- Goda Y (1992) Uncertainty of design parameters from viewpoint of extreme statistics. *J offshore mech arctic eng ASME* 114(2): 76-82. doi:10.1115/1.2919962.
- Goda Y, Watanabe N (1991) A longshore current formula for random breaking waves. *Coast Eng Japan* 34(2): 159-175. doi: <https://doi.org/10.1080/05785634.1991.11924547>.
- Goda Y, Hawkes P, Mansard E, Marti, MJ, Mathiesen M, Peltier E, van Vledder G (1993) Intercomparison of extremal wave analysis methods using numerically simulated data. *Ocean Wave Measurement and Analysis ASCE* : 963-977.
- Goda Y, Kudaka M, Kawai H (2010) Incorporation of weibull distribution in L-moments method for regional frequency analysis of peaks-over-threshold wave heights. *Coastal Engineering Proceedings*, 1(32):62. doi: <http://dx.doi.org/10.9753/icce.v32.waves.62>
- Gumbel EJ (1959) *Statistics of extremes*. Columbia Univ. press, New York.
- Hosking JRM, Wallis JR, Wood EF (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27(3):251-261. <http://dx.doi.org/10.1007/s11069-012-0229-y>.
- Jenkinson AF (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 81(348):158-171.
- Komen GJ, Cavaleri L, Doneland M, Hasselmann K, Hasselmann S, Janssen PAEM (1994) *Dynamics and Modelling of Ocean Waves*. Cambridge University Press, Cambridge.
- Kumar VS, Johnson G, Dora GU, Chempalayil SP, Singh J, Pednekar P (2012) Variations in nearshore waves along Karnataka, west coast of India. *Journal of earth system science*, 121(2), 393-403. doi:10.1007/s12040-012-0160-3.
- Kumar VS, Glejin J, Dora GU, Sajiv CP, Singh J, Pednekar P (2012) Variations in nearshore waves along Karnataka, west coast of India. *J. Earth System Science* 121, 393-403.
- Kumar VS, Shanas, PR, Dubhashi, KK(2014) Shallow water wave spectral characteristics along the eastern Arabian Sea. *Nat Hazards* 70:377–394.
- Li Y, Simmonds D, Reeve D (2008) Quantifying uncertainty in extreme values of design parameters with resampling techniques. *Ocean Engineering*, 35(10): 1029-1038. doi:10.1016/j.oceaneng.2008.02.009.
- Méndez FJ, Menéndez M, Luceño A, Losada IJ (2006) Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. *Journal of Geophysical Research: Oceans* (1978–2012), 111(C7) doi: 10.1029/2005JC003344.
- Méndez FJ, Menéndez M, Luceño A, Medina R, Graham NE (2008) Seasonality and duration in extreme value distributions of significant wave height. *Ocean Engineering* 35(1):131-138. doi:10.1016/j.oceaneng.2007.07.012.
- Moritz HP, Moritz HR (2004) Regional Analysis of Extremal Wave Height Variability Oregon Coast, USA. In *Proceedings 8th International Workshop in Wave Hindcasting and Forecasting*, Oahu, Hawaii, USA 14-19 2004, p1-15.

- Neelamani S, Al-Salem K, Rakha K (2007) Extreme waves for Kuwaiti territorial waters. *Ocean engineering*, 34(10):1496-1504. doi:10.1016/j.oceaneng.2006.08.013
- Palutikof JP, Brabson BB, Lister DH, Adcock ST (1999) A review of methods to calculate extreme wind speeds. *Meteorological applications* 6(02):119-132.
- Panchang V, Jeong CK, and Demirbilek Z (2013) Analyses of Extreme Wave Heights in the Gulf of Mexico for Offshore Engineering Applications, *Journal of Offshore Mechanics and Arctic Engineering*, Transactions of the ASME, Volume 135, pp 031104-1 - 031104-15. doi: 10.1115/1.4023205.
- Pickands III J (1975) Statistical inference using extreme order statistics. *The Annals of Statistics*, 119-131.
- Premkumar K, Ravichandran M, Kalsi SR, Sengupta D, Gadgil S (2000) First results from a new observational system over the Indian Seas. *Curr Sci.* 78(3):323-330.
- Regional specialized meteorological centre-tropical cyclones, India meteorological department, RMSC (2013) Web (01 May 2016) <http://www.rsmcnewdelhi.imd.gov.in/index.php?lang=en>.
- Singh OP, Khan TA, Rahman MS (2000) Changes in the frequency of tropical cyclones over the North Indian Ocean. *Meteorology and Atmospheric physics*, 75(1-2), pp.11-20.
- Sivakholundu KM, Jossia JK, Jena BK (2014) *Wave Atlas of the Indian Coast*, National Institute of Ocean Technology, Chennai, India, ISBN-81901338-4-5, 130p.
- Soares CG, Teixeira AP (2000) Structural reliability of two bulk carrier designs. *Marine Structures*, 13(2):107-128. doi:10.1016/S0951-8339(00)00004-6
- Sørensen OR, Kofoed-Hansen H, Rugbjerg M, Sørensen LS (2004) A third-generation spectral wave model using an unstructured finite volume technique. *Proc. 29th Intl. Conf. on Coastal Engineering*, Lisbon, Portugal, ASCE, 894–906.
- Teena, N. V, Kumar, V. S, Sudheesh, K, Sajeev, R (2012) Statistical analysis on extreme wave height. *Natural hazards*, 64(1):223-236. doi:10.1007/s11069-012-0229-y.

## Figure captions

Figure 1. Map showing the locations considered in the study

Figure 2. Scatter plot of wave hind cast data with the buoy measured data a) at AD02 in Arabian Sea and b) DS3 in Bay of Bengal. The solid line is the exact match line and the dotted lines are 15% deviation

Figure 3. QQ plot for fitted GEV (a) and corresponding QQ plot for fitted GPD (b) for a typical location (W06)

Figure 4. A typical SME plot (top) and corresponding Parameter stability plot (bottom) for location W06 along western shelf seas of India. From these plots, 3.34 m selected as threshold for the location

Figure 5. The parameters estimated for GEV using different data series for locations in the western (left panel) and the eastern (right panel) shelf seas of India. a) and b) are shape parameters, c) and d) are scale parameters and e) and f) are location parameters

Figure 6. The parameters estimated for GPD using different data series for locations in the western (left panel) and the eastern (right panel) shelf seas of India. a) and b) are shape parameters, and c) and d) are scale parameters

Figure 7. The 50-year return period significant wave height estimated using GEV (DM, MM and AM) and GPD (6-h and DM) by different data series for the locations along western shelf seas of India

Figure 8. Estimated SWH values for 50-year return period obtained by different case studies with tropical cyclone Yemyin at the location W02 using GEV and GPD

Figure 9. Estimated 50-year return period significant wave height using GEV (DM, AM, MM) and GPD (6-h and DM) by different data series (and 6-h) for the locations along eastern shelf seas of India

Figure 10. Estimated SWH for 50-year return period obtained for different case studies with tropical cyclones at location E03 using GEV and GPD

Figure 11. Estimated SWH for 50-year return period obtained by different case studies with tropical cyclones at location E08 using GEV and GPD

Table 1. The locations considered in the study, the annual mean value of significant wave height and the estimated significant wave height for 50-year return period by using different data series and corresponding month of occurrence

Location	Position	SWH <0.5 m (%)	SWH <1 m (%)	Mean SWH (m)	Daily Maximum Data				Monthly Maximum Data			AM data	6-h data GPD (m)
					Month	GEV (m)	Month	GPD (m)	Month	GEV (m)	Month	GEV (m)	
W01	22.50° N;68.25° E	6.3	52.0	1.3	Jun	3.68	Jun	3.75	Jun	6.26	Jun	6.16	5.96
W02	21.00° N; 69.00° E	2.3	48.3	1.4	Jun	4.22	Jun	4.53	Jun	7.01	Jun	6.91	6.25
W03	19.50° N; 71.25° E	5.3	54.2	1.3	Jun	4.06	Jun	4.36	Jun	6.00	Jun	5.97	5.74
W04	18.00° N; 72.00° E	4.3	55.8	1.4	Jun	4.46	Jun	4.67	Jun	7.09	Jun	7.15	6.94
W05	16.50° N; 72.75° E	5.6	57.0	1.3	Jun	4.40	Jun	4.51	Jun	6.81	Jun	6.80	6.86
W06	15.00° N; 73.50° E	6.6	58.6	1.3	Jun	4.12	Jun	4.30	Jun	6.84	Jun	6.53	6.60
W07	13.50° N; 73.50° E	2.0	51.1	1.4	Jun	4.14	Jun	4.28	Jun	6.28	Jun	6.24	6.33
W08	12.00° N; 74.25° E	3.2	51.2	1.3	Jun	3.85	Jun	4.00	Jun	5.59	Jun	5.31	5.34
W09	10.50° N; 75.00° E	2.1	48.2	1.3	Jun	3.58	Jun	3.83	May	5.08	May	5.22	5.37
W10	09.00° N; 75.75° E	2.6	45.9	1.3	Jun	3.33	Jun	3.62	May	4.52	Jun	4.67	4.81
W11	07.50° N; 77.25° E	0.2	21.4	1.6	Jun	3.05	Jun	3.19	May	3.90	Jun	3.87	4.04
E01	10.50° N; 81.00° E	2.5	50.2	0.9	Dec	2.29	Dec	2.22	Dec	4.70	Dec	4.77	5.12
E02	12.00° N; 81.00° E	2.3	54.2	1.1	Dec	2.29	Dec	2.18	Dec	5.98	Dec	6.82	6.84
E03	13.50° N; 81.00° E	3.0	62.2	1.0	Dec	2.17	Nov	2.20	Dec	5.94	Nov	7.36	7.42
E04	15.00° N, 81.00° E	3.7	64.8	1.0	Dec	1.98	Nov	2.04	Nov	4.97	Nov	6.08	6.15
E05	16.50° N, 83.00° E	3.8	49.8	1.2	Jun	2.34	Jun	2.29	Dec	4.79	Dec	5.37	5.47
E06	18.84° N; 85.00° E	7.6	52.8	1.4	Jul	2.70	Jun	2.67	Aug	6.00	Sep	5.96	6.33
E07	19.50° N; 87.00° E	14.5	54.6	1.4	Jul	3.19	Jun	3.23	Jun	5.80	Aug	5.44	5.85
E08	20.25° N; 88.50° E	18.8	55.5	1.4	Jun	3.51	Jun	3.67	Jun	6.77	Sep	7.23	7.41



Table 2. Statistical inter comparison of return value estimation by different data series using GEV/GPD methods

No	Comparison	RMSD	Correlation
Locations along the western shelf seas of India			
01	GEV <sub>DM</sub> vs GEV <sub>MM</sub>	0.63	0.94
02	GEV <sub>MM</sub> vs GEV <sub>AM</sub>	0.14	0.99
03	GEV <sub>DM</sub> vs GPD <sub>DM</sub>	0.35	0.79
04	GPD <sub>DM</sub> vs GPD <sub>6-h</sub>	0.49	0.91
05	GEV model vs data	0.30	0.95
06	GPD model vs data	0.43	0.91
Locations along the eastern shelf seas of India			
01	GEV <sub>DM</sub> vs GEV <sub>MM</sub>	0.49	0.69
02	GEV <sub>MM</sub> vs GEV <sub>AM</sub>	0.56	0.76
03	GEV <sub>DM</sub> vs GPD <sub>DM</sub>	0.08	0.99
04	GPD <sub>DM</sub> vs GPD <sub>6-h</sub>	0.82	0.29
05	GEV model vs data	0.46	0.87
06	GPD model vs data	0.73	0.76

Table 3. Estimated return values obtained by different case studies for locations W02, E03 and E08. Case 1 is maximum value of series replaced by average value of the series. Case 2 is minimum value of the series is replaced by the maximum value. Case 3 is the series after removal of the cyclone/storm event related to maximum value. Table also shows the percentage difference occurred among case results with actual result.

Data series	Actual result		Case 1				Case 2				Case 3				
	GEV	GPD	GEV	%	GPD	%	GEV	%	GPD	%	GEV	%	GPD	%	
Location W02															
Yearly	6.91	-	4.58	-33.7	-	-	8.35	+20.8	-	-	4.39	-36.5	-	-	
Monthly	7.01	-	4.70	-33.0	-	-	8.58	+22.4	-	-	4.69	-33.1	-	-	
Daily	4.22	4.74	4.13	-2.1	4.52	-4.6	4.33	+2.6	5.42	+14.3	4.13	-2.1	4.52	-4.6	
Location E03															
Yearly	7.36	-	5.40	-26.6	-	-	7.66	+4.1	-	-	4.84	-34.2	-	-	
Monthly	5.94	-	4.36	-26.6	-	-	6.92	+16.5	-	-	4.13	-30.5	-	-	
Daily	2.39	2.96	2.30	-3.8	2.38	-19.4	4.55	+90.4	5.15	+74.0	2.29	-4.1	2.41	-18.6	
Location E08															
Yearly	7.23	-	-	-19.1	-	-	7.39	+2.2	-	-	5.33	-26.3	-	-	
Monthly	6.77	-	-	-16.0	-	-	7.34	+8.4	-	-	5.69	-16.0	-	-	
Daily	3.88	4.18	3.71	-4.5	4.18	0.0	6.22	+60.3	6.50	+55.5	3.69	-4.9	4.21	+0.7	

Table 4. The Characteristic features of tropical cyclones and depressions considered in the case studies related to eastern shelf sea. (source: IMD annual cyclone review and RSMC reports)

No.	Cyclonic Storm Or Depression	Date, Time & place of genesis (Lat °N/Lon °E)	Date, Time (UTC) and place of landfall	Estimated maximum wind(kt), Date & Time	Maximum T. No attained	Max SWH observed (m) and the date
E03-1	Severe Cyclonic Storm, "LAILA" over the BoB (17-21 May, 2010).	17 May, 0600 UTC near 10.5/88.5	Crossed Andhra Pradesh coast near 16.0/80.5 between 1100 & 1200 UTC of 20 May	55 kt at 0600 UTC of 19 May near 13.5/81.5	T 3.5	5.44 19-May-2010
E03-2	Severe Cyclonic Storm "JAL" over the BoB (04-08 November 2010)	04 November, 0000 UTC near 8.0/92.0	Crossed north Tamil Nadu and south Andhra Pradesh coast near 13.3/80.2 around 1600 UTC of 07 November	60 kt at 1200 UTC of 06 November near 11.0/84.5 .	T 3.5	4.99 7-Nov-2010
E03-3	Very Severe Cyclonic Storm "THANE" over the BoB (25-31 December, 2011)	25 December, 1200 UTC near 8.5/88.5	Crossed north Tamil Nadu & Puducherry coast within 0100 and 0200 UTC of 30 December, 2011	75 kts at 0600 UTC of 29 December near 12.0/82.0	T 4.5	5.12 29-Dec-2011
E08-1	Deep depression over BoB June 11-14	11 June, 0300 UTC near 15.5/90.0	Crossed Orissa coast near Puri between 0400 UTC to 0500 UTC June 13	30 kt on June 12 at 0300 UTC	T 2.0	5.87 13-Jun-2004
E08-2	Deep Depression over the BoB 28-30 June	28 June, 0000 UTC near 18.5/87.0	Crossed Odisha coast near Puri between 0000-0100 UTC of 29 June	30 kt at 0300 UTC 28 June.	T 2.0	5.98 29-Jun-2007
E08-3	Depression over the BoB 21-24 September	21 September 1200 UT near 18.0/ 86.5	Crossed Odisha coast near Puri between 1300-1400 UTC of 22 September.	25 kt at 1200 UTC of 21 September.	T 1.5	5.42 23-Sep-2007
E08-4	Deep Depression over the BoB during 15-19 September.	15 September 1200 UTC near 19.5/88.5	Crossed Odisha coast near Chandbali between 1600-1700 UTC of 16 September.	30 kt at 0300 UTC of 16 September	T 2.0	4.27 16-Sep-2008

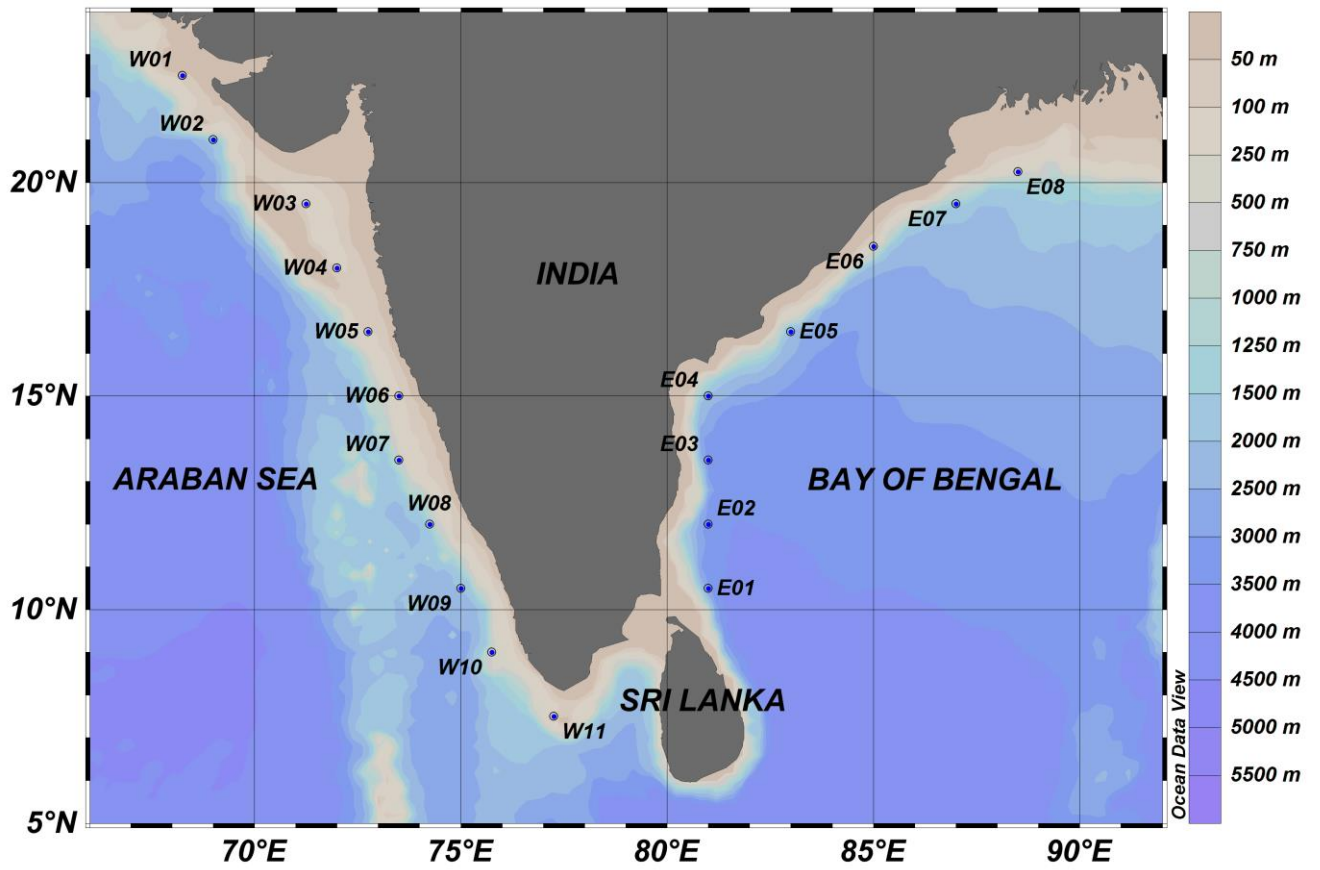


Figure 1. Map showing the locations considered in the study

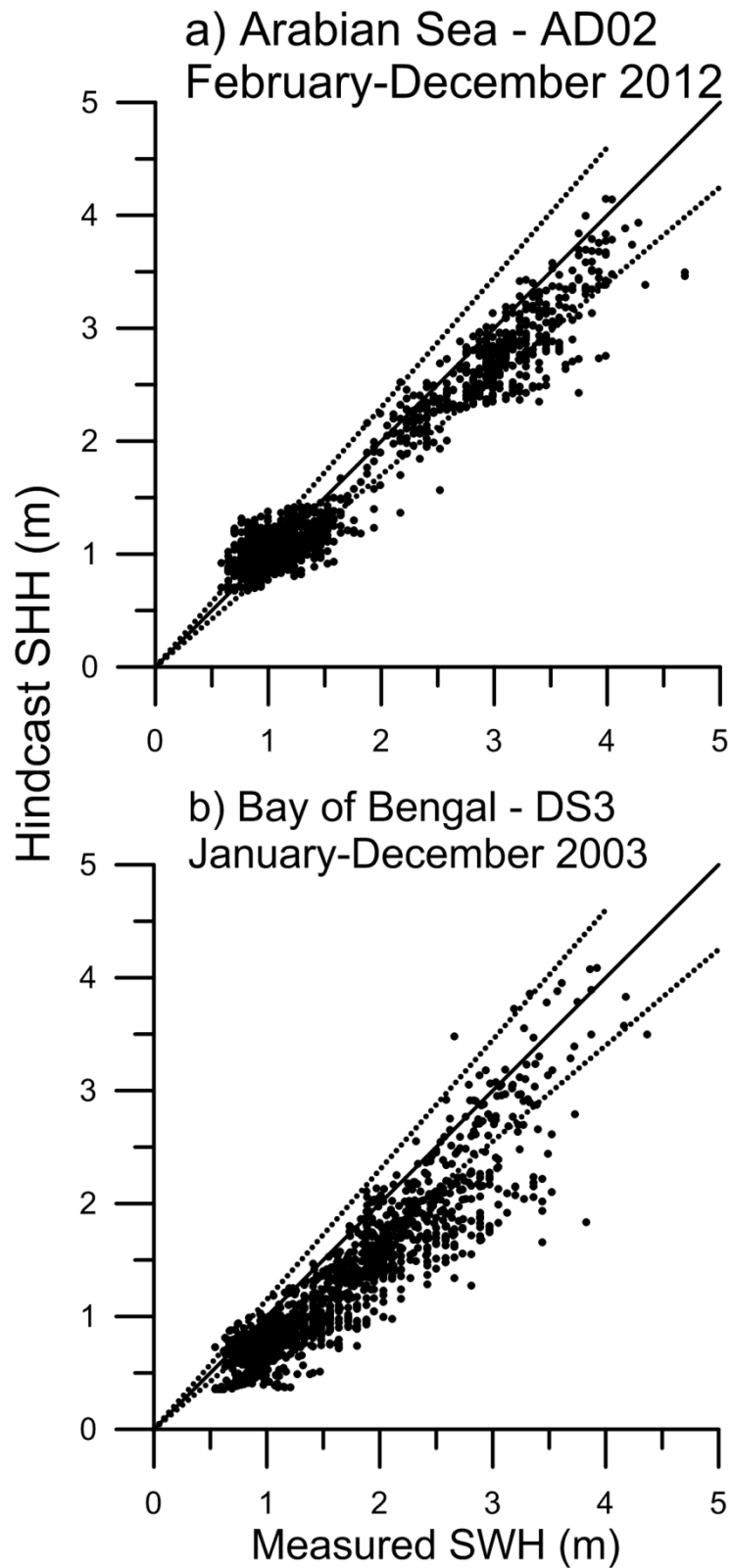


Figure 2. Scatter plot of wave hind cast data with the buoy measured data a) at AD02 in Arabian Sea and b) DS3 in Bay of Bengal. The solid line is the exact match line and the dotted lines are 15% deviation

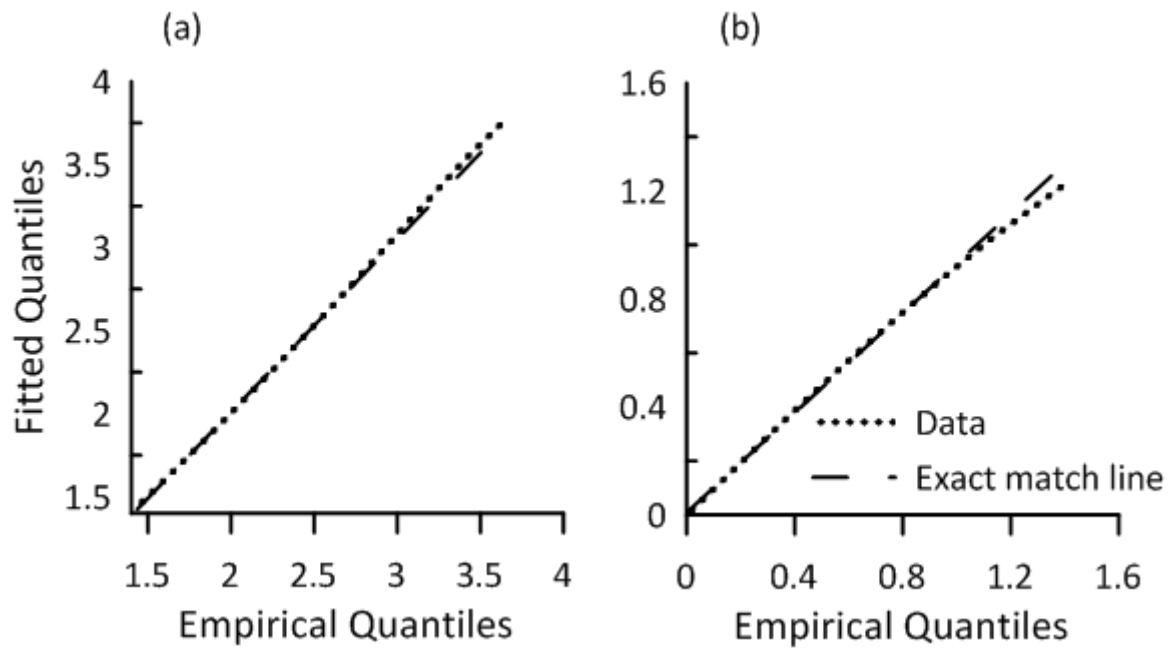


Figure 3. QQ plot for fitted GEV (a) and corresponding QQ plot for fitted GPD (b) for a typical location (W06)

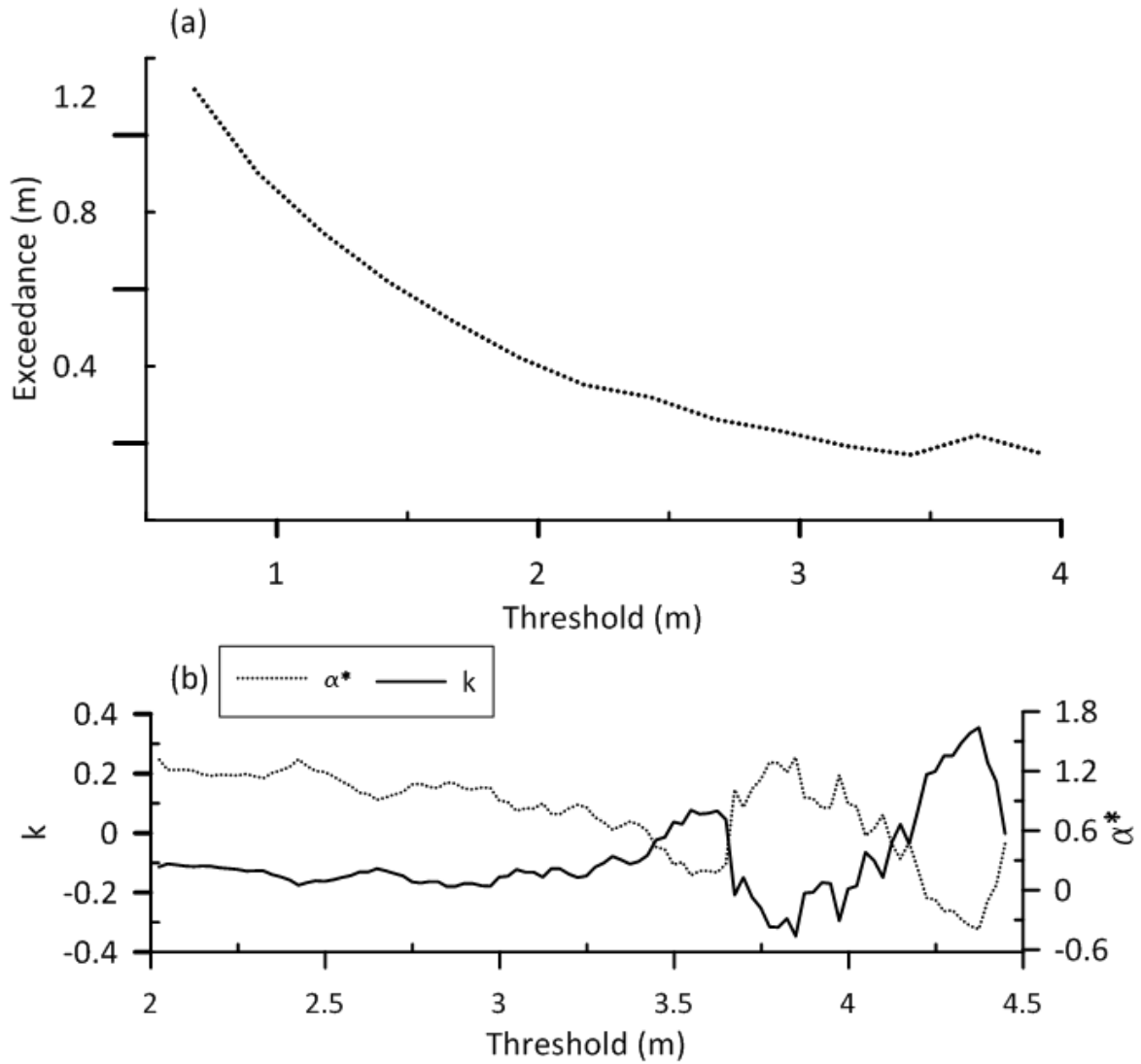


Figure 4. A typical SME plot (top) and corresponding Parameter stability plot (bottom) for location W06 along western shelf seas of India. From these plots, 3.34 m selected as threshold for the location

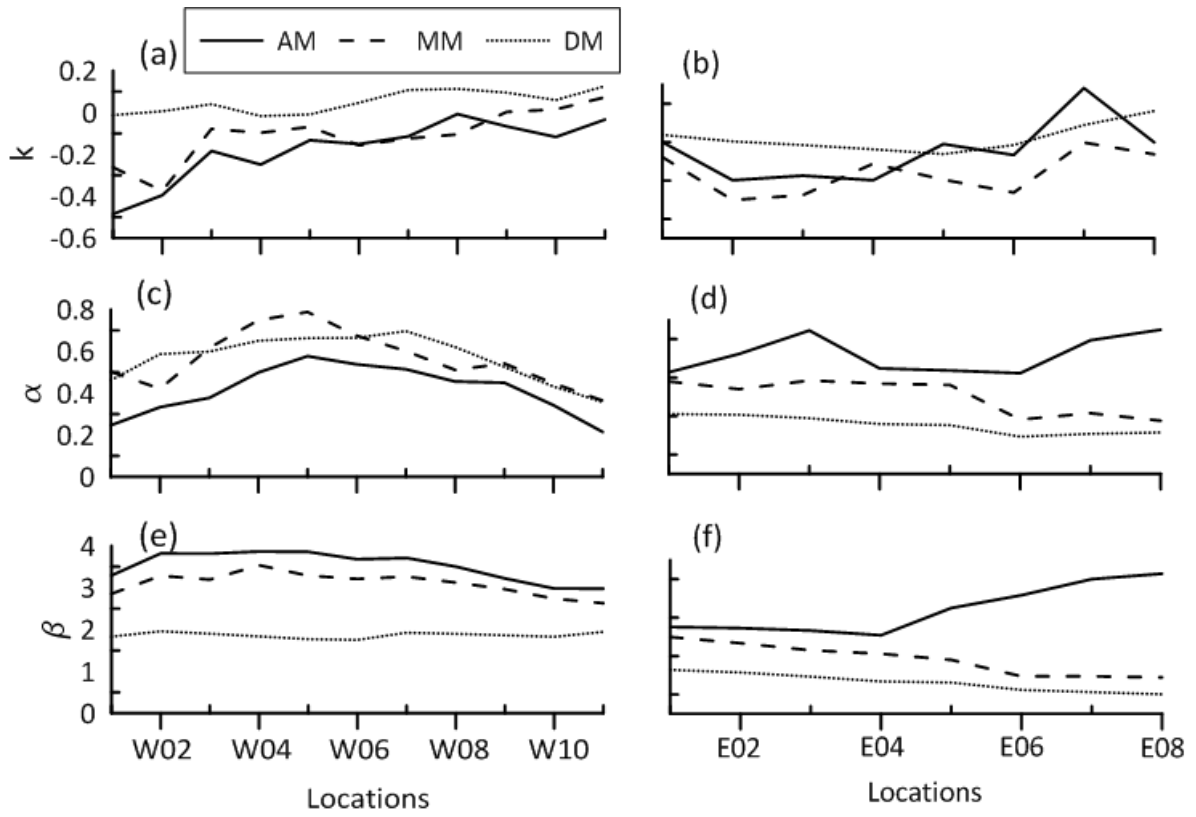


Figure 5. The parameters estimated for GEV using different data series for locations in the western (left panel) and the eastern (right panel) shelf seas of India. a) and b) are shape parameters, c) and d) are scale parameters and e) and f) are location parameters



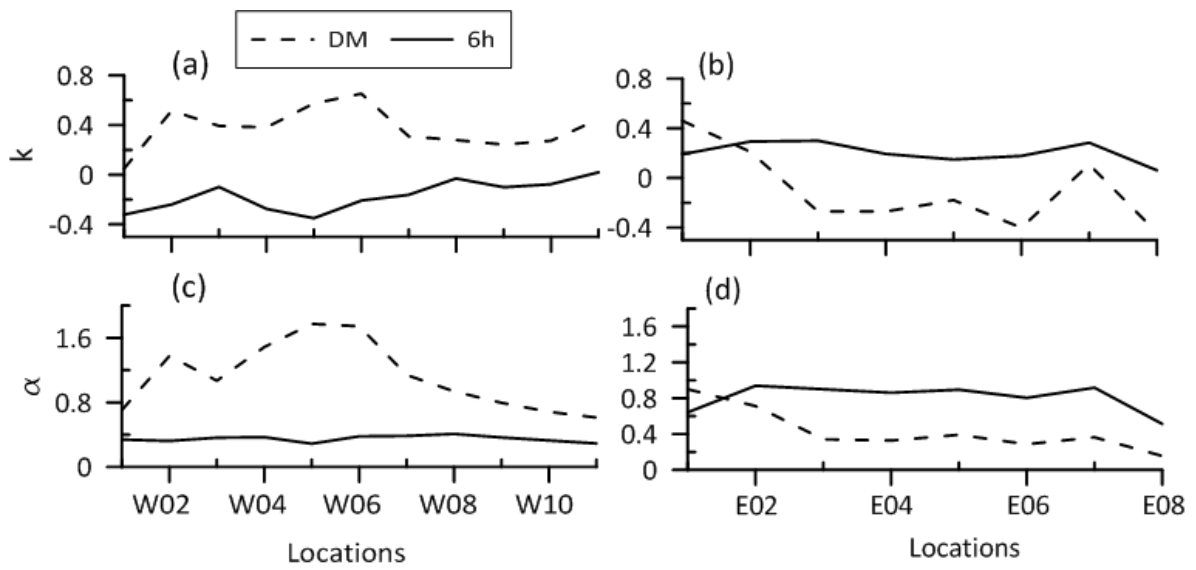


Figure 6. The parameters estimated for GPD using different data series for locations in the western (left panel) and the eastern (right panel) shelf seas of India. a) and b) are shape parameters, and c) and d) are scale parameters

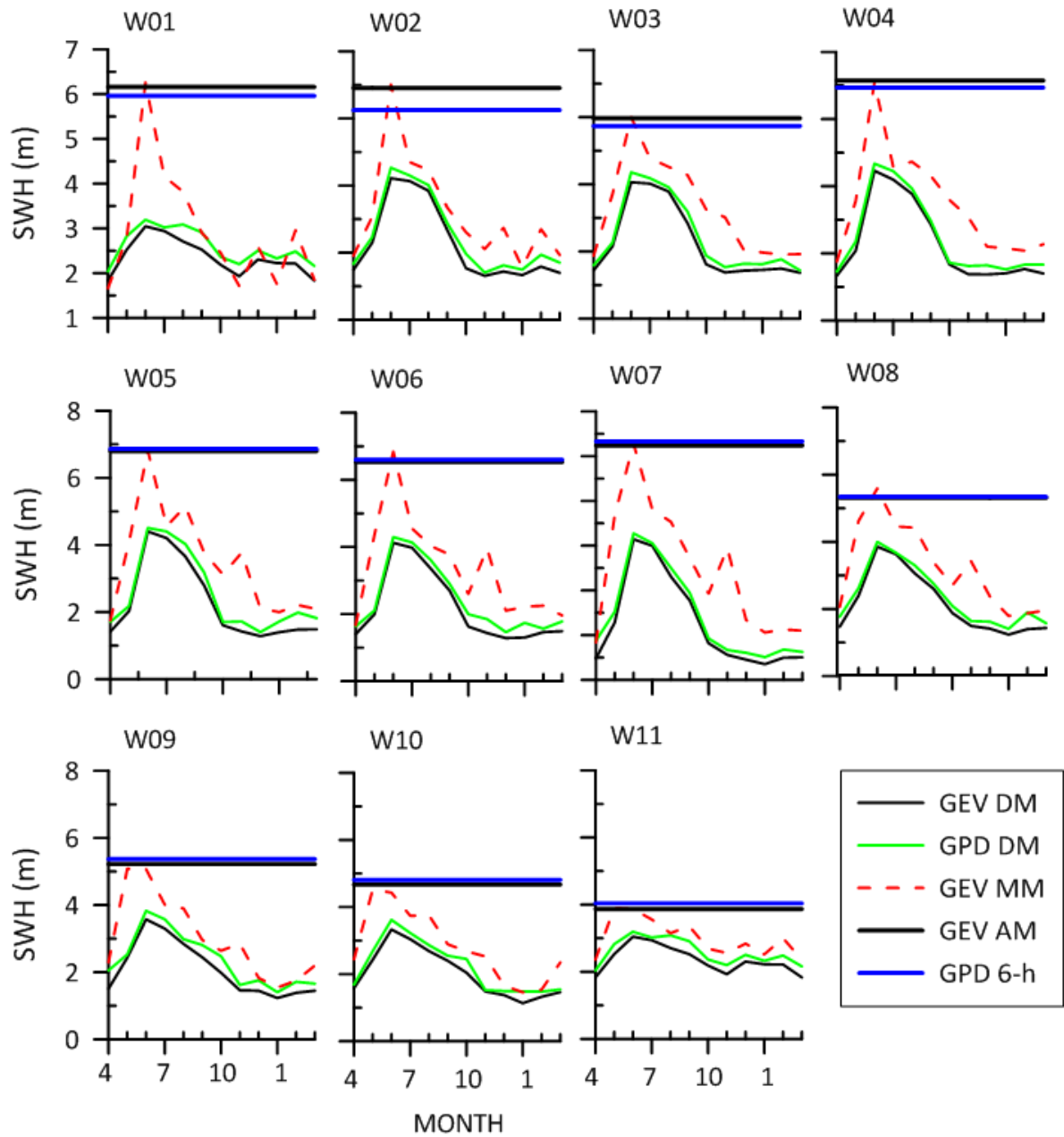


Figure 7. The 50-year return period significant wave height estimated using GEV (DM, MM and AM) and GPD (6-h and DM) by different data series for the locations along western shelf seas of India

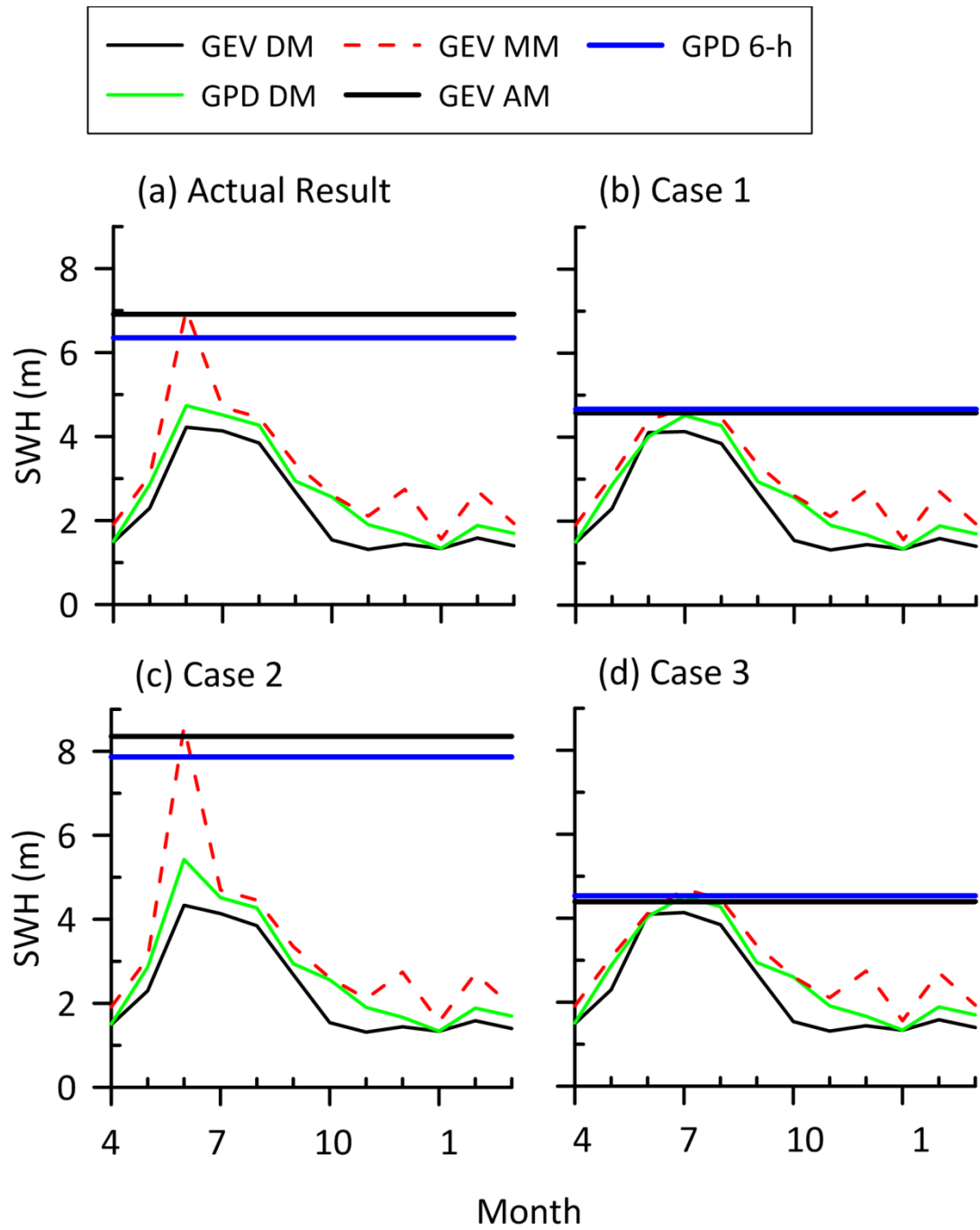


Figure 8. Estimated SWH values for 50-year return period obtained by different case studies with tropical cyclone Yemyin at the location W02 using GEV and GPD

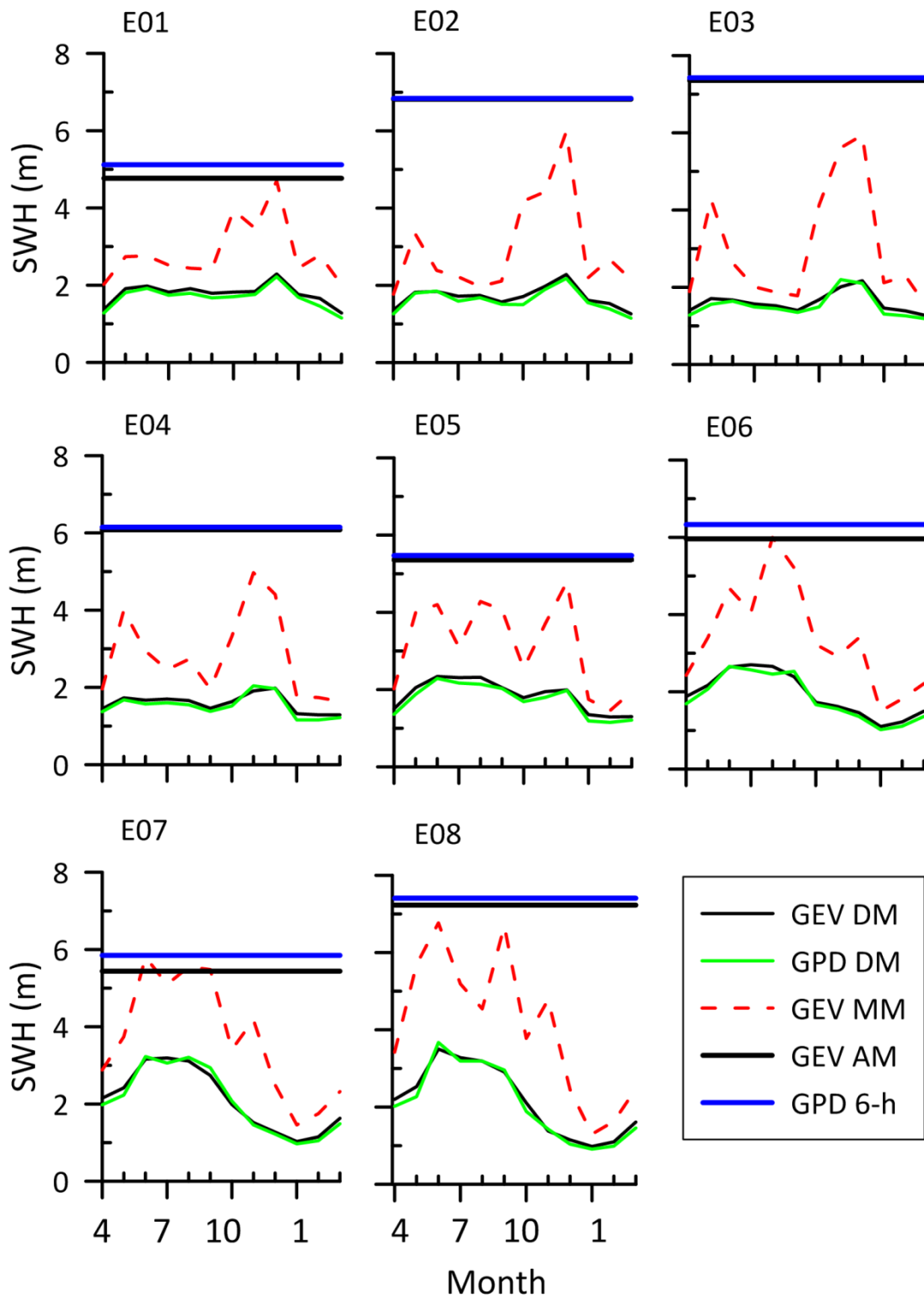


Figure 9. Estimated 50-year return period significant wave height using GEV (DM, AM, MM) and GPD (6-h and DM) by different data series (and 6-h) for the locations along eastern shelf seas of India

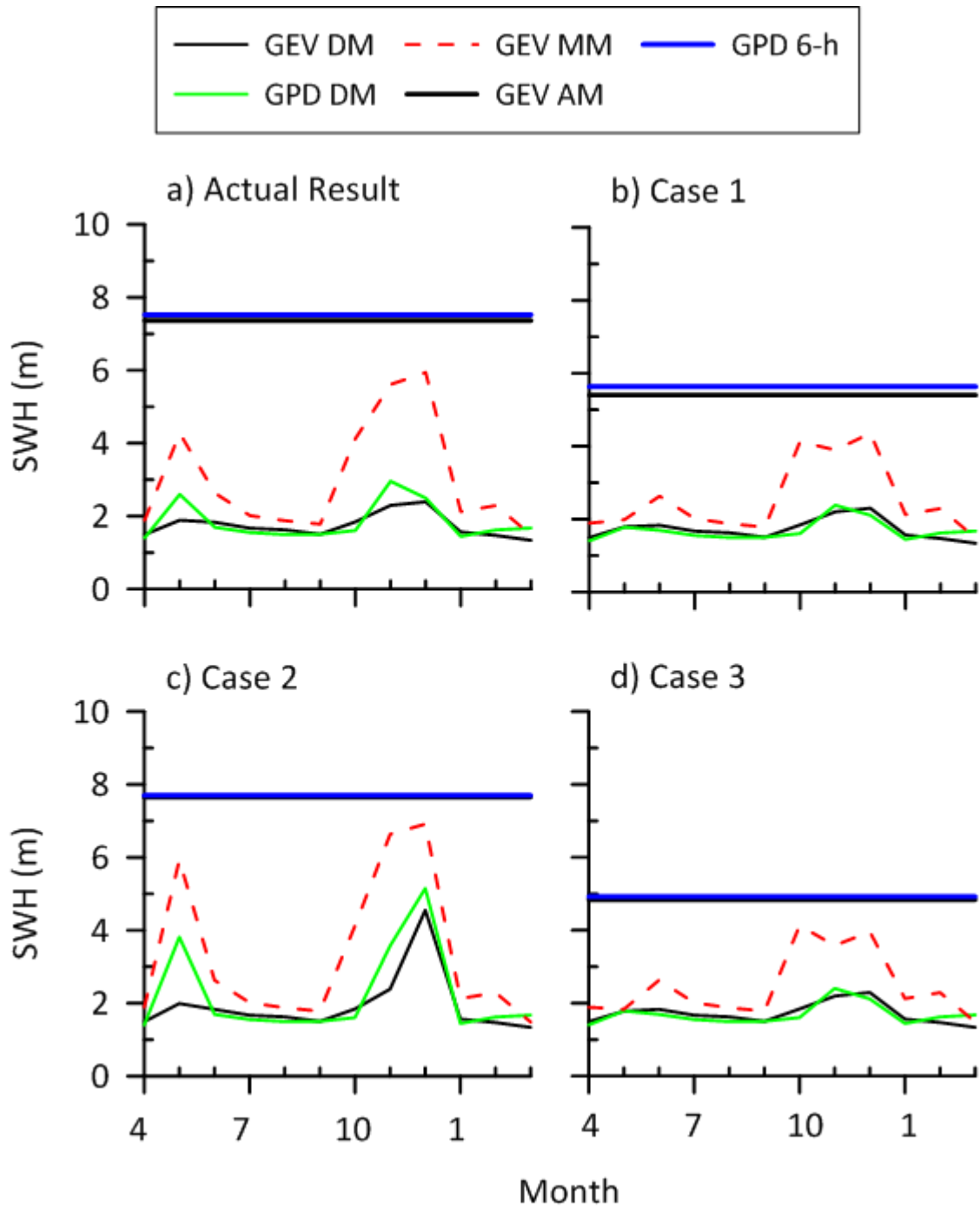


Figure 10. Estimated SWH for 50-year return period obtained for different case studies with tropical cyclones at location E03 using GEV and GPD

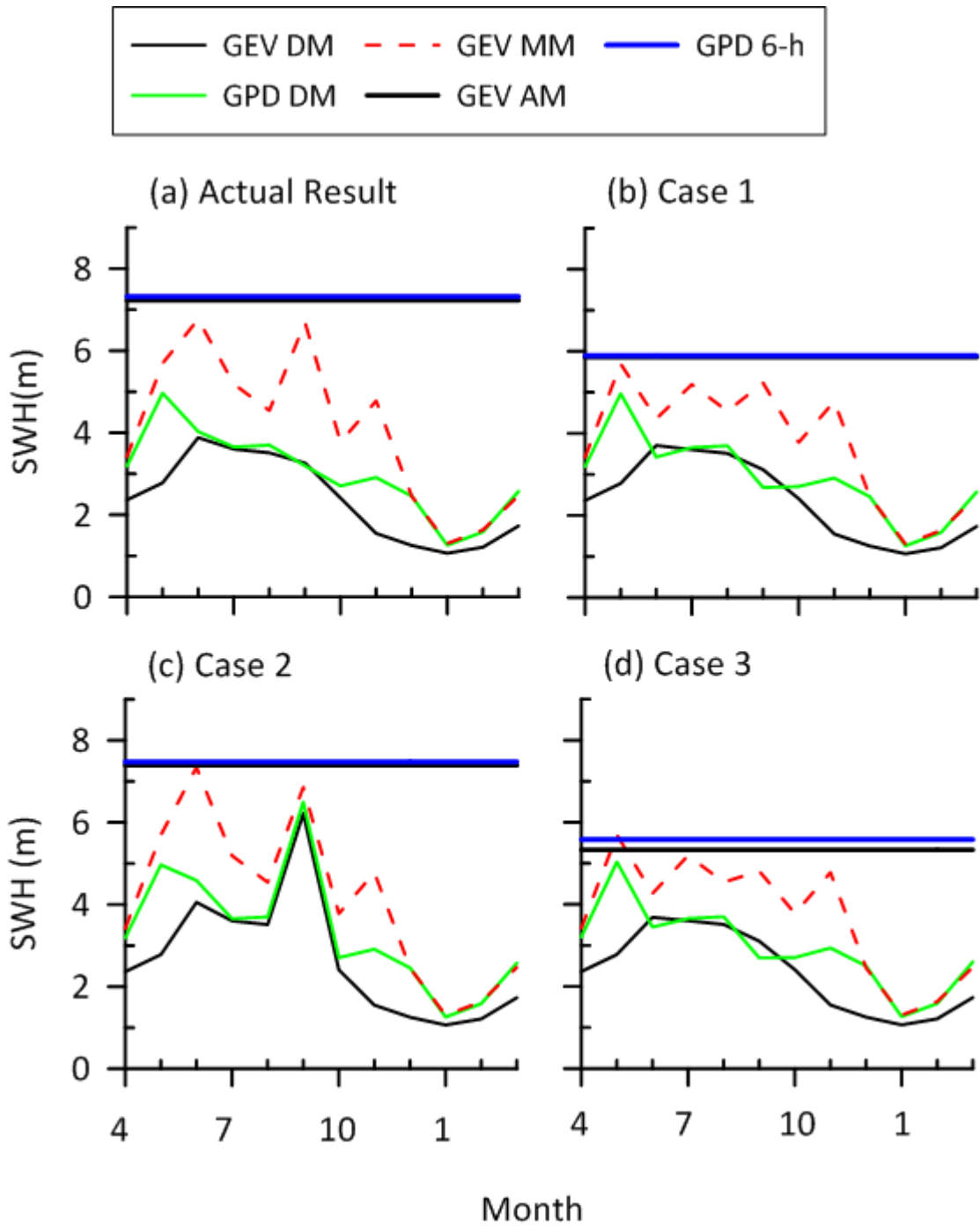


Figure 11. Estimated SWH for 50-year return period obtained by different case studies with tropical cyclones at location E08 using GEV and GPD