



# Variations in return value estimate of ocean surface waves – a study based on measured buoy data and ERA-Interim reanalysis data

T. Muhammed Naseef and V. Sanil Kumar

Ocean Engineering Division, Council of Scientific and Industrial Research (CSIR)–National Institute of Oceanography, Dona Paula 403 004, India

Correspondence to: V. Sanil Kumar (sanil@nio.org)

Received: 9 May 2017 – Discussion started: 19 June 2017  
Accepted: 12 September 2017 – Published: 19 October 2017

**Abstract.** An assessment of extreme wave characteristics during the design of marine facilities not only helps to ensure their safety but also assess the economic aspects. In this study, return levels of significant wave height ( $H_s$ ) for different periods are estimated using the generalized extreme value distribution (GEV) and generalized Pareto distribution (GPD) based on the Waverider buoy data spanning 8 years and the ERA-Interim reanalysis data spanning 38 years. The analysis is carried out for wind-sea, swell and total  $H_s$  separately for buoy data. Seasonality of the prevailing wave climate is also considered in the analysis to provide return levels for short-term activities in the location. The study shows that the initial distribution method (IDM) underestimates return levels compared to GPD. The maximum return levels estimated by the GPD corresponding to 100 years are 5.10 m for the monsoon season (JJAS), 2.66 m for the pre-monsoon season (FMAM) and 4.28 m for the post-monsoon season (ONDJ). The intercomparison of return levels by block maxima (annual, seasonal and monthly maxima) and the  $r$ -largest method for GEV theory shows that the maximum return level for 100 years is 7.20 m in the  $r$ -largest series followed by monthly maxima (6.02 m) and annual maxima (AM) (5.66 m) series. The analysis is also carried out to understand the sensitivity of the number of observations for the GEV annual maxima estimates. It indicates that the variations in the standard deviation of the series caused by changes in the number of observations are positively correlated with the return level estimates. The 100-year return level results of  $H_s$  using the GEV method are comparable for short-term (2008 to 2016) buoy data (4.18 m) and long-term (1979 to 2016) ERA-Interim shallow data (4.39 m). The 6 h interval data tend to miss high values of  $H_s$ , and hence there

is a significant difference in the 100-year return level  $H_s$  obtained using 6 h interval data compared to data at 0.5 h interval. The study shows that a single storm can cause a large difference in the 100-year  $H_s$  value.

## 1 Introduction

Coastal zones are relatively dynamic compared to the rest of the regions due to numerous natural as well as anthropogenic activities. Events such as extreme waves, storm surges and coastal flooding cause large catastrophes in the coastal region. The long-term (climate) behavior of sea state variables can be studied using non-stationary multivariate models that represent the time dependence of the variables (Solari and Losada, 2011). Various marine activities such as the design of coastal and offshore facilities, planning of harbor operations and ship design require detailed assessment of wave characteristics with certain return periods (Caires and Sterl, 2005; Menéndez et al., 2009; Goda et al., 2010). Generally, extreme value theory (EVT) is used for the determination of return levels by adopting a statistical analysis of historic time series of wave heights obtained from various sources such as in situ buoy measurements (e.g., Soares and Scotto, 2004; Méndez et al., 2008; Viselli et al., 2015), satellite data (e.g., Alves et al., 2003; Izaguirre et al., 2010), and hindcasted or reanalysis data by numerical models (e.g., Goda et al., 1993; Caires and Sterl, 2005; Teena et al., 2012; Jonathan et al., 2014). EVT consists of two types of distributions, viz. the generalized extreme value (GEV) distribution family which includes the Gumbel, Fréchet and Weibull distributions (Gumbel, 1958; Katz et al., 2002) and generalized Pareto distribution (GPD)

which incorporates the peak over threshold (POT) approach (Pickands, 1975; Coles et al., 2001).

GEV distribution by annual maxima (AM) observations (Goda, 1992) is one of the widely used methods in the EVT analysis. The main difficulty with using this method is the unavailability of reliable observations at a location of interest. To overcome the data scarcity, two different alternatives have been used by various authors: (i) the initial distribution method (IDM), in which all the data are used (Alves and Young, 2003); and (ii) the  $r$ -largest approach (Smith, 1986), where a number of the largest observations from a block period are considered rather than one observation as used in the AM method. The POT method (Abild et al., 1992) provides a good number of observations available for the analysis. Although there have been various proposals to automate threshold selection, threshold estimation for the application of the POT method to a single sample is still not resolved (Solari and Losada, 2012; Solari et al., 2017). GPD is another class of distribution introduced by Pickands (1975) and has been used by several authors such as Caires and Sterl (2005) and Thevasiyani et al. (2014). Teena et al. (2012) and Samayam et al. (2017) have carried out the EVT analysis of ocean surface waves in the northern Indian Ocean based on wave hind-cast data and ERA-Interim reanalysis data.

The most reliable source of ocean wave data is buoy measurements, and it can be used for EVT analysis (Panchang et al., 1999). In this paper, data from a directional Waverider buoy located in the central western shelf of India are used. Seasonality is one of the important aspects of climate data, and, therefore, it should be incorporated in the EVT analysis of waves, especially in a region such as the Arabian Sea. Seasonal analysis of the extremes helps in the planning of short-term marine activities such as offshore explorations and maintenance of coastal facilities. In the present paper, the EVT analysis is carried out by following both the GEV and GPD methods considering wind-sea, swell and total significant wave height ( $H_s$ ) separately. The IDM and POT methods are used for total wave height analysis, and block maxima (annual and monthly maxima) and the  $r$ -largest method are used in wind-sea and swell height analysis. Since the measured buoy data are for a short period of 8 years, the ERA-Interim reanalysis data from 1979 to 2016 are also used for comparing the  $H_s$  value with the 100-year return period.

The paper is organized as follows. Section 2 deals with the data and methodology used in the analysis. It also presents the threshold selection adopted in the study, and Sect. 3 explains the results obtained in the analysis, categorized into seasons using total  $H_s$  data, and comparison of return level estimation by different GEV approaches using wind-sea and swell height data. A case study is also included in the section for realizing the uncertainty related to observations in the AM approach when a limited number of observations are available. The influence of length of wave data on the estimated  $H_s$  return value is also covered under this section. Section 4 provides the concluding remarks.

## 2 Data and methodology

### 2.1 Data

Data used in the analysis are from a Datawell directional Waverider buoy deployed off Honnavar (14.304° N; 74.391° E) at a water depth of 9 m. The half-hourly sampled data cover the period from March 2008 to February 2016. The waves at the location show strong intra-annual variations due to the prevailing wind system during monsoon and non-monsoon seasons (Sanil Kumar et al., 2014). To understand the local and remote influences on the design wave characteristics, we analyzed  $H_s$  of wind-sea, swell and total waves separately. A season-wise study is also carried out since it will provide insight into the design wave heights for short-term coastal activities.

The  $H_s$  data from ERA-Interim (Dee et al., 2011), the global atmospheric reanalysis product of the European Centre for Medium-Range Weather Forecast (ECMWF), from 1979 to 2016 (38 years) are also used to evaluate the 100- and 50-year return period wave height in the shallow (water depth  $\sim 20$  m) and deep water. The shallow region is close to the buoy location, and the deep water location is at a water depth of  $\sim 4000$  m (Table 1). The ERA-Interim reanalysis used in the study has a spatial resolution of  $0.125^\circ \times 0.125^\circ$  and a temporal resolution of 6 h.

### 2.2 Methodology

EVT analysis is carried out by following the GEV distribution model and the POT method in which exceedance over a reliable threshold wave height can be fit into GPD. In the POT method, distribution of excess,  $x$ , over a threshold  $u$  is defined as

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

where  $y = x - u$ . Pickands (1975) shows that the distribution function of excess,  $F_u(y)$ , for a sufficiently high threshold  $u$  converges to GPD, having a cumulative distribution function (CDF) as follows:

$$G(x; k, \alpha, \beta) = \left\{ 1 - \left( 1 - k \frac{x - \beta}{\alpha} \right)^{1/k} \right\} \quad k \neq 0 \quad (2)$$

$$= 1 - e^{-(x-\beta)/\alpha} \quad k = 0. \quad (3)$$

GEV has a CDF as follows:

$$F(X) = \exp \left\{ - \left( 1 - k \left( \frac{X - \beta}{\alpha} \right)^{1/k} \right) \right\} \quad k \neq 0 \quad (4)$$

$$= \exp \left\{ - \exp \left( - \frac{X - \beta}{\alpha} \right) \right\} \quad k = 0, \quad (5)$$

where  $\alpha$  is scale parameter in the range of  $\alpha > 0$ ,  $\beta$  is the location parameter with possible values of  $-\infty < \beta < \infty$  and

**Table 1.** The comparison of 50- and 100-year significant wave height return levels based on buoy, ERA-Interim shallow and ERA-Interim deep data at 6 h intervals along with data statistical parameters.

Distribution	Particulars	Buoy	ERA-Interim shallow	ERA-Interim deep	
	Location	74.391° E 14.304° N	74.380° E 14.250° N	69.250° E 14.250° N	
GEV	Total number of data points	11 479	55 520	55 520	
	Data period	2008–2016	1979–2016	1979–2016	
	No. data points between different ranges	> = 5 m	0	5	70
		> = 4.5 m & < 5 m	0	11	229
		> = 3.5 m & < 4.5 m	32	275	2846
		> = 3 m & < 3.5 m	137	1224	3263
	Data max (m)		4.11	5.45	7.13
	Second highest value (m)		4.03	5.37	6.09
	Data mean (m)		1.12	1.24	1.67
	Data SD (m)		0.73	0.70	0.88
	Return levels for 50-year period (m)	3.88	3.52	4.58	
	Return levels for 100-year period (m)	4.18	4.39	5.67	
GPD	No. data points between different ranges after decluster	> = 5 m	0	2	15
		> = 4.5 m & < 5 m	0	4	42
		> = 3.5 m & < 4.5 m	9	46	253
		> = 3 m & < 3.5 m	23	133	235
	Threshold (m)		3.19	3.50	4.41
	Return levels for 50-year period (m)		4.36	5.55	6.69
	Return levels for 100-year period (m)		4.46	6.18	7.28

$k$  is the shape parameter in the range of  $-\infty < k < \infty$ . GPD can be further categorized into three distributions based on its tail features. When  $k = 0$ , GPD corresponds to an exponential distribution (medium-tailed or Pareto type I) with mean  $\alpha$ ; when  $k > 0$ , GPD is short-tailed, also known as Pareto type II; when  $k < 0$ , the distribution takes the form of ordinary Pareto distribution, having a long-tailed distribution (also known as Pareto type III). Parameter estimation and statistical distribution fitting are carried out by using the WAFO toolbox (Brodtkorb et al., 2000), developed by Lund University, Sweden.

The analysis is carried out by using the wind-sea, swell and total  $H_s$  data covering  $\sim 8$  years (2008–2016). From the measured data, to separate the wind seas and swells, the method proposed by Portilla et al. (2009) is used. The separation algorithm is based on the assumption that the energy at the peak frequency of a swell cannot be higher than the value of a Pierson–Moskowitz (PM) spectrum with the same frequency. If the ratio between the peak energy of a wave system and the energy of a PM spectrum at the same frequency is above a threshold value of 1, the system is considered to represent wind sea – otherwise it is taken to be a swell. A separation frequency  $f_c$  is estimated following Portilla et al. (2009), and the swell and wind-sea parameters are obtained for frequencies ranging from 0.025 Hz to  $f_c$  and from  $f_c$  to 0.58 Hz respectively. The GPD method is used for seasonal analysis of different period data series. The

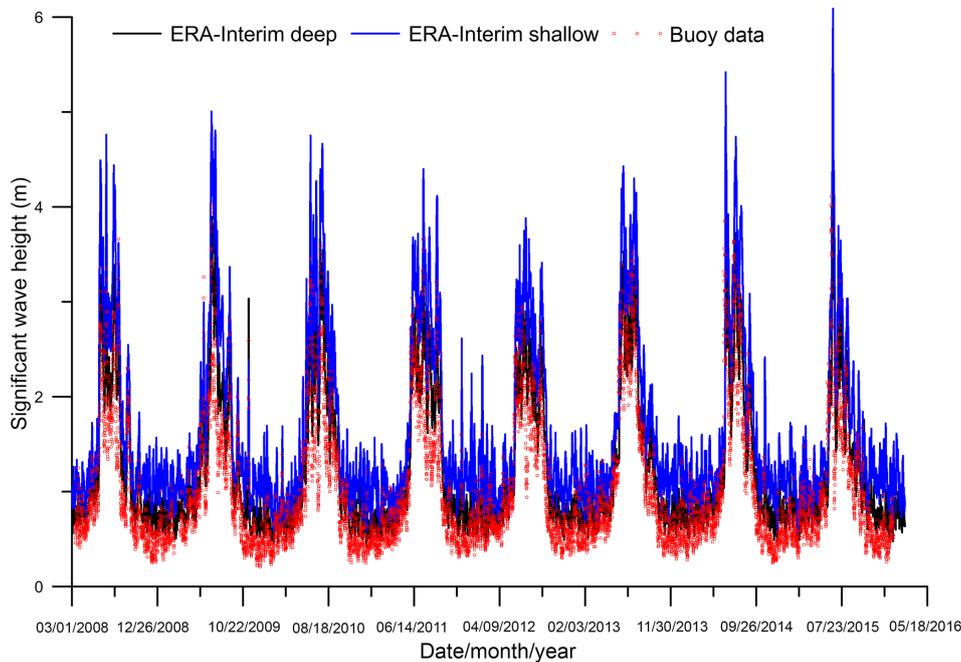
GEV method is used for intercomparison of return level estimation among wind-sea, swell and resultant data sets by extracting different block maxima series: (i) seasonal maxima, which contain the highest observations from each season; (ii) monthly maxima, which contain one highest observation from each month; and (iii) annual maxima. The parameters are estimated using the probability-weighted moment (PWM) method since the data set duration is very limited, and the PWM method holds good results compared to other methods such as the maximum likelihood method (Hosking et al., 1985).

To study the uncertainties related to the length of the observation, we extracted 3, 6, 12 and 24 h data series from the half-hourly original data and carried out EVT analysis. Since the wave climate in the study location is strongly characterized by the prevailing seasonal behavior of wind system, we took further consideration of uncertainties related to a seasonal aspect of wave climate by extracting three seasonal data sets, viz. pre-monsoon (FMAM), monsoon (JJAS) and post-monsoon (ONDJ) seasons.

The major drawback of EVT analysis using the block maxima method, especially the annual maxima, is that it does not consider the significant amount of observations which are closely related to storm features of the data set. Those omissions of observations would cause significant variations in the final results of EVT analysis, especially in the cases where EVT analysis is performed for a very limited data set.

**Table 2.** Different goodness of fit tests used for selecting threshold values of POT analysis.  $H = 0$  indicates that the test does not reject the hypothesis at the 5% significance level (i.e.,  $p$  value  $> 0.05$  or test statistics is less than critical value), and  $H = 1$  indicates that the hypothesis is rejected. KS test represents the Kolmogorov–Smirnov test and CM test represents the Cramér–von Mises test.

Seasons	Time interval	$H_{s-max}$ (m)	Threshold (m)	KS test				CM test			
				$p$ value	Test statistics	Critical value	$H$	$p$ value	Test statistics	Critical value	$H$
Full year	0.5 h	4.70	3.31	0.332	0.167	0.242	0	0.320	0.178	0.459	0
	3 h	4.28	3.31	0.920	0.114	0.287	0	0.808	0.062	0.458	0
	6 h	4.11	3.19	0.402	0.183	0.281	0	0.490	0.122	0.458	0
	12 h	4.11	2.72	0.745	0.092	0.187	0	0.595	0.098	0.460	0
	24 h	4.00	2.74	0.525	0.126	0.213	0	0.739	0.072	0.459	0
FMAM	0.5 h	1.94	1.32	0.952	0.081	0.218	0	0.985	0.027	0.459	0
	3 h	1.88	1.19	0.258	0.126	0.170	0	0.222	0.226	0.460	0
	6 h	1.83	1.19	0.203	0.151	0.192	0	0.210	0.234	0.460	0
	12 h	1.83	1.19	0.447	0.143	0.227	0	0.446	0.134	0.459	0
	24 h	1.83	1.19	0.296	0.210	0.294	0	0.423	0.142	0.458	0
JJAS	0.5 h	4.70	3.49	0.562	0.158	0.275	0	0.665	0.085	0.458	0
	3 h	4.28	3.36	0.722	0.141	0.281	0	0.657	0.087	0.458	0
	6 h	4.11	2.94	0.766	0.084	0.174	0	0.758	0.069	0.460	0
	12 h	4.11	3.20	0.890	0.117	0.281	0	0.906	0.046	0.458	0
	24 h	4.00	2.78	0.961	0.070	0.194	0	0.990	0.024	0.460	0
ONDJ	0.5 h	2.81	1.06	0.131	0.123	0.144	0	0.193	0.247	0.460	0
	3 h	2.61	1.00	0.247	0.106	0.142	0	0.307	0.183	0.460	0
	6 h	2.59	0.98	0.488	0.092	0.151	0	0.451	0.133	0.460	0
	12 h	2.18	0.84	0.197	0.102	0.129	0	0.350	0.166	0.461	0
	24 h	2.18	0.87	0.195	0.155	0.196	0	0.207	0.237	0.460	0

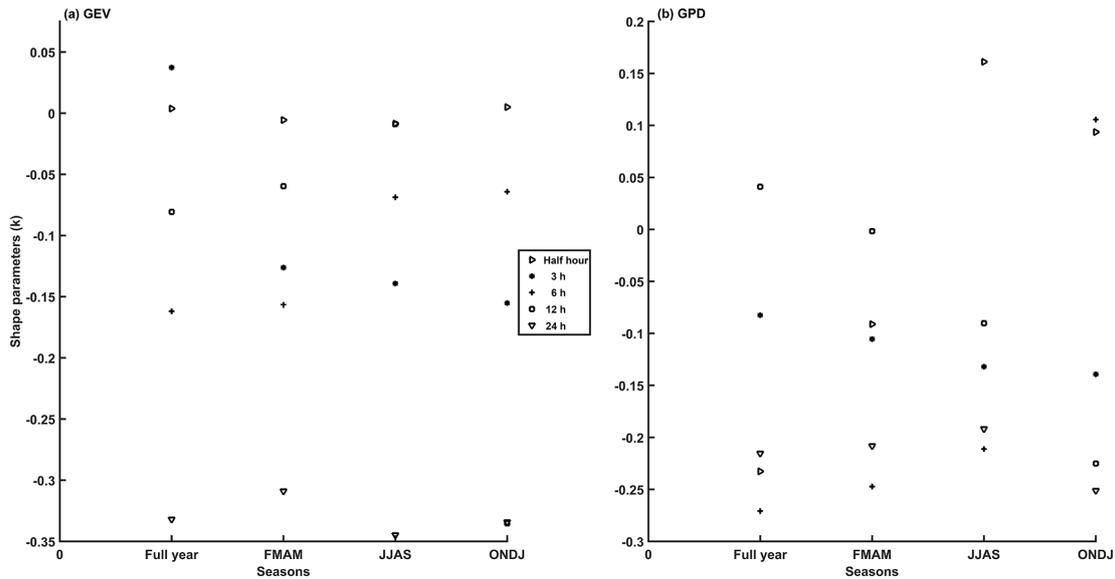


**Figure 1.** Time series plot of the significant wave height measured by buoy and from ERA-Interim data at shallow and deep water.

EVT is based on the assumption that the observations under consideration are independent and identically distributed (Coles et al., 2001). We can expect identical status of ocean wave observations for a large extent. Since the POT approach resamples the data over a threshold value, making identical and independent observations is a tedious task. A suitable

combination of threshold and minimum separation time between the resampled observations must be taken into account to establish independence among the observations.

The average duration of tropical storms in the Arabian Sea is 2–3 days (Shaji et al., 2014). Therefore, in the present analysis, we fixed a minimum of 48 h of separation time in



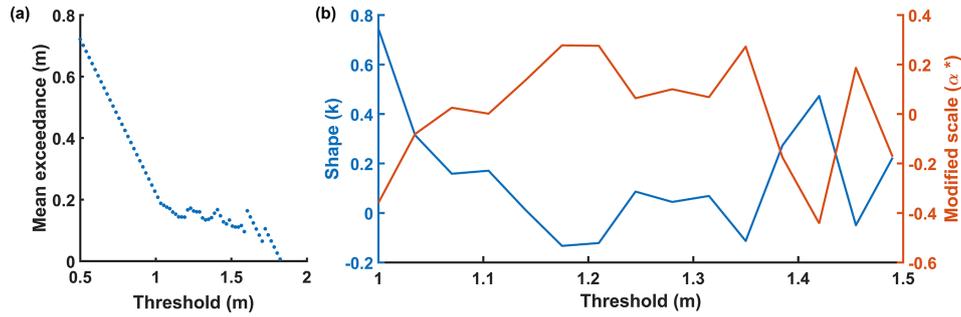
**Figure 2.** Estimated shape parameters for different seasonal data with different sampling intervals used in the (a) GEV and (b) GPD model.

**Table 3.** Table showing different parameters and corresponding RMSEs of data and the estimated CDF used during each data series analysis.

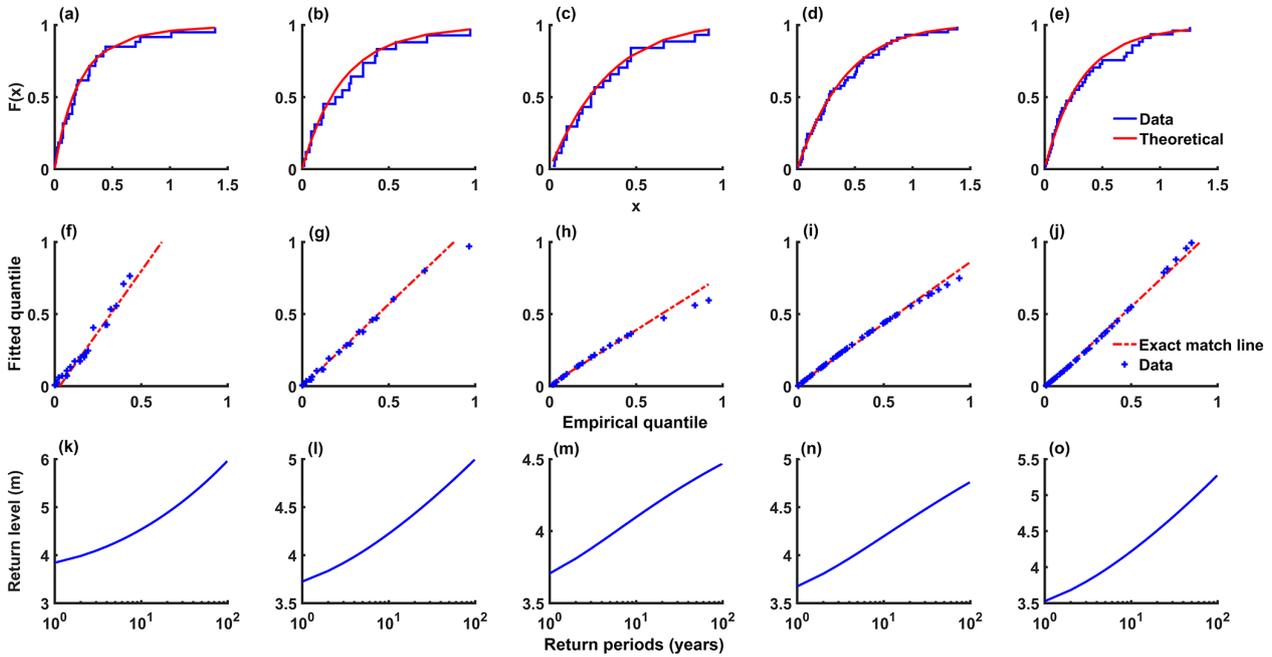
Seasons	Data	GPD			GEV			
		$k$ (m)	$\alpha$ (m)	RMSE (m)	$k$ (m)	$\alpha$ (m)	$\beta$ (m)	RMSE (m)
Full year	0.5 h	-0.233	0.208	0.222	0.004	0.416	2.462	0.023
	3 h	-0.091	0.232	0.028	-0.006	0.381	2.439	0.008
	6 h	0.161	0.346	0.110	-0.008	0.418	2.223	0.004
	12 h	0.094	0.420	0.102	0.005	0.416	2.206	0.020
	24 h	-0.082	0.314	0.071	0.037	0.458	2.015	0.060
FMAM	0.5 h	-0.105	0.130	0.037	-0.126	0.115	1.134	0.090
	3 h	-0.132	0.125	0.078	-0.139	0.104	1.143	0.098
	6 h	-0.139	0.123	0.077	-0.155	0.099	1.147	0.100
	12 h	-0.271	0.095	0.167	-0.162	0.108	0.998	0.125
	24 h	-0.247	0.099	0.082	-0.157	0.114	0.872	0.142
JJAS	0.5 h	-0.184	0.216	0.124	-0.069	0.298	2.782	0.088
	3 h	-0.046	0.280	0.068	-0.069	0.274	2.786	0.074
	6 h	0.041	0.328	0.051	-0.081	0.288	2.583	0.118
	12 h	-0.002	0.265	0.042	-0.060	0.281	2.598	0.065
	24 h	-0.090	0.267	0.083	-0.009	0.312	2.423	0.007
ONDJ	0.5 h	-0.225	0.189	0.393	-0.335	0.117	1.023	0.631
	3 h	-0.215	0.178	0.333	-0.332	0.116	1.025	0.533
	6 h	-0.208	0.177	0.284	-0.309	0.114	0.912	0.525
	12 h	-0.192	0.167	0.267	-0.345	0.104	0.911	0.523
	24 h	-0.251	0.183	0.315	-0.334	0.111	0.780	0.498

between two consecutive storm peaks to ensure the independence of the data points for the analysis. Then, we selected a tentative threshold value in such a way as to ensure the presence of at least 15 peak values per year on average. This resulted in at least 120 data points in each sub-data sets used for the seasonal analysis. The resulting data series are used in further POT analysis. Further adjustment of the threshold

is carried out by sample mean excess (SME) plots and parameter stability plots (PS plot). From these plots, we selected four probable thresholds and fitted the corresponding GPD. A final threshold value is chosen by analyzing results obtained in different goodness of fit (GOF) tests such as the Kolmogorov–Smirnov (KS) test, Anderson–Darling



**Figure 3.** Typical (a) SME and (b) PS plots used for selecting a range of thresholds required for POT analysis. In this particular case, a range of 1.19 to 1.32 m was selected.



**Figure 4.** Figure corresponding to the full-year analysis. (a) to (e) represent the CDF plots for half, 3, 6, 12 and 24-hourly data respectively; (f) to (j) correspond to  $Q-Q$  plots; (k) to (o) correspond to return levels estimated using the GPD model.

(AD) test and Cramér–von Mises (CM) test (Stephens, 1974; Choulakian and Stephens, 2001).

The distributions used in the analysis are validated using graphical tools such as quantile–quantile ( $Q-Q$ ) plots and CDF plots. In addition to above graphical tools, we checked the reliability of the chosen thresholds for the POT method by using different GOF tests such as KS, AD and CM tests (Table 2). A  $p$  value  $> 0.05$  indicates that the selected distribution does not show a significant difference from the original data within the 5% significance interval.

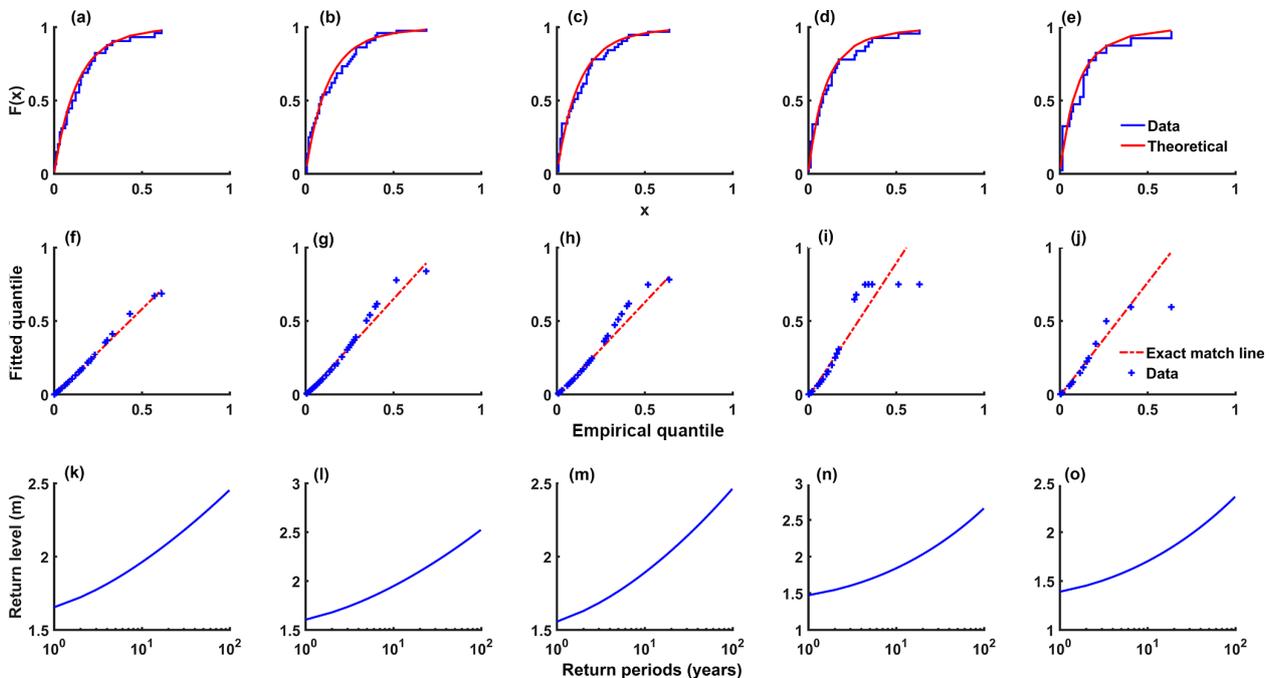
### 3 Results and discussion

#### 3.1 Long-term statistical analysis of total $H_s$

The mean wave climate at the study location is characterized by an annual mean  $H_s$  of 1.04 m. The maximum  $H_s$  of the data during 2008–2016 is 4.70 m, and the next highest  $H_s$  is 4.34 m (Fig. 1), whereas the highest wind-sea  $H_s$  and swell  $H_s$  are 4.29 and 4.28 m respectively. A statistical analysis of  $H_s$  was carried out by considering the seasonal characteristics of the wave climate. To study the seasonal aspects of the return level estimation, the data are grouped into three different seasonal series, viz. FMAM, JJAS and ONDJ seasons, in addition to full-year data. Since the study is located off the central west coast of India, the wave climate shows distinct variability throughout the year. Previous studies such as

**Table 4.** Estimated return values corresponding to different seasons using total wave height ( $H_s$ ) following the GEV and GPD methods. Here the GEV method follows the initial distribution approach.

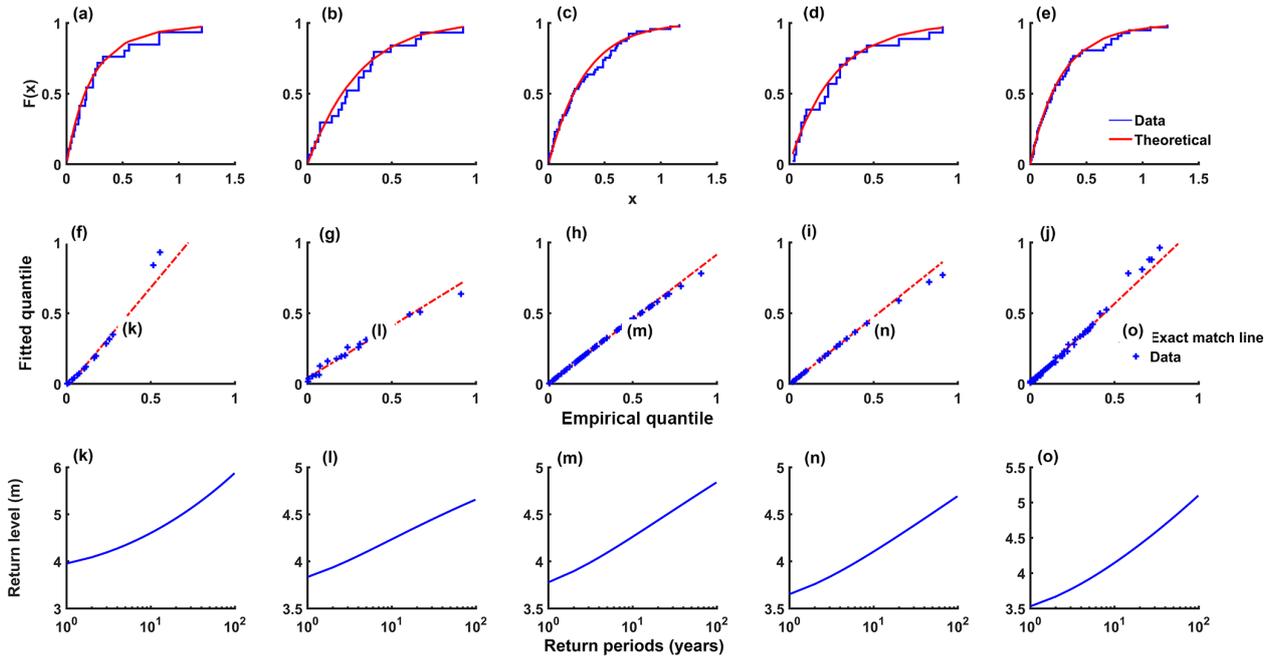
Seasons	Data	GPD			GEV		
		10 years (m)	50 years (m)	100 years (m)	10 years (m)	50 years (m)	100 years (m)
Full year	0.5 h	4.24	4.73	4.96	3.37	4.00	4.26
	3 h	4.01	4.36	4.50	3.28	3.88	4.13
	6 h	4.08	4.37	4.47	3.17	3.88	4.18
	12 h	4.17	4.59	4.76	3.14	3.81	4.10
	24 h	4.18	4.92	5.27	3.00	3.68	3.95
FMAM	0.5 h	1.94	2.29	2.45	1.43	1.71	1.85
	3 h	1.93	2.33	2.52	1.42	1.68	1.81
	6 h	1.87	2.26	2.46	1.41	1.68	1.81
	12 h	1.82	2.35	2.66	1.29	1.59	1.74
	24 h	1.68	2.12	2.36	1.18	1.48	1.64
JJAS	0.5 h	4.24	4.62	4.80	3.48	4.04	4.29
	3 h	4.02	4.21	4.27	3.42	3.91	4.12
	6 h	4.24	4.66	4.84	3.29	3.90	4.19
	12 h	4.08	4.51	4.69	3.27	3.83	4.09
	24 h	4.11	4.78	5.10	3.13	3.66	3.89
ONDJ	0.5 h	2.64	3.69	4.28	1.41	1.96	2.30
	3 h	2.46	3.41	3.93	1.41	1.95	2.28
	6 h	2.35	3.23	3.71	1.28	1.77	2.07
	12 h	2.22	3.03	3.47	1.27	1.77	2.09
	24 h	2.16	3.16	3.74	1.15	1.67	2.00



**Figure 5.** Same as in Fig. 4 but corresponding to the pre-monsoon season.

that of Anoop et al. (2015) reported that average  $H_s$  attains its peak at around 3 m during JJAS and that the FMAM season is relatively calm (0.5–1.5 m) compared to ONDJ (1.5–

2 m). The seasonal analysis is carried out using  $H_s$  data following both the GEV and GPD methods. Here, the initial distribution method is considered in the GEV method rather



**Figure 6.** Same as in Fig. 4 but corresponding to the monsoon season.

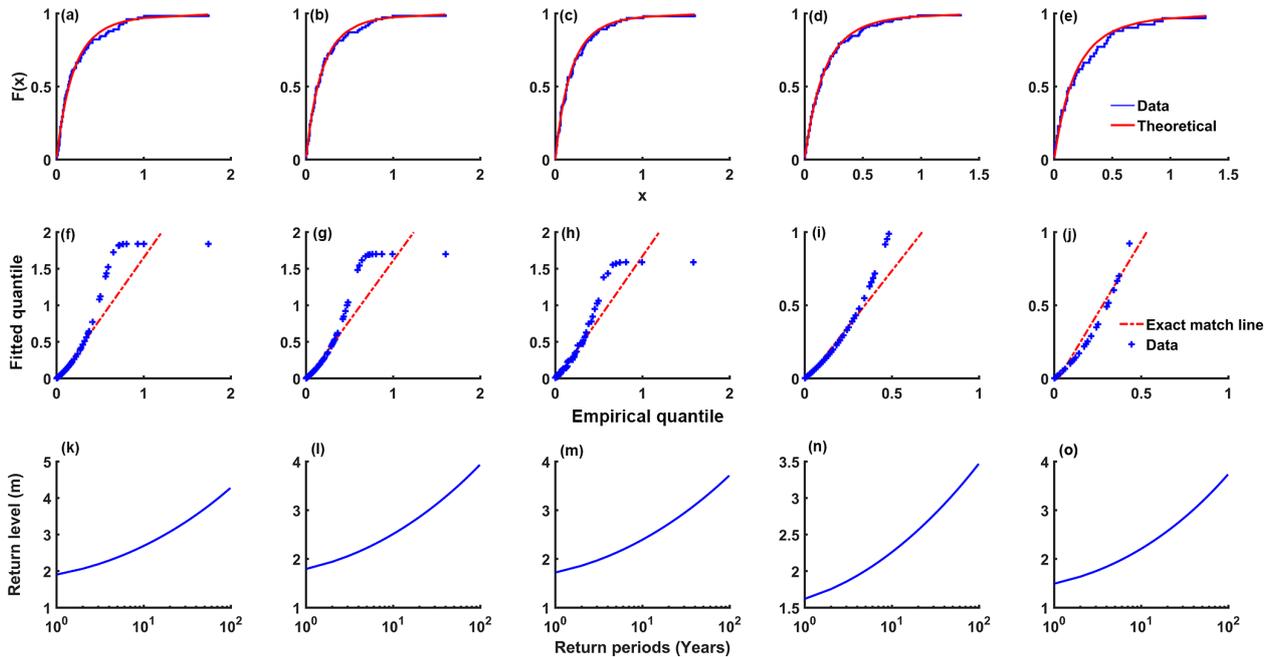
than block maxima (Mathiesen et al., 1994). One of the challenging tasks for GPD modeling is the selection of a suitable threshold value. The threshold should be high enough for observations to be independent, and data after POT analysis must have the necessary number of observations in order to converge the POT analysis into GPD. SME plots and PS plots are used to select a range of initial thresholds. Upon analyzing the resultant GPD fit for those thresholds, the final thresholds are chosen with the help of GOF tests, which are presented in Table 2. Figure 2 and Table 3 show the estimated parameters using the PWM method for both GEV and GPD. It is clear that shape parameters in both cases are negative, indicating that the models are a type III distribution for GPD and a Weibull distribution for GEV. Table 3 also shows the RMSE in the chosen model for each data series with estimated CDF. It is evident that the JJAS season has a lower RMSE ( $\sim 0.07$  m on average) when considering the GPD model, while, in the case of the GEV model, the full-year data series has a lower RMSE ( $\sim 0.02$  m on average). The ONDJ season shows a higher discrepancy in both cases, resulting in an average RMSE of 0.31 and 0.54 m for GPD and GEV respectively. Figure 3 shows the typical SME and PS plots used for choosing a range of thresholds before fixing the final threshold for POT analysis on each series. In this particular case (6 h data series of FMAM season), a range of thresholds from 1.10 to 1.32 m was selected, and the final threshold of 1.19 m was fixed for analyzing the GOF test results (Table 2).

### 3.1.1 Full year

Here, we considered full-year data series without dealing with seasonality, and both the GEV and GPD are used in the analysis. Initially, a range of thresholds from 2.5 to 3.4 m was selected, and further adjustment of the threshold is carried out by analyzing the GOF test results. Table 2 shows the selected thresholds and the corresponding GOF test results for each series in the full-year data analysis. It is clear that the selected thresholds are in good agreement with the GOF test results. Both the KS test and CM test give a  $p$  value  $> 0.32$ . Moreover, both CDF plots and  $Q-Q$  plots (see Fig. 4, first and second rows respectively) show that the selected GPD models exhibited a good performance for the particular POT series. After acquiring the best fit model, return levels (Table 4) were estimated for 10, 50 and 100 years. The GPD model estimates a 10-year return level smaller than that of the maximum measured total  $H_s$  value by 5 to 15%. An underestimation of 10 to 25% from the maximum measured value was reported by Samayam et al. (2017) compared to the 36- and 30-year return levels based on ERA-Interim re-analysis data for deep waters around the Indian mainland. The initial distribution approach underestimates the return levels in such a way that even the 100-year return level does not cross the highest observation (4.70 m) in the data and the largest 100-year return level is reported as 4.26 m when dealing with half-hourly data series (Table 4). The large number of observations having a very low  $H_s$  in the data series used in the analysis leads to the underestimation in the initial distribution method, whereas the GPD model estimated 4.73 and 4.96 m as the 50- and 100-year return levels respec-

**Table 5.** Return levels estimated by the GEV model using total, wind-sea and swell data for different block maxima series.

Data	Total $H_s$ (m)			Wind-sea $H_s$ (m)			Swell $H_s$ (m)		
	10 years	50 years	100 years	10 years	50 years	100 years	10 years	50 years	100 years
Monthly maxima	3.21	5.28	6.02	2.45	3.43	3.88	2.92	4.77	5.72
Two maximum values from each season	3.66	5.58	6.56	2.68	3.78	4.29	3.31	5.07	5.95
One maximum value from each season	3.85	6.04	7.20	2.91	4.32	5.06	3.51	5.40	6.35
Annual maxima	4.50	5.28	5.66	3.27	4.86	6.16	3.97	4.83	5.35



**Figure 7.** Same as in Fig. 4 but corresponding to the post-monsoon season.

tively. When considering different time interval data, both 12 and 24 h data series estimate lower return levels compared to other series by the GEV model. It is evident that there are uncertainties related to the sampling interval adopted for the return value estimation. The standard deviation for GPD estimation when considering different time intervals is 0.57 m, which is highest among the other seasonal data. GEV estimation reports an even lower spread of return levels with 0.16 m standard deviation.

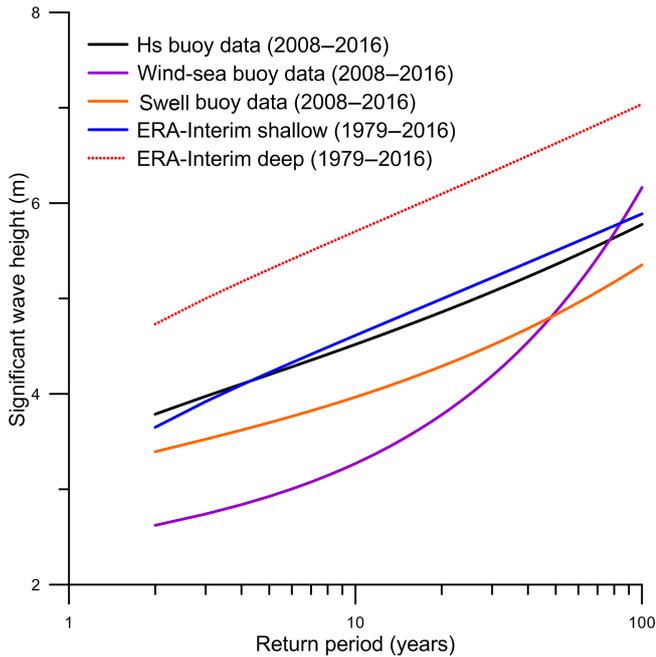
**3.1.2 Pre-monsoon season**

The data from February to May constitute the pre-monsoon data set. Pre-monsoon is the calmest season in the study location, with a maximum and an average  $H_s$  of around 1.94 and 0.73 m respectively. Using SME and PS plots, a range of thresholds from 1.19 to 1.32 m is selected for each time series and fitted to the corresponding GPD by using the resultant POT values. The final threshold selected by the help of GOF tests is presented in Table 2. KS and CM tests give a  $p$  value of more than 0.43 and 0.45 respectively on average (Table 2). Since the  $p$  values are more than 0.05, the chosen

POT is not significantly different from the time series data. CDF plots and  $Q-Q$  plots (Fig. 5) for the different data series of the season illustrate the reliability of the chosen model. Return levels for different return periods using a particular GPD are presented in Table 4. GEV estimation exhibits the same characteristics of underestimation as shown in the full-year analysis. Average 100-year return levels estimation using different time interval data using the GEV model attained a value of only 1.77 m, which is less than the highest observed data point in the season, whereas GPD reports an average 100-year return level of 2.49 m. Time interval analysis for the season exhibits the least discrepancies among the return level estimations compared to other seasons. Standard deviations of 0.11 and 0.08 m for GPD and GEV estimations respectively were observed for 100-year return levels considering different time series data.

**3.1.3 Monsoon season**

The monsoon season data set covers observations from June to September, and this season is characterized by rough wave climate at the study location.  $H_s$  values of 4.70 and 1.77 m



**Figure 8.** Return levels of significant wave heights for different return periods based on buoy data (2008–2016), ERA-Interim shallow water data and ERA-Interim deep water data (1979–2016) at 6 h intervals by the GEV model using annual maxima series.

are recorded for the maximum and average respectively during the season. A range of thresholds (2.78 to 3.49 m) is selected for preliminary GPD fitting as a result of interpreting SME and PS plots of each data series, and the corresponding final thresholds were selected after clarifying with the GOF test results (Table 2). Both KS and CM tests report a  $p$  value  $> 0.56$ , indicating that the resulting POT series for the selected threshold converges into GPD. CDF and  $Q-Q$  plots in Fig. 6 shows the reliability of the adopted threshold value. Return levels for the distinct return period were estimated using the resultant POT series. Table 4 provides 10-, 50- and 100-year return period values estimated using GPD and GEV models. For half-hourly data, GPD projects a value of 4.80 m for the 100-year return level, whereas GEV underestimates it, with a value of 4.29 m. The GPD model shows a 0.36 m standard deviation among the return levels for different time interval data. Both the 12 and 24 h series gave lower return levels compared to other series.

### 3.1.4 Post-monsoon season

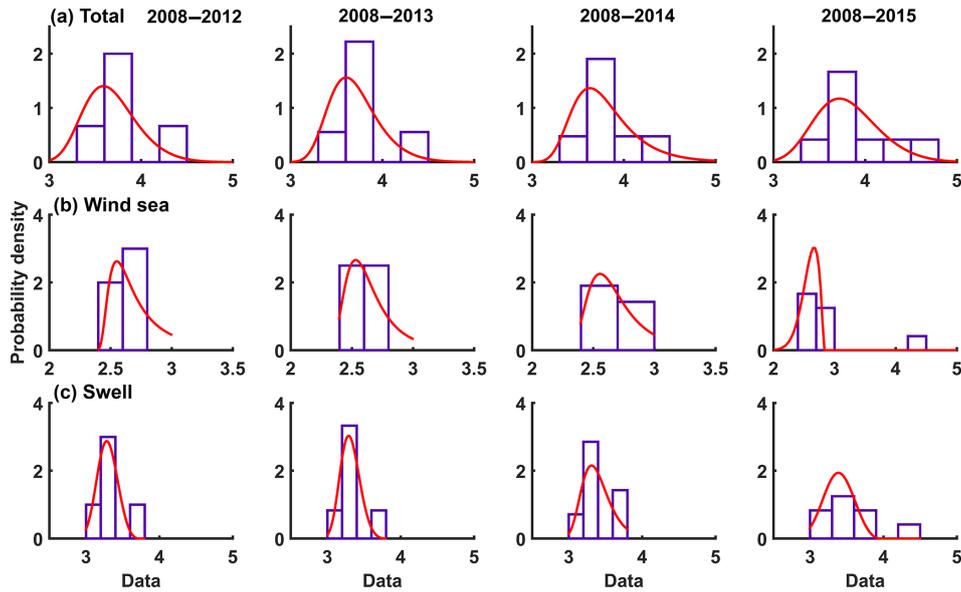
The post-monsoon season constitutes data from the October to January months of the year, and the observed maximum  $H_s$  in this season is 2.41 m. The majority of observations during this season lie below the average value of  $H_s$ . Only 32% of the observations lie above 1.13 m, and 8% of the data are above 1.5 m. Hence, selecting the best threshold for the season was more difficult. GPD was fitted for a range of

thresholds (0.7 to 1.3 m) selected from SME and PS plots corresponding to each series. Most suitable thresholds were selected after checking the goodness of fit of GPD (Table 2). The GOF test results show that the ONDJ series holds maximum uncertainties on threshold selection due to lower  $p$  values for the KS test ranges from 0.13 to 0.48 and from 0.19 to 0.45 for the CM test. Figure 7 shows the CDF and  $Q-Q$  plots. The GEV and GPD estimations for the post-monsoon season show very large difference among return levels (Table 4). The average percentage difference between the 100-year return values obtained from GEV and GPD estimations is  $\sim 60\%$ . This shows that the GEV model clearly underperforms during the ONDJ season, when the initial distribution methods were adopted. The highest return level reported by the GPD model is 4.28 m, whereas GEV estimated about 2.3 m for the season. The ONDJ season has a standard deviation of 0.30 and 0.13 m for the GPD and GEV estimation respectively while using different sampling intervals.

### 3.2 Long-term statistical analysis of wind seas and swells

In this section, we relied on the GEV method based on block maxima. For that purpose, we extracted total, wind-sea and swell  $H_s$  data into different block maxima, viz. monthly, seasonal and annual maxima series. Two seasonal maxima series are considered in such a way that one includes the highest two observations in a season and another one consist of the highest observation from each season. Therefore, the monthly maxima series includes 96 data points. Both seasonal maxima series (seasonal maxima 1 and 2) consist of 24 and 48 data points respectively. The annual maxima series covers eight data points. Table 5 shows the estimated return levels corresponding to various return periods. It is clear that both seasonal maxima series provide the highest return levels for total  $H_s$  (6.56 and 7.20 m) and swell  $H_s$  (5.95 and 6.35 m), whereas wind-sea  $H_s$  is 6.16 m when the annual maxima series is considered. The GEV-AM model shows an underestimation of the 10-year return level compared to the maximum measured data. The annual maxima series resulted in a value of 5.66 m as the 100-year return level for the total  $H_s$  (Fig. 8), which is comparable to Teena et al. (2012) estimation for the location off the central west coast of India.

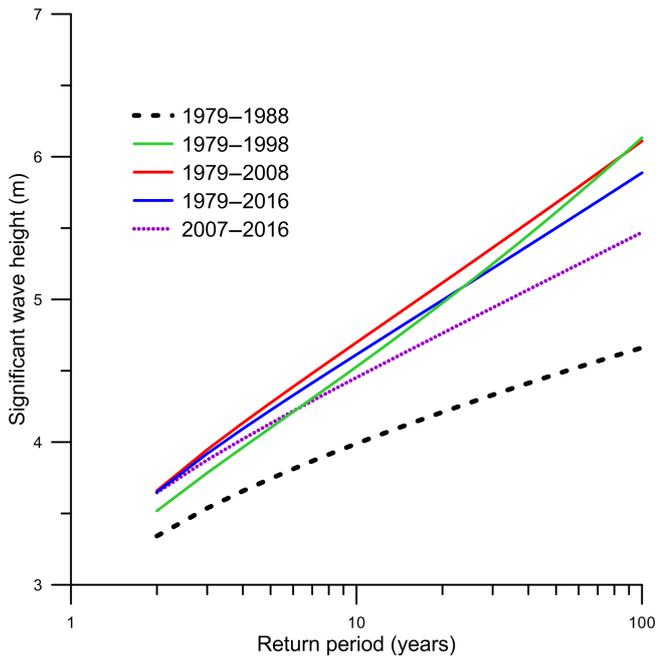
We performed a separate analysis of the annual maxima series to get insight into the abnormal results observed for wind-sea data series. Here, we considered four unique series of different lengths by taking annual maxima observations from 2008 to 2016; that is, the first series (S1) consists of five data points (2008–2012) and second series (S2) consists of six data points (2008–2013) and so on. The density plots showing the probability for different wave height class are presented in Fig. 9 along with the corresponding GPD fit. We calculated the standard deviation for each series and the percentage difference between each series and the parent series (S0). The result shows that return levels are positively



**Figure 9.** Density plots showing the probability for different wave height class. Total, wind-sea and swell  $H_s$  are presented row wise. Columns correspond to the selected number of data points (5 to 8 years). The solid curve is the corresponding GPD fit.

**Table 6.** Table showing the results of the case study. The standard deviations (SDs) of each data series considered are provided, and percentage differences among the SDs of each series with parent series (S0) are given in the brackets. The percentage difference in the corresponding return level estimation is also shown in the brackets of the respective return periods.

Data set	Series (years)	Maximum observed (m)	Standard deviation (% difference)	Return levels		
				10 years (m)	50 years (m)	100 years (m)
Total	S1 (2008–2012)	4.32	0.36 (21.8)	4.24 (6.1)	4.89 (8.6)	5.20 (10.42)
	S2 (2008–2013)	4.32	0.32 (32.7)	4.17 (8.0)	4.67 (13.1)	4.90 (15.5)
	S3 (2008–2014)	4.32	0.32 (34.5)	4.23 (6.4)	4.65 (13.5)	4.83 (17.2)
	S0	4.70	0.45	4.50	5.28	5.66
Wind sea	S1 (2008–2012)	2.80	0.13 (128.90)	2.82 (14.81)	2.88 (51.29)	2.89 (72.30)
	S2 (2008–2013)	2.80	0.14 (125.06)	2.81 (15.00)	2.95 (48.96)	3.00 (69.16)
	S3 (2008–2014)	2.89	0.16 (114.08)	2.89 (12.35)	3.05 (45.80)	3.11 (66.00)
	S0	4.29	0.60	3.27	4.86	6.16
Swell	S1 (2008–2012)	3.47	0.23 (48.17)	3.65 (8.23)	4.16 (14.93)	4.45 (18.36)
	S2 (2008–2013)	3.47	0.20 (58.53)	3.62 (9.18)	4.01 (18.53)	4.22 (23.56)
	S3 (2008–2014)	3.47	0.22 (50.80)	3.71 (6.62)	4.05 (17.53)	4.21 (23.97)
	S0	4.28	0.37	3.97	4.83	5.35



**Figure 10.** Return levels of significant wave heights for different return periods based on ERA-Interim shallow water data in different block years by the GEV model using annual maxima series.

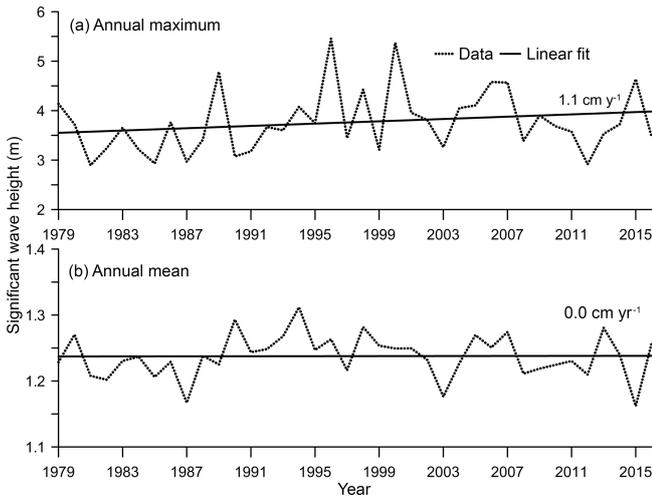
correlated with standard deviation (Table 6). In the case of the total  $H_s$ , the correlation between the changes in standard deviation and the corresponding changes in 100-year return levels is 0.997, whereas for wind sea and swell they are 0.964 and 0.647 respectively. The annual maxima of wind sea (4.29 m) for the year 2015 caused an abrupt change in the standard deviation of the series by about 0.46 m, which is more than 17 % of the average of the series excluding 2015. Therefore, the 100-year return level for wind sea overshoots by about 6.16 m, resulting in a 66 % difference from return value obtained for S3 series. In this case study, the length of the special series under consideration does not influence the estimated return levels; that is, in the case of the total  $H_s$  series, the 100-year return level for the S1 series is greater than for both the S2 and S3 series. The same characteristics can also be seen in the case of swell  $H_s$ . Therefore, return levels for annual maxima by the GEV model have greater influence over how a single data point, i.e., the annual maxima, alters the standard deviation of the series rather than the changes in the length of the series.

### 3.3 Influence of length of wave data on the estimated significant wave height return value

An analysis is carried out to check uncertainties in return level estimation related to the length of the wave record. From the 0.5 h buoy-measured data, data at 6 h intervals are extracted and used for the analysis, and the return levels obtained by using 6 h measured buoy data are compared

with the return level obtained from the 6 h ERA-Interim data at shallow and deep locations (Fig. 8). The 6-hourly ERA-Interim reanalysis 38-year data (1979–2016) are used in this analysis. Buoy data consist of 11 479 data points, and ERA-Interim data consist of 55 520 data points (Table 1). The highest observed  $H_s$  in the 6-hourly buoy data is 4.11 m followed by 4.03 m, while the maximum  $H_s$  in ERA-Interim shallow water is 5.45 m and in ERA-Interim deep water it is 7.13 m. The  $H_s$  values at the deep location are  $\sim 1.4$  times the values at the shallow location, and this resulted in a higher return level of  $H_s$  at the deep location. Sanil Kumar and Muhammed Naseef (2015) observed that ERA-Interim overestimates the  $H_s$  for shallow water locations along the west coast of India due to swell height overestimation, and the difference between the ERA-Interim  $H_s$  and the buoy  $H_s$  is up to 15 %. For the study location, the storm-induced wave heights during the non-monsoon period are less than the monsoon-induced waves. The 1st week of June is the onset of the Indian summer monsoon, and the maximum  $H_s$  in the study area is due to monsoon influence; in all years, this occurs during June to September. The 100-year return levels using the GEV method are comparable for buoy data (4.18 m) and ERA-Interim shallow data (4.39 m), while that for ERA-Interim deep is 5.67 m (Fig. 8). It is clear that the 100-year  $H_s$  return level using GEV for ERA-Interim data is lower than the maximum  $H_s$  in the data, while, in the case of buoy data, the 100-year return level is slightly higher than the highest  $H_s$  value. The return levels obtained by the GPD method show significant discrepancy among 100-year estimates. The 100-year return level obtained for buoy data is 4.46 m, but that using ERA-Interim shallow data is 6.18 m and that for ERA-Interim deep is 7.28 m. The 100-year  $H_s$  return level for deep water has closer values following GEV and GPD, while, in the shallow water, a significant difference is obtained. The 6 h interval data tend to miss 18 values of  $H_s$  between 4.11 and 4.70 m, and hence there is a significant difference in the 100-year return level of  $H_s$  based on GEV-AM obtained using these data compared to that based on the data at 0.5 h intervals.

We have examined the difference in the return level of  $H_s$  by considering data in different blocks; i.e., 10, 20, 30 and 38 years using the ERA-Interim shallow water data. The study indicates a large underestimation ( $\sim 18$  %) in the return level estimate if we consider only the first 10 years of (1979–1988) data in place of the 38 years (Fig. 10). The large difference in the values of  $H_s$  return level is due to the occurrence of a tropical storm in the Arabian Sea during 9–12 June 1996, which resulted in high wave heights,  $H_s$  values of up to 5.46 m, whereas the maximum  $H_s$  excluding this storm is 4.63 m. During the 1996 storm, an  $H_s$  of 5.69 m is measured by a Datawell directional Waverider buoy moored at 23 m water depth off Goa (Sanil Kumar et al., 2006), which is  $\sim 150$  km north of the present study area. The data blocks containing this storm data, i.e., the 20-year (1979–1998) and 30-year (1979–2008) data, did not show much difference in



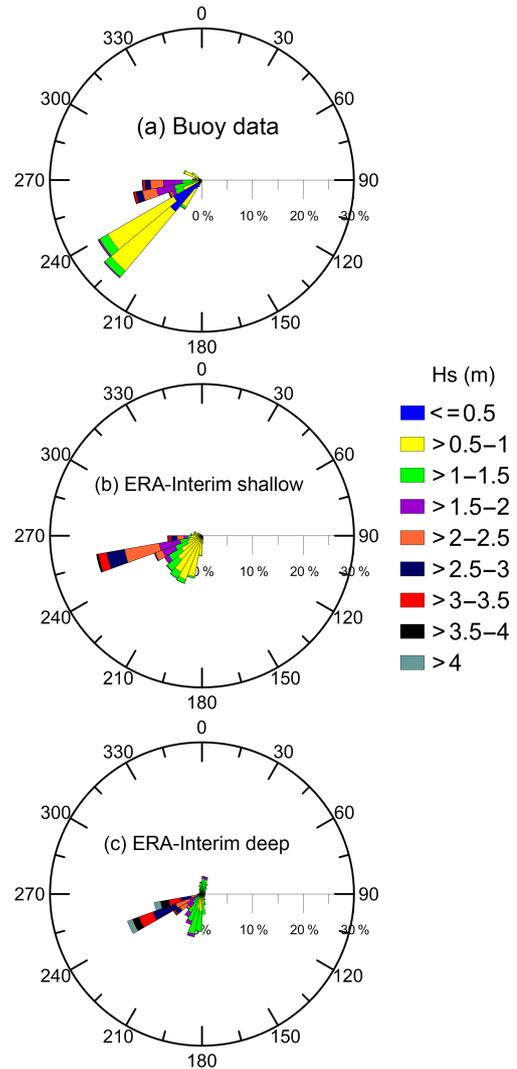
**Figure 11.** Variation of the (a) annual maximum and (b) annual mean  $H_s$  at the shallow locations based on ERA-Interim data. The solid line indicates the trend in  $H_s$  during 1979 to 2016.

the 100-year  $H_s$  value compared to the 38-year data. If we consider only the last 10 years (2007–2016), it resulted in a 7% underestimation in the 100-year  $H_s$  value. The study shows that a single storm can create a large difference in the 100-year  $H_s$  value, compared to the differences in values that resulted from a different length of the data block.

The long-term and decadal trend of wave climate in the different parts of major oceans is studied (Young et al., 2011). We have examined the trend in  $H_s$  at the shallow location based on the ERA-Interim data from 1979 to 2016. The study shows that the annual maximum  $H_s$  shows a weak increasing trend ( $1.1 \text{ cm yr}^{-1}$ ), whereas there is no significant trend in the annual mean value (Fig. 11). Sanil Kumar and Anoop (2015) observed that during 1979 to 2012 the average trend of annual mean  $H_s$  for all the locations in the western shelf seas was  $0.06 \text{ cm yr}^{-1}$ .

### 3.4 Influence of water depth on the measured buoy data

The relative water depth based on the spectral peak period ( $d/L_p$ ) indicates that most of the time (97.8 to 99.3%) the wave regime is in intermediate water (Table 7). Only during 0.1 to 0.8% of the time do the waves satisfy the deep water condition. Hence, the waves measured by the buoys are influenced by the bathymetry, and the wave characteristics are different in the deep water. The wave rose plots from March 2008 to February 2016 based on the measured buoy data and the ERA-Interim reanalysis data at shallow and deep water locations are presented in Fig. 12. As the waves move from deep to shallow waters, the direction of high waves shifted from southwest to west. The limiting value of wave height based on breaker criteria is 0.6 to 0.78 times the water depth (Massel, 1966). The maximum  $H_s$  in the measured



**Figure 12.** Wave rose plots from March 2008 to February 2016 based on the measured buoy data and the ERA-Interim reanalysis data at shallow and deep water locations.

buoy data is 4.70 m, and some of the waves containing this record are very steep or broken at 9 m water depth since the maximum wave height is 1.65 to 1.8 times the  $H_s$ .

## 4 Conclusions

Long-term statistical analysis of extreme waves is carried out based on GEV and GPD models using measured buoy data from March 2008 to February 2016 and the ERA-Interim data from 1979 to 2016. Return levels are calculated for resultant, wind-sea and swell  $H_s$  separately. The analysis is also conducted for data under three different seasons. The parent data are resampled into 3-, 6-, 12- and 24-hourly series and are used to estimate the discrepancy in return level estimation between the different series. Selection of appro-

**Table 7.** The percentage of time of the waves in the shallow, intermediate and deep water regime in different years along with the mean wave period and mean peak wave period.

Year	Mean wave period (s)	Criteria based on ratio of water depth and wave length corresponding to mean wave period			Mean peak wave period (s)	Criteria based on ratio of water depth and wave length corresponding to peak wave period		
		Shallow water	Intermediate water	Deep water		Shallow water	Intermediate water	Deep water
2008–2009	5.5	0	98.7	1.3	12.1	1.0	98.9	0.1
2009–2010	5.6	0	98.3	1.6	12.0	0.5	99.3	0.2
2010–2011	5.4	0	97.5	2.5	11.7	0.6	99.2	0.2
2011–2012	5.7	0	99.5	0.5	11.9	0.9	98.5	0.6
2012–2013	5.5	0	99.4	0.6	12.0	0.3	99.6	0.1
2013–2014	5.0	0	95.0	5.0	11.8	1.4	97.8	0.8
2014–2015	5.7	0	98.7	1.3	12.6	1.8	98.1	0.1
2015–2016	5.5	0	98.0	2.0	12.3	0.8	99.0	0.2

appropriate thresholds for the POT method is justified using different GOF tests results. Analysis of the total  $H_s$  shows that the IDM approach underestimates return levels for different seasons compared to the corresponding GPD. The 100-year return levels estimated by IDM are almost comparable with the corresponding GPD estimation for the 10-year period, but there is a significant difference in the return level estimates when considering different sampling intervals. IDM estimates largely underestimate return levels for the post-monsoon season since the majority of the observation in this season lies away from its tail of the distribution.

Long-term statistics of wind-sea and swell data are calculated by the GEV model following block maxima and the  $r$ -largest methods. Annual maxima and monthly maxima are considered for block maxima series, and two seasonal maxima series are considered for the  $r$ -largest method. It is shown that these methods give higher return levels than the GPD models. The  $r$ -largest method provides 7.20 m as the 100-year return level when compared to 5.27 m of the GPD model. The sensitivity analysis of the GEV-AM model shows that change in the standard deviation of data series under consideration causes discrepancies in the return level estimates rather than a change in the length of the series. Both GEV and GPD models underestimate 10-year return levels compared to maximum measured data. The 100-year return levels acquired by using the GEV method are comparable for short-term (2008 to 2016) buoy data (4.18 m) and the long-term (1979 to 2016) ERA-Interim shallow data (4.39 m). The 6 h interval data tend to miss high values of  $H_s$ , and hence there is a significant difference in the 100-year return level  $H_s$  obtained using these data compared to data at 0.5 h intervals. The ERA-Interim data show that from 1979 to 2016 the annual maximum  $H_s$  shows a weak increasing trend ( $1.1 \text{ cm yr}^{-1}$ ). The study shows that a single storm can create a large difference in the 100-year  $H_s$  value, compared to the differences in values obtained from a different length of the data block.

*Data availability.* The measured wave data used in the study can be requested from the corresponding author for joint research work. The long-term data on significant wave height and wind speed are from the ERA-Interim global atmospheric reanalysis data set of the ECMWF and are available at <http://www.ecmwf.int/en/research/climate-reanalysis/era-interim> (Dee et al., 2011).

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The director of the CSIR–National Institute of Oceanography, Goa, provided the facilities to carry out the study. Shri Jai Singh, technical officer, CSIR–NIO, assisted in the data analysis. This work forms part of the PhD thesis of the first author and is CSIR–NIO contribution number 6103 under the institutional project MLP1701. We thank the editor Mauricio Gonzalez and the two anonymous referees for their suggestions for improving the manuscript.

Edited by: Mauricio Gonzalez

Reviewed by: two anonymous referees

## References

- Abild, J., Andersen, E. Y., and Rosbjerg, D.: The climate of extreme winds at the Great Belt, Denmark, *J. Wind. Eng. Ind. Aerod.*, 41, 521–532, 1992.
- Alves, J. H. G. M. and Young, I. R.: On estimating extreme wave heights using combined Geosat, Topex/Poseidon and ERS-1 altimeter data, *Appl. Ocean. Res.*, 25, 167–186, 2003.
- Anoop, T. R., Kumar, V. S., Shanas, P. R., and Johnson, G.: Surface wave climatology and its variability in the north Indian Ocean Based on ERA-interim reanalysis, *J. Atmos. Ocean. Tech.*, 32, 1372–1385, <https://doi.org/10.1175/JTECH-D-14-00212.1>, 2015.
- Brodtkorb, P. A., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., and Sjö, E.: WAFO – a Matlab toolbox for analysis of random

- waves and loads, Proc. 10th Int. Offshore and Polar Eng. Conf., Seattle, USA., Vol III, 343–350, 2000.
- Caires, S. and Sterl, A.: 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data, *J. Climate.*, 18, 1032–1048, 2005.
- Choulakian, V. and Stephens, M. A.: Goodness-of-Fit Tests for the Generalized Pareto Distribution, *Technometrics*, 43, 478–484, 2001.
- Coles, S., Bawa, J., Trenner, L., and Dorazio, P.: An introduction to statistical modeling of extreme values, Springer-Verlag, London, UK, 205 pp., 2001.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- Goda, Y.: Uncertainty of design parameters from viewpoint of extreme statistics, *J. Offshore Mech. Arct.*, 114, 76–82, 1992.
- Goda, Y., Hawkes, P., Mansard, E., Martin, M. J., Mathiesen, M., Peltier, E., and Van Vledder, G.: Intercomparison of extremal wave analysis methods using numerically simulated data, Proc. WAVES'93, New Orleans, USA, 963–977, 1993.
- Goda, Y., Kudaka, M., and Kawai, H.: Incorporation of Weibull distribution in L-moments method for regional frequency of peaks-over-threshold wave heights, Proceedings of 32nd international conference on coastal engineering, American Society of Civil Engineers, Proceedings of 32nd Conference on Coastal Engineering, Shanghai, China, 2010, edited by: McKee Smith, J. and Lynett, P. <https://doi.org/10.9753/icce.v32.waves.62>, 2010.
- Gumbel, E. J.: Statistics of extremes, Columbia Univ. Press., New York, 1958.
- Hosking, J. R. M., Wallis, J. R., and Wood, E. F.: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, *Technometrics*, 27, 251–261, 1985.
- Izaguirre, C., Mendez, F. J., Menendez, M., Luceño, A., and Losada, I. J.: Extreme wave climate variability in southern Europe using satellite data, *J. Geophys. Res.*, 115, C04009, <https://doi.org/10.1029/2009jc005802>, 2010.
- Jonathan, P., Randell, D., Wu, Y., and Ewans, K.: Return level estimation from on-stationary spatial data exhibiting multidimensional covariate effects, *Ocean Eng.*, 88, 520–532, <https://doi.org/10.1016/j.oceaneng.2014.07.007>, 2014.
- Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, *Adv. Water Resour.*, 25, 1287–1304, [https://doi.org/10.1016/s0309-1708\(02\)00056-8](https://doi.org/10.1016/s0309-1708(02)00056-8), 2002.
- Massel, S. R.: On the largest wave height in water of constant depth, *Ocean Eng.*, 23, 553–573, 1966.
- Mathiesen, M., Goda, Y., Hawkes, P. J., Martín, M. J., Peltier, E., and Edward, F.: Recommended practice for extreme wave analysis, *J. Hydraul. Res.*, 32, 803–814, 1994.
- Méndez, F. J., Menéndez, M., Luceño, A., Medina, R., and Graham, N. E.: Seasonality and duration in extreme value distributions of significant wave height, *Ocean Eng.*, 35, 131–138, 2008.
- Menéndez, M., Méndez, F. J., Izaguirre, C., Luceño, A., and Losada, I. J.: The influence of seasonality on estimating return values of significant wave height, *Coast Eng.*, 56, 211–219, <https://doi.org/10.1016/j.coastaleng.2008.07.004>, 2009.
- Panchang, V., Zhao, L., and Demirebilek, Z.: Estimation of extreme wave heights using GEOSAT measurements, *Ocean Eng.*, 26, 205–225, [https://doi.org/10.1016/s0029-8018\(97\)10026-9](https://doi.org/10.1016/s0029-8018(97)10026-9), 1999.
- Pickands, J.: Statistical inference using extreme order statistics, *Ann. Stat.*, 3, 119–131, 1975.
- Portilla, J., Ocampo-Torres, F. J., and Monbaliu, J.: Spectral Partitioning and Identification of Wind Sea and Swell, *J. Atmos. Oceanic Technol.*, 26, 107–122, 2009.
- Samayam, S., Laface, V., Annamalaisamy, S. S., Arena, F., Vallam, S., and Gavrilovich, P. V.: Assessment of reliability of extreme wave height prediction models, *Nat. Hazards Earth Syst. Sci.*, 17, 409–421, <https://doi.org/10.5194/nhess-17-409-2017>, 2017.
- Sanil Kumar, V. and Anoop, T. R.: Spatial and temporal variations of wave height in shelf seas around India, *Nat. Hazards*, 78, 1693–1706, <https://doi.org/10.1007/s11069-015-1796-5>, 2015.
- Sanil Kumar, V. and Muhammed Naseef, T.: Performance of ERA-Interim wave data in the nearshore waters around India, *J. Atmos. Ocean. Tech.*, 32, 1257–1269, <https://doi.org/10.1175/JTECH-D-14-00153.1>, 2015.
- Sanil Kumar, V., Pathak, K. C., Pednekar, P., Raju, N. S. N., and Gowthaman, R.: Coastal processes along the Indian coastline, *Curr. Sci. India*, 91, 530–536, 2006.
- Sanil Kumar, V., Shanas, P. R., and Dubhashi, K. K.: Shallow water wave spectral characteristics along the eastern Arabian Sea, *Nat. Hazards*, 70, 377–394, <https://doi.org/10.1007/s11069-013-0815-7>, 2014.
- Shaji, C., Kar, S. K., and Vishal, T.: Storm surge studied in North Indian Ocean: A review, *Indian J. Geo-Mar. Sci.*, 43, 125–147, 2014.
- Smith, R. L.: Extreme value theory based on the  $r$  largest annual events, *J. Hydrol.*, 86, 27–43, [https://doi.org/10.1016/0022-1694\(86\)90004-1](https://doi.org/10.1016/0022-1694(86)90004-1), 1986.
- Soares, C. G. and Scotto, M.: Application of the  $r$ -largest-order statistics for long-term predictions of significant wave height, *Coast Eng.*, 51, 387–394, <https://doi.org/10.1016/j.coastaleng.2004.04.003>, 2004.
- Solari, S. and Losada, M. A.: Non-stationary wave height climate modeling and simulation, *J. Geophys. Res.*, 116, C09032, <https://doi.org/10.1029/2011JC007101>, 2011.
- Solari, S. and Losada, M. A.: Unified distribution models for met-ocean variables: Application to series of significant wave height, *Coast. Eng.*, 68, 67–77, 2012.
- Solari, S., Egüen, M., Polo, M. J. and Losada, M. A.: Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit  $p$ -value, *Water Resour. Res.*, 53, 2833–2849, 2017.
- Stephens, M. A.: EDF Statistics for Goodness of Fit and Some Comparisons, *J. Am. Stat. Assoc.*, 69, 730–737, <https://doi.org/10.1080/01621459.1974.10480196>, 1974.

- Teena, N. V., Sanil Kumar, V., Sudheesh, K., and Sajeev, R.: Statistical analysis on extreme wave height, *Nat Hazards*, 64, 223–236, <https://doi.org/10.1007/s11069-012-0229-y>, 2012.
- Thevasiyani, T. and Perera, K.: Statistical analysis of extreme ocean waves in Galle, Sri Lanka, *Weather Clim. Extrem.*, 5–6, 40–47, <https://doi.org/10.1016/j.wace.2014.07.003>, 2014.
- Viselli, A. M., Forristall, G. Z., Pearce, B. R., and Dagher, H. J.: Estimation of extreme wave and wind design parameters for offshore wind turbines in the Gulf of Maine using a POT method, *Ocean Eng.*, 104, 649–658, <https://doi.org/10.1016/j.oceaneng.2015.04.086>, 2015.
- Young, I. R., Zieger, S., and Babanin, A.: Global trends in wind speed and wave height, *Science*, 332, 451–455, 2011.