Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

ASSESSMENT OF RELIABILITY OF EXTREME WAVE HEIGHT PREDICTION MODELS

Satish Samayam[1], Valentina Laface[2], Sannasiraj Sannasi Annamalaisamy[1], Felice Arena[2], Sundar Vallam[1], Polnikov Vladislav Gavrilovich[3].

[1]Department of Ocean Engineering, Indian Institute of Technology Madras, India.

[2]Mediterranea University of Reggio Calabria, Italy.

[3]Obukhov Institute for Physics of Atmosphere of the Russian Academy of Sciences, Russia.

Correspondence to: Sannasiraj S.A. (sasraj@iitm.ac.in)

Abstract

Extreme waves influence coastal engineering activities and have an immense geophysical implication. Therefore, their study, observation and extreme wave prediction are decisive for planning for mitigation measures against natural coastal hazards, ship routing, design of coastal and offshore structures. In this study, the estimates of design wave heights associated with return period of 30 and 100 years are dealt with in detail. The design wave height is estimated based on four different models to obtain a general and reliable model. Different locations are considered to perform the analysis: four sites in Indian waters, one in Mediterranean Sea and two in North America. For the Indian water domain European Centre for Medium-Range Weather Forecasts (ECMWF) global atmospheric reanalysis ERA-interim wave hind cast data covering a period of 36 years have been utilized for this purpose. For the locations in Mediterranean Sea and North America both ERA-interim wave hind cast and buoy data are considered. The reasons for the variation in return value estimates of the ERA-interim data and the buoy data using different estimation models are assessed in detail.

1. Introduction

Indian Ocean with two horns of the Arabian sea and the Bay of Bengal has been playing a significant role in the regional economic development. This rapid progress is attributed to a variety of activities in the coastal and offshore sectors that include construction and development of major ports and fishing harbours, establishment of power plants, offshore exploration and exploitation of oil and gas, and tampering of ocean wave and tidal energy. To sustain these developments along the coast, the aforementioned activities require a variety of coastal and offshore structures such as groins, sea walls, breakwaters, offshore platforms, intake and outfall structures, submarine pipelines etc. to be constructed in the marine environment. It is hence mandatory to design these structures for its life span which could be

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

achieved by considering its survival conditions. The most dominant environmental forces that dictates this design of structure is due to the maximum probable wave height of a site of interest (Massel, 1978).

Depending on the importance and life span of the structure, the return period of the extreme events could be selected as 30 years or 100 years. The lesser would be associated with lesser wave height but more risk and vice versa. It demands a better understanding of hydrodynamic characteristics of local wave environment, especially the extreme conditions. In the design of any marine structures, the first step is the extreme wave analysis for the determination of design wave heights with certain return periods (Goda, 2000). Estimation of appropriate design values indicate the level of protection and the scale of investment during the construction of structure.

Fundamentally, extreme values are scarce and are necessarily outside the range of the available observations, implying that an extrapolation from the observed sea states to unknown territories is required. An estimate of anticipated wave height can be furnished using historical wave hind-cast data or field observed data with the help of various distribution models, which enable extrapolation under the Extreme value theory framework (Goda, 2000; Coles, 2001; Caires, 2011). Ferreira and Soares, 2000 suggested that the estimation of extreme values should rely on methods based on extreme value theory which makes use of the largest of the observations in the sample. Coles, 2001 obtained the detailed statistical results of extreme value prediction using the annual maximum (AM) (Castillo, 1988) and Peaks Over Threshold (POT) (Ferreria and Soares, 1998) sampled observations. Caires, 2011 rigorously compared the commonly used extreme value statistical methods (like GEV and GPD) with different parameter estimation methods for combination of different data sampling techniques.

Another approach that may be applied starting from a wave data time series is that of equivalent storm models (Boccotti 1986, 2000; Arena and Pavone, 2009; Fedele and Arena, 2010; Laface and Arena, 2016) which is based on the concept of sea storm. Specifically these models consist in substituting the sequence of sea storms at a given site (actual sea) with a sequence of equivalent storms (equivalent sea) from a statistical perspective. The equivalent storms have very simple geometric shapes such as triangular (Boccotti 1986, 2000; Arena and Pavone, 2009), power (Fedele and Arena, 2010) or exponential (Laface and Arena, 2016). Depending on the shape the related model gives analytical or numerical

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

solution for the calculation of the return period $R(H_s>h)$ of a sea storm whose maximum H_s is greater than a given threshold h. Specifically the triangular and exponential equivalent storm models give closed form solution for $R(H_s>h)$, while the Equivalent power storm model requires numerical calculation. In the paper the Equivalent Triangular (ETS) model is utilized.

The accuracy of any methodology for extreme values significantly depends on the length of the recorded time series data. It is believed that measurements from wave rider buoy offer the most reliable long historical record. However, the availability of such buoy data is limited to a certain specific locations, mainly in the northern hemisphere. At a particular location of interest, the availability of buoy data is usually scarce, and often there will be no data. The oceanographic community has recognized the hind-casts with ocean wave models to complement the limited buoy observational records.

In the recent years, the performance of wave models has appreciably improved, with better quality of the wind fields and enhancement in numerical wave modeling. The meteorological centres like European Centre for Medium-Range Weather Forecasts (ECMWF), Australian Bureau of Meteorology and Meteo-France that operate global wave models are currently using altimeter wind data for data assimilation purposes. The process combines numerical wave model and observations of diverse sorts in the best possible ways to generate a consistent, global estimate of the various atmospheric, wave and oceanographic parameters. At present, in numerous meteorological centres, wind and wave simulated data are assimilated on a daily basis.

The simulated hindcast data have been adopted in numerous studies for the estimation of extreme wave conditions. Teena et al., 2012 applied a generalized extreme value distribution and generalized Pareto distribution to the 31 years assimilated wave hindcast data based on MIKE-21, a spectral wave model for a location in the eastern Arabian sea and extracted extreme wave for several return periods. Li et al., 2016 used a third generation wave model, WAMC4 and simulated 35 years of wave hindcast data from two sets of reanalysis wind data, NCEP and ECMWF. In their study, Pearson-III distribution method is used to analyse the extreme wave climate in the East China seas. Polnikov and Gomorev, 2015 proposed to use the extrapolation of polynomial approximation constructed for the shorter part of the tail of probability function to estimate the return values of wind speed and wind-wave height. The

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences Discussions

wave field was computed from wind-wave model, WAM-C4M from ECMWF global atmospheric reanalysis ERA-interim wind field data.

Even though several studies have been carried out, the study on the identification of the most suitable approach for estimating extreme wave heights for a particular source of assimilated wave hindcast data is still lacking. In the present study, the investigation of different existing approaches and models is carried to assess its application and reliability for the Indian domain. Increased uncertainty in the model outputs questions the reliability of the estimation model, which is an important issue. Thus, the present study introduces a statistical approach to validate the reliability of the design wave height return values resulting from a particular extreme wave estimation method by considering variability criterion as measured maximum value. The variation in the extreme value estimates of the ERA-interim data and the buoy data for different estimation models is also considered and examined. The objective of the present study is to identify a robust extreme wave height estimation method for Indian domain using global atmospheric reanalysis ERA-interim wave hindcast data.

2. Datasets

2.1 Study Locations

Four offshore locations along the Indian coast (Fig.1) are considered. The selection of these particular locations is based on their distance from the nearest coast and the water depth, two each on east and west coasts of the Indian peninsula. Both deep and shallow water locations are chosen to examine the application of the estimation model based on water depth.

The projected estimates using ERA-Interim data are compared with those obtained from data from various buoys to validate the performance of ERA-Interim data in extreme wave analysis. The choice of the locations was according to the size of wave data that were available. Further, two locations in North America, National Data Buoy Center Station 44005 in Gulf of Maine, National Data Buoy Center Station 46050 West of Newport and one of the most energetic sites in the coasts of Central Mediterranean Sea (Liberti et al., 2013; Vicinanza et al., 2013; Arena et al., 2015) from the Italian buoys network (RON) locations, Alghero (West coast of Sardinia Island) are considered. A comprehensive comparison has been carried out by extracting the ERA-Interim data of resolution 0.125^{o}x0.125^{o} nearest to the selected buoy locations. The coordinates, period of data availability, interval and number of data points for these locations are presented in Table 1.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

2.2 Wave Data

2.2.1 ERA-Interim data

ERA-Interim data is produced by the ECMWF, which is a global atmospheric reanalysis from 1979, continuously updated in real time and is one among the most recent re-analysis data available (Berrisford et al., 2009). ERA-Interim is the first to perform re-analysis using adaptive and fully automated bias corrections of observations (Dee and Uppala, 2008). The parameters such as significant wave height (H_s), mean wave direction and mean wave period can be obtained with 6-hourly fields covering the whole globe, with the best space resolution of 0.125^ox0.125^o.

There have been several studies comparing the values of H_s between ERA-Interim dataset and buoy data at different locations around the world to evaluate the model performance. It has been found that at certain locations in the Arabian sea, the maximum H_s based on ERA dataset in deep water is about 15% less than that of buoy measured data, whereas, in shallow waters, ERA dataset over predicts the maximum H_s by about 9% (Shanas and Kumar, 2014). The under prediction in deep water suggests that extreme events attained mainly during cyclones are difficult to be captured by the model. The results show that H_s of model data set are reliable in both deep and shallow water locations with a good degree of accuracy. The estimates in this study are based on ERA-Interim wave hind cast data, covering a period of 36 years (1979-2014). For nearest intersection buoy locations, the data period was selected based on buoy data availability.

2.2.2 Buoy data

The most reliable data for significant wave height are from the buoy measurements. The available length of buoy data is usually limited and the data prior to 1978 is scanty. The available buoy data further requires significant quality control on account of large gaps of missing data and outlier, flagship measurements. In the paper data from two different buoys networks are processed: RON (Rete Ondametrica Nazionale) Italian network and the National Oceanic and Atmospheric Administration's National Data Buoys Center (NOAA-NDBC).

The Italian buoys network (RON) started measurements in 1989, with 8 directional buoys located off the coasts of Italy. Later it has reached the number of 15 buoys moored in deep water. For each record, the data of significant wave height, peak and mean period and dominant direction are given.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences Discussions

159 The NOAA manages the NDBC, which consists of many buoys moored along the US coasts,
160 both in the Pacific Ocean and in the Atlantic Ocean. Some buoys were moored in the late
161 1970s, so that more than 35 years of data are available. The historical wave data give hourly
162 significant wave height, peak and mean period. The NOAA buoy observations are readily
163 available which are of proven quality. The measurements have passed through quality control
164 by NOAA. It is however always recommended to perform some basic quality checks.

165 The return value estimates acquired from the ERA-Interim data are compared with that of
166 NDBC Stations 44005, 46050 and at Alghero along the coast of Central Mediterranean Sea.
167 Table-1 provides the coordinates and data details of these buoy stations. ERA-Interim wave
168 hindcast data has been used to assess the estimates in Indian waters.

169 **3. Extreme wave height Estimation Methods**

170 **3.1 General**

171 The estimation models used in this study to obtain extreme wave return values include the
172 generalised extreme value (GEV) and the generalised Pareto distribution (GPD), which are
173 currently being adopted for the standard practice in mainstream extreme statistics. Each
174 distribution was fit to the data using the Maximum likelihood method (MLE) and the
175 Probability weighted moments method (PWM). Further, new polynomial approximation
176 model prescribed by Polnikov and Gomorev, 2015 and Equivalent Triangular storm model
177 (Boccotti, 2000) based on the concept of replacing the sequence of actual storms extrapolated
178 from a given time series of H_s with a sequence of equivalent triangular storms are used.
179

180 **3.2 Generalised extreme value distribution model**

181 According to extreme value theory, to form a valid distribution, the sampled observations
182 should be independent which would mean that successive observations should not be
183 correlated with one another and identically distributed (Goda, 2000). In general, for the
184 sampling of data to be used for extreme wave analysis three different approaches are
185 available. The first approach uses all the recorded data of H_s during a number of years and fits
186 a cumulative distribution to this data. This approach is called the initial distribution method
187 (IDM). For the other two approaches, only the peaks of wave heights are engaged. The
188 method of block maxima consists of partitioning recorded data in blocks, where in, maximum
189 value of each block is considered. Normally a block could be chosen as one year (Lionello et
190 al., 1992). The POT (Peaks Over Threshold) method, consists of the peaks of clustered data
191 exceeding over a given threshold. IDM observations violate the conditions of identity and

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences Discussions

192 independence in distribution, which invalidates the application of the common statistical
193 methods as well as the definition of return values (Anderson et al., 2001). The annual maxima
194 method and POT method both satisfy the obligatory of independency.
195 According to theory of the generalized extreme value (GEV) distribution, the sample has
196 been selected by means of annual maxima (AM) method.

197 The generalized extreme value (GEV) distribution has the cumulative distribution function
198 (CDF) as:

199 $$GEV(x;\mu,\sigma,\xi) = \begin{cases} exp\left(-\left(1-\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{1}{\xi}}\right), & for\ \xi \neq 0 \\ exp\left(-exp\left(\frac{-(x-\mu)}{\sigma}\right)\right), & for\ \xi = 0 \end{cases} \quad (1)$$

200 where, μ, σ and ξ represent the location, scale and shape parameters of distribution,
201 respectivelyand within the range of $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$ .By setting the shape
202 parameter, ξ, one can obtain the most common distributions like Gumbel (ξ=0), Frechet
203 (ξ>0) and Weibull (ξ<0).

204 The 1/T yr wave height return value, X_T based on the GEV distribution model is given as

205 $$X_T = \begin{cases} \mu - \frac{\sigma}{\xi}\left\{1-\left[-log\left(1-\frac{1}{T}\right)\right]^{\xi}\right\}, & for\ \xi \neq 0 \\ \mu - \sigma ln\left[-log\left(1-\frac{1}{T}\right)\right], & for\ \xi = 0 \end{cases} \quad (2)$$

206 **3.3 Generalised Pareto distribution model**

207 This approach is based on fitting the generalized Pareto distribution (GPD) to the POT
208 sampled data. The observations in a cluster above the threshold are considered and
209 calculating return values has been done by taking into account the rate of occurrence of
210 clusters (Davidson and Smith, 1990; Coles, 2001).
211 The cumulative distribution function of GPD is given as:

$$GPD(x;\mu,\sigma,\xi) = \begin{cases} 1-\left(1-\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{1}{\xi}}, & for\ \xi \neq 0 \\ 1-exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right), & for\ \xi = 0 \end{cases} \quad (3)$$

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

where $0 < x < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. where μ is the threshold, scale (σ) and shape (ξ) parameters. When $\xi = 0$ the GPD is said to amounts to the exponential distribution with mean σ; when $\xi > 0$, it is the Pareto distribution; and when $\xi < 0$ it is a special case of the beta distribution.

The 1/T yr wave height return value based on the GPD distribution model ,X_T, is given as

$$X_T = \begin{cases} \mu + \frac{\sigma}{\xi}\{1 - (\lambda T)^{-\xi}\}, & for\ \xi \neq 0 \\ \\ \mu + \sigma ln(\lambda T)\xi, & for\ \xi = 0 \end{cases} \quad (4)$$

where, $\lambda = N_u/N$, with N_u being the total number of exceedances above the selected threshold u and N are the number of years in the record.

There are several parameter estimation methods for fitting the above candidate distribution functions to the sampled wave data (Goda, 2000). The method of moments (MM), probability weighted moments (PWM) method and the maximum likelihood method (MLE) are more preferred estimation methods since these are more flexible, particularly when the number of parameters is increased. The MM yields a large bias particularly for small size samples and this method was not used in the present study.The parameters of the above distributions are derived according to the methods of maximum likelihood method (MLE) and probability weighted moments (PWM).

The threshold selection in GPD analysis is an important practical problem, which is analogous to the block size in the block maxima approach. The threshold value represents a compromise between bias and variance. Too low a threshold violates the asymptotic basis of the GPD model, leading to a bias. Too high a threshold will generate fewer values of excess to estimate the model, leading to high variance. There is an extensive literature on the attempt to choose an optimal threshold by Neelamani, 2009; Caires, 2011. In this study, the threshold selection is based on the Mean residual life plots introduced by Davison and Smith, 1990.

The mean residual life plot is based on the theoretical mean of the GPD given as:

$$E[x] = \mu + \frac{\sigma}{1-\xi}, \text{ for } \xi < 1 \quad (5)$$

The mathematical basis for Mean residual life plots method is

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences Discussions

$$E[X - y/X > y > 0] = \frac{\sigma + \xi y}{1 - \xi} \text{ , for } \xi < 1 \quad (6)$$

If X is GPD, then the mean excess over a threshold y (for y > 0) with slope $\xi/(1 - \xi)$ is a linear function of y. Thus, we can draw a plot in which the ordinate is the sample mean of all excesses over that threshold and the abscissa is the threshold.

A mean residual life plot consists in representing points:

$$\left\{\left(\mu, \frac{1}{n}\sum_{i=1}^{n} x_i, n - \mu\right) : \mu \le x_{\max}\right\} \quad (7)$$

where n is the number of observations (x_i, i=1,2,…n) above the threshold μ, and x_{max} is the maximum of the observations. According to the Central Limit Theorem, confidence intervals are added to this mean residual life plot as the empirical mean to be normally distributed. However, this normality doesn't hold for high threshold as there are less and less excesses.

3.4 Polynomial approximation model

Polnikov and Gomorev, 2015 proposed to use the extrapolation of polynomial approximation constructed for the shorter part of the tail of probability function to estimate the return values of wind speed and wave height.

This method involves the construction of an analytical approximation $F_{ap}(H)$, aimed for its extrapolation beyond the observed maximum value H_M. The approximation should be restricted to a shorter domain lying above the uppermost mode of the histogram considered of the function $F(H)$.The domain suitable for approximation can be determined by the condition

$$H_l \le H \le H_h \le H_M \quad (8)$$

where H_l and H_h are the lower and the upper edges of the domain of $F(H)$, used for constructing approximation $F_{ap}(H)$. The number of points (N_M) considered in the histogram is $H_M/\Delta H$ and N_S is defined as,

$$N_S = \left(H_M - H_h\right)/\Delta H \quad (9)$$

And the number of points (N_T) used for building approximation $F_{ap}(H)$ is defined as,

$$N_T = \left(H_h - H_l\right)/\Delta H + 1 \quad (10)$$

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

The approximation, $F_{ap}(H)$ should be built in the logarithmic coordinates due to few values in the tail of $F(H)$, providing importance to the tail values. It allows to assess the strong variability of the tail of function $F(H)$ near the maximum value of the series, depending on the length of the series. To exclude the application of fixed statistics, the approximation function $F_{ap}(H)$ in the form of a polynomial of degree n, the value of which may vary is considered. The varying n allows obtaining the approximation $F_{ap}(H,n)$ with an accuracy higher than the case of using the fixed statistical distributions.

The statistical distribution with the provision function is of the form,

$$F_{ap}(H) = \exp\left[\sum_{i=0}^{i=n} a_k H^k\right] \tag{11}$$

Once the approximation function, $F_{ap}(H)$ is obtained from Eq.(11), the return value, X_R, appearing once for N_R years, can be deduced by inversing the formula,

$$F(X_R) = \Delta t / 8760 \cdot N_R \tag{12}$$

where, Δt is the interval of discrete of data observations.

Another, principal features of P-approximation $F_{ap}(W)$ are given as the standard deviation δ (hereinafter, the deviation), defined by the formula:

$$\left(\frac{1}{N_T}\cdot\sum_{H_i=H_l}^{H_i=Hh}\left[\ln(F(H_{iap}))-\ln(F\ (H\ ))\right]^2\right)^{1/2}=\delta \tag{13}$$

Obviously for, the lesser δ the higher accuracy of approximation can be achieved and it is more preferable.

3.5 Equivalent Triangular Storm model

The Equivalent Triangular Storm (ETS) model (Boccotti, 2000; Arena and Pavone, 2006, 2009) is applied for calculating return values of significant wave height for given thresholds of return period. The ETS approach is based on the assumption that given a sequence of actual storms it may be replaced by an equivalent storm sequence maintaining the same wave risk. The validity of the above assumption is guaranteed by the statistics equivalence between the actual storm and the related Equivalent Triangular (ETS) one. The ETS associated with a

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences Discussions

given storm is achieved by means of two parameters: the triangle height *a* and its base *b* (Fig. 2). The former is an intensity parameter and is equaled to the maximum significant wave height during the actual storm, the latter is a duration parameter and it is determined following an iterative procedure imposing the equality between the maximum expected wave heights of actual and triangular storms. It has been experimentally proved that imposing this equality not only the area under the exceedance probability curves of the maximum wave height are the same, but those curves tend to coincide.

Considering all these aspects emerges that the actual storm and the ETS sequences (actual and Equivalent Triangular seas) have the same number of storms, each of them characterized by the same maximum significant wave height and the same probability $P(H_{max}>H)$ that the maximum wave height is greater than a fixed threshold *H*. The considerations above enable to affirm that the return period of a sea storm with given characteristics is the same if calculated starting from the actual storm sequence or the ETSs one. Referring to equivalent triangular sea, an analytical solution for the calculation of the return period $R(H_s>h)$ of a sea storm whose maximum significant wave height is greater than a given threshold *h* has been developed by Boccotti, 2000.

$$R(H_s > h) = \frac{\bar{b}(h)}{hp(H_s = h) + P(H_s > h)} \tag{14}$$

Where $\bar{b}(h)$ is the base-height regression function of ETSs, $P(H_s > h)$ is the probability of exceedance of the significant wave height H_s at the considered site and $p(H_s = h) = -\frac{dP(H_s > h)}{dh}$ is the probability density function of H_s.

The calculation of return values of H_s by means of Eq. (14) requires the determination of two functions:

- the base-height regression function, $\bar{b}(h)$ which gives the average value of the base *b* of ETSs for a given storm height *h*;
- the probability $P(H_s > h)$.

The function $\bar{b}(h)$ is determined starting from the ETSs sequence. Specifically, starting from a given storm sequence, the related ETSs sequence is determined by calculating for each event the intensity and duration parameters *a* and *b* of the ETS. Then two different

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences Discussions

assumptions can be made. The simplest is to consider an average value of *b*, in other words to assume a constant base-height regression. The other is to calculate a linear or exponential regression function. In general, the function $\bar{b}(h)$ presents a decaying pattern, because of which, an exponential function is preferable in order to guarantee positive values of the base *b* for any storm threshold. To calculate the regression, initially the data have to be divided in classes of triangle height and then for each class the average values of *a* and *b* have to be calculated. Then the regression may be easily determined representing data in a Cartesian plot with parameters *b* in y axis and parameter *a* in x axis respectively. By fitting the data by means of an exponential law as per the relationship given below:

$$\bar{b}(a) = k_1 \exp(k_2 a) \tag{15}$$

where k_1 (hours) and k_2 (m^{-1}) are parameters depending on the storms characteristics at the considered site.

The determination of the base-height regression function despite very simple from a computational and mathematical point of view, requires careful attention because of its sensibility to the time interval between the data of H_s used in the analysis. In this regard, it is worth noting that ETS duration parameter *b*, is strongly dependent on the actual storm structure close to the storm peak. Specifically it tends to increase as the storm structure became flat and it is quite small for steep storms. When data sampling interval is more than one to three hours, one may have very flat storms. This involves that the calculation may lead to big values of duration *b*. This aspect causes that return values of H_s may be underestimated (Arena et al., 2013). This aspect strongly affect predictions when wave model data are processed (3 to 6 hours between two successive data of H_s). To overcome this problem, a good practice is to do the analysis in conjunction with buoy data close to the location under study. In these cases, the base height regression function calculated from buoys is utilized for correcting the base height regression function obtained from model data.

Concerning the distribution of the significant wave height $P(H_s > h)$, three-parameters Weibull distribution is considered.

$$P(H_s > h) = \exp\left[-\left(\frac{h - h_l}{w}\right)^u\right] \tag{16}$$

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

where u, h_l and w are the characteristic parameters at the considered location. In particular u and h_l are the shape parameters and w is the scale parameter of the distribution.

4. Results

In this study, hindcast results for ERA-interim data are compared with buoy measurements for different estimation models. Further study of the various uncertainties due to the parameter estimation method, the sample size, sample interval and location conditions involved in this analysis are also examined.

4.1 GEV analysis

In the application of generalised extreme value distribution to the sampled annual maxima data, the scale, shape and location parameters can be used to make statements about the probability of the annual maximum exceeding a particular level. A change in either parameter can affect the long-period return levels.

The parameter estimation is done by maximum likelihood estimate method (MLE) and probability weighted moments (PWM) method (Hosking et al., 1985) and resulted parameters are shown in Table 2. It has been observed that the shape parameter is positive for ERA-interim data indicating that this data would follow Frechét distribution and the tail of the cumulative distribution function decreases more slowly.

The influence of estimated parameters in fitting the data to the GEV model is presented in Fig. 3a. It shows the level of fitting of the empirical CDF with the GEV PWM and GEV MLE models. The difference in the normal coordinates in their fitting with empirical CDF is insignificant. Figure 3b shows the variation in tail estimates of the PWM and MLE parameter estimation methods in logarithmic scale. The results show for both buoy and ERA-interim data sets, the PWM method of parameter estimation yields better estimates compared to the MLE method.

The statistical parameter, root mean square error was estimated in order to check the level of fitting of sampled data to the GEV distribution model. The root mean square error is a residual between the empirical cumulative distribution obtained from the actual observed data and the theoretical GEV model cumulative distribution. Lower the value of RMSE i.e., near to zero indicates a better fit of sampled data to the GEV distribution model. The fitting of GEV to buoy and ERA-interim data is found to be good for both PWM and MLE methods.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

The variation RMSE value of the MLE estimates is usually smaller than those of the PWM estimates for both buoy and ERA-interim data.

4.2 GPD analysis

In POT method, the selection of a suitable threshold value is the key in achieving a robust sample data set. The mean residual plot, between the mean excess GPD and threshold helps in determining a proper range of threshold to be selected (Coles, 2001). Such plots with 95% confidence for the data ERA IN-1, (Fig. 4) appear to have two slopes with major transition at the threshold range of 1.5 to 2.5 indicates the range of threshold could possibly be selected. However, attention should be made as too high threshold can results in less sampled data set which results in a higher variance of the GPD model.

The sample used in the peaks over threshold method has to be extracted in such a way that the data can be modelled as independent observations. A process of de-clustering helps collecting only the peaks within the clusters of successive exceedances of a specified threshold and are retained in such a way that they are sufficiently apart (so that they belong to 'independent storms'). Specifically, in the present applications we have treated cluster maxima at a distance of less than 48 hours apart as belonging to the same cluster (Caires, 2011). Table 3 provides the selected threshold and the number of exceedences of that specified threshold with 48h interval. It is seen that the threshold values are observed to be dependent on the length, location and interval of the datasets. The major factor has to be the location, since the higher latitude locations are exposed to more severe wave and wind conditions than those at the lower latitudes.

For parameter estimation, the PWM and MLE methods are used. The MLE has a considerable statistical motivation but can turn out to be poor estimators, especially in the case where the number of estimated parameters is large. So the approach chosen here was to utilize a variety of techniques like PWM and MLE for exploratory fitting for the probability model and chose the best possible parameters.

To verify the estimated parameters for the GPD model, quantile-quantile (QQ) plots were used. In Fig. 5a, the QQ plots for the dataset NOAA44005 is shown, comparing the estimated GPD with the sample data for PWM parameter estimation method. In order to check the influence of parameters resulting from PWM and MLE parameter estimation models, the Root mean square error was estimated for GPD model also and presented in Table 3.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

Comparing the estimates and the fits, one can conclude that the MLE fits seem less adequate and that the shape parameter estimates are lower than those of the PWM fits. These results support the recommendations of Hosking et al., 1985 to preferably use the PWM method for GPD or GEV estimation from relatively short duration of data with limited heavy-tailed cumulative distributions. Figure 5b shows the return value GPD plot of PWM fit to the dataset NOAA44005.

4.3 Polynomial approximation method analysis

Polynomial approximation (P-approximation) method has a distinct advantage of selecting the optimum choice of the parameters N_S, N_T, and n. The detailed analysis demonstrates that all approximation parameters (n, N_T, and N_S) are equally important. Figure 6 shows the application of P-approximation method for both buoy and ERA-interim data at Alghero buoy station. In the above mentioned figure, the bottom level (ln(1-F)= -12.6) indicates probability of occurrence once for 100 years and can be deduced by Eq. 12 with discrete of data observations of 3hr interval. For 1hr and 6hr interval of data observations, the probability of occurrence once in 100 years can obtained as -13.7 and -11.9 respectively.

One can see the adaptation of P-approximation to the real behavior of the tails for provision functions. For the Alghero location buoy data, the optimized parameters obtained are N_S=0, N_T=8 (points used for approximation), n=2 (degree of approximation function) to arrive at the optimum return value as shown in the Fig. 6.

The optimum choice of parameters will also depends on standard deviation δ (Eq. 13) which resembles the residual between the actual tail of the provision function and the Polynomial approximation tail fitted to it. Lower the value of δ i.e., near to zero indicates a better fit between actual tail of the provision function and the Polynomial approximation tail fitted. The parameters N_S, N_T, and n for all the datasets including the resulted standard error δ are provided in Table 4.

4.4 Analysis of ETS Model

The calculation of the 100 yrs return values via ETS model is done by means of Eq. (13), known the base-height regression function Eq. (14) and the probability distribution Eq. (15) of H_s at the examined location. The base height regression function is determined staring from the storm sequence at the considered site, while the probability distribution is achieved

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

processing the whole data set of H_s. An important aspect to be taken into account in estimating the parameters of both Eq. (14) and Eq. (15) is the time interval between two successive data of H_s. A value of 3-6 hours should be appropriate for estimating the parameters of the probability distribution, in order to guarantee the stochastic independence between successive events, but could be too high for determining the parameters of the base-height regression function.

In fact, Arena et Al., 2013 has shown that as the time interval between two successive H_s increases, the peak of the storm may not be well identified, involving flat storm history that led to an increase of the duration b of ETSs respect to the case with lower time interval between H_s data. Such situation are those of wave model data. In this paper both wave model and buoy data are considered.

To determine the base-height regression function parameters, the actual storm sequence is identified starting from H_s time series, and for each actual storm the parameters a and b of ETS are calculated (Boccotti, 2000). Then the ETS are divided into classes of H_s of 1m width and the average value a_m and b_m of a and b for each class is considered. The sequence a_m, b_m is plotted in a Cartesian diagram and fitted by an exponential law as the Eq. (14). For the case of wave model data a further step is required. Specifically, considering an increase of b due to high time interval between H_s data, the regression should be corrected considering a reducing factor r, defined as the ratio between the average values of the base calculated starting from buoy data moored close to the considered site and the average value calculated by means of wave model data. The regression parameters k_1 and k_2 at each considered site are summarized in Table 5 in conjunction with the parameters u, w and h_l of the probability distribution Eq. (15).

5. Discussions

From the results it is observed that the estimates from buoy observations are higher compared to the estimates for ERA-interim datasets. This trend is being observed from all the estimation models. Shanas and kumar, 2014 have found that at certain location in Arabian sea the maximum H_s based on ERA dataset in deep water is about 15% less than that of buoy measured data. Herein, also a similar phenomenon of variation of 20-30% while comparing maximum observed H_s of Buoy data and ERA-interim at NOAA440050, NOAA46050 and Alghero buoy locations is observed. This in turn will result in under estimation of return value of ERA interim data.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

The under prediction of ERA interim data suggests, that high wave events mainly due to the cyclone events are difficult to capture by ECMWF numerical model. It is a familiar phenomenon and challenge that the smoothing effect implanted in numerical model will lead to the flattened variability at relatively high frequencies, resulting in the missing peaks. An additional potential explanation for the under prediction is that the simulated ERA-Interim data contains 6- hourly intervals of H_s data. It is possible because of the lower sampling rate, the maximum wave heights in a storm occurs between observations will not be recorded. To overcome this, it is obvious that ECMWF numerical modeling system need further improvement in correction or calibration of the ERA-interim data especially when this hindcast is used for the extreme wave analysis.

Final results on the 30,100 year extreme wave estimates, obtained by the GEV, GPD, ETS and Polynomial approximation methods described above are presented in Table 6 and 7. The variation of these estimates from the measured maximum wave heights will give a statistical validation of the performance of the estimation models. The percentage of variation of 30-yr and 100-yr return value estimates from measured 36 year maximum wave height are calculated for this analysis. Here one can observe the following principal peculiarities from the results of above mentioned statistical validation methodology.

The GEV and GPD methods, shows the 30 year return values smaller than the measured maximum H_s for all the locations mostly by an extent of 10-25%. In the cases of simulated data these models exhibit high deviations from measured maximum H_s. This peculiarity is because of the reason of neglecting the behavior of the tails for provision functions, accepted in GEV and GPD methods. As a result, this leads to underestimating the return values. This is a reasonable shortage of these methods, as far as one cannot forecast extreme smaller return values than ones observed already.

The GEV model with annual maxima sample resulted in over estimation of return values compared to the GPD model with peaks over threshold approach. As the GEV estimation model considers only the highest H_s in the year, which might lead to the overestimation of Annual maxima based approach in comparison with the other method. For most of the locations, there is not much variation in the results of the PWM, MLE parameter estimated GEV and GPD models. But Hosking et al., 1985 recommended to always applying the PWM

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

parameter estimation method for GPD and GEV distribution models from relative short datasets with not too heavy-tailed distributions. Furthermore, PWM works for a wider range of parameter values than MLE method.

The results from the P-method are remarkably closer to the measured maximum values than those obtained by the GEV ,GPD and ETS method, which variation is as high as 5% to as less as -7%. The Polynomial approximation method shows consistency in 100-yr estimated return values for both simulated and buoy wave height datasets, as these varies consistently between 7-13% from the measured maximum values. GEV, GPD and ETS methods fails in the above mentioned criterion as variation is as high as 56% to as low as -19% which are not possible in nature.

This consistency of polynomial approximation method estimates is due the dependence of return values on the actual kind of the tail for provision function, which could vary and is dependent on the sample of the series. The only disadvantage of P-app method ($F_{ap}(Hs,n)$) is the necessity to control reliability of its extrapolation, as far as the extrapolation of polynomial with the order n > 1 may have twists and extremes. This well-known fact could be provided by a considerable variability of the "tail" for *F(W)*. Such an extrapolation is implausible, of course. Therefore, it is necessary to vary the parameters N_S, N_T, and the order of polynomial n in such a manner, the twists of extrapolation could be avoided.

6. Conclusions

In this study we chose the simulated ERA-Interim wave data, for the two following reasons. First, they have more regular coverage for the whole World ocean, and the Indian coast, in particular. Second, numerical simulated datasets have long and regular continuous series, what is very important for the extreme value statistical aims.

This study focused on the estimation of the extreme significant wave heights only. The analyses carried out and results produced will aid in the preparation of a 100 year extreme wave map for the Indian water domain which may serve as a quick guide to identify regions where extremes lie within the design criteria of the coastal and offshore structures to be constructed.

We have considered four different approaches to the return values estimating: the GEV distribution model based on annual maxima sample, the GPD distribution model based on

peaks over threshold sample, ETS model based on storms and the polynomial approximation method for the tail of provision function. All of them have their own advantages and shortages.

The main shortage of the GEV and GPD methods are the high variation in underestimating or overestimating return values with respect to ones presenting in the time-series. The shortage of the P-app method is related to the ambiguity of the return values estimations, obtained from different parts of the full time-series. It is also found that the values estimated based on GEV model were slightly larger than those from the GPD. But GPD method with peaks over threshold sample is preferable in the locations of multiple storm events in a single year. In turn, the estimates through the Polynomial approximation method, depend on the actual kind of the tail for provision function, which could vary and is dependent on the sample of the series resulted in showing the consistency in 100-yr estimated return values for both simulated and buoy wave height datasets, as these varies consistently between 7-13% from the measured maximum values.

It is observed that the return value estimates from buoy observations are higher when compared to the estimates for ERA-interim datasets.The under prediction of ERA interim data suggests, that high wave events mainly due to the cyclone events are difficult to capture by ECMWF numerical model. To overcome this, it is obvious that ECMWF numerical modeling system need further improvement in correction or calibration of the ERA-interim data especially when this hindcast is used for the extreme wave analysis.

Acknowledgments

This paper has been developed during the Marie Curie IRSES project “Large Multi- Purpose Platforms for Exploiting Renewable Energy in Open Seas (PLENOSE)” funded by European Union (Grant Agreement Number: PIRSES-GA-2013-612581).

This work was partially supported by the joint Russian-Indian (RFBR-DST) grant, # 14-05-92692-Ind-a.

References

Anderson, C. W., D. J. T. Carter, and P. D. Cotton, 2001: Wave climate variability and impact on offshore design extremes. Shell International and the Organization of Oil and Gas Producers Rep., 99 p.

568 Arena, F. & Pavone, D. (2006) The return period of non-linear high wave crests, Journal of
569 Geophysical Research, Vol. 111, No. C8, paper C08004, doi: 10.1029/2005JC003407

570 Arena, F. & Pavone, D. (2009) A generalized approach for the long-term modelling of
571 extreme sea waves, Ocean Modelling, Vol. 26, Issue 4, pp. 217-225.
572 doi:10.1016/j.ocemod.2008.10.003

573 Arena, F., Laface, V., Barbaro, G., Romolo, A., 2013. Effects of sampling between data of
574 significant wave height for intensity and duration of severe sea storms, International Journal
575 of Geosciences, Vol. 4, Issue 1A, pp. 240-248 doi:10.4236/ijg.2013.41A021.

576 Arena F., Laface V., Malara G., Romolo A., Viviano A., Fiamma V., Sannino G., Carillo A.,
577 2015. Wave climate analysis for the design of wave energy harvesters in the Mediterranean
578 Sea, Renewable energy , volume 77, pp. 125-141, ISSN: 0960-1481.

579 Berrisford, P., D. Dee, K. Fielding, M. Fuentes, P. Kalberg, S. Kobayashi, and S. Uppala, The
580 ERA-Interim Archive.ERA report series, 2009(1): p. 1-16.

581 Boccotti, P., 1986. On coastal and offshore risk analysis. Excerpta of the Italian Contribution
582 to the Field of Hydraulic Eng., 19–36.

583 Boccotti, P., 2000. Wave mechanics for ocean engineering. Elsevier, Amsterdam, The
584 Netherlands, 496 pp.

585 Castillo,E., 1988: Extreme value theory in Engineering. Academic Press, 389 pp.

586 Coles, S., 2001: An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag,
587 208 pp.

588 Davidson & Smith (1990), Models for Exceedances over high Thresholds, Journal of the
589 Royal Statistical Society B,52, pp. 393-442.Grimshaw, S. D. (1993)

590 Dee, D. and S. Uppala, Variational bias correction in ERA-Interim2008: ECMWF.

591 Fedele, F., and Arena, F., 2010. The equivalent power storm model for long-term predictions
592 of extreme wave events. J. Phys. Oceanogr. 40, 1106–1117.

593 Ferreira, J.A., and C. Guedes Soares, 1998: An application of the peaks over threshold
594 method to predict extremes of significant wave height. J. Offshore Mech. Arct. Eng., 120,
595 165–176.

596 Ferreira, J. A., and C. Guedes Soares, 2000: Modelling distributions of significant wave
597 height. Coast. Eng., 40, 361–374.

598 Goda Y. Statistical analysis of extreme waves (Chapter 11). In: Random seas and design of
599 maritime structures, 15th volume of advanced series on ocean engineering, World Scientific
600 Publishing Co., 2000; 377- 425.

601 Hosking, J.R.M, Wallis, J.R. and Wood E.F. (1985), Estimation of the generalized extreme-
602 value distribution by the method of probability-weighted moments Technometrics (27), pp.
603 251-261

604 Jiangxia Li, Yongping Chen, Shunqi Pan, Yi Pan, Jiayu Fang, Derrick M.A. Sowa,
605 Estimation of mean and extreme waves in the East China Seas, Applied Ocean Research
606 56(2016) 35-47.

607 Laface, V. and Arena, F., (2016), A new equivalent exponential storm model for long-term
608 statistics of ocean waves, Coastal Engineering, Volume 116, Pages 133-151.

609 Liberti, L., Carillo, A., Sannino, G., (2013), Wave energy resource assessment in the
610 Mediterranean, the Italian perspective. Renewable Energy; 50: 938-49.

611 Lionello, P., H. Gunther, and P. A. E. M. Janssen. 1992. Assimilation of altimeter data in a
612 global third-generation wave model. J. Geophys. Res. 97(◦C9):14453–14474.

613 Massel RS (1978) Needs for Oceanographic Data to Determine Loads on Coastal and
614 Offshore Structures. IEEE Journal of Oceanic Engineering vol oe-3 no. 4.

615 Neelamani, S., (2009), Influence Of threshold value on Peak Over Threshold Method on the
616 predicted extreme significant wave heights in Kuwaiti territorial waters. J. Coast. Res.,
617 Special Issue 56 564-568.

618 P.R.Shanas and V. Sanilkumar, Comparison of ERA-Interim waves with buoy data in eastern
619 Arabian sea during high waves, Indian journal of Marine sciences vol.43(7), July 2014,pp.

620 Sofia Caires (2011),Extreme value analysis: wave data, JCOMM Technical Report No. 57

621 N.V. Teena, V. Sanil Kumar, K. Sudheesh, R. Sajeev, Statistical analysis on Extreme wave
622 Height, Natural Hazards, vol.64; 2012; 223-236.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

623 V.G. Polnikov and I.A. Gomorev, On Estimating the return values for time series of wind
624 speed and wind wave height, ISSN 1068-3739, Russian Meteorology and Hydrology, 2015,
625 Vol. 40, No. 12, pp. 820–827. Ó Allerton Press, Inc., 2015.

626 Vicinanza, D., Contestabile, P. & Ferrante, V., (2013), Wave energy potential in the north-
627 west of Sardinia (Italy). Renewable Energy 50: 506-521.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

628 **Table 1:** ERA-Interim data locations and Buoy stations

Data Point	Coordinates	Availability	Interval (Hr)	No. of data points
ERA IN-1	*19.50N, 85.75E*	1979-2014	6	52596
ERA IN-2	*15.50N, 81.00E*	1979-2014	6	52596
ERA IN-3	*10.25N, 75.75E*	1979-2014	6	52596
ERA IN-4	*14.50N, 73.50E*	1979-2014	6	52596
NDBC 44005	*43.204N, 69.128W*	1978-2014	1	254221
ERA 44005	*43.25N, 69.125W*	1979-2014	6	52596
NDBC 46050	*44.656N, 124.526W*	1991-2014	1	180231
ERA 46050	*44.625N, 124.50W*	1991-2014	6	35064
RON Alghero	*40.548N,8.107E*	1989-2008	3	125443
ERA Alghero	*40.5N,8.125E*	1989-2008	6	29220

629

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

630 **Table 2:** PWM and MLE parameter estimators fitting GEV

Data	PWM method				MLE method			
	ξ	**μ**	**σ**	**RMSE**	**ξ**	**μ**	**σ**	**RMSE**
ERA IN-1	0.1125	3.1523	0.3849	0.053	0.1157	3.1572	0.3779	0.045
ERA IN-2	0.2085	1.9181	0.3108	0.081	0.4971	1.8838	0.2499	0.039
ERA IN-3	0.0311	2.8386	0.3279	0.032	0.0296	2.8413	0.3270	0.035
ERA IN-4	0.1169	3.6889	0.4553	0.033	0.1118	3.6975	0.4485	0.029
NOAA 44005	-0.1642	6.7735	1.0880	0.052	-0.1811	6.7958	1.0571	0.023
ERA 44005	-0.0866	5.0506	0.5649	0.031	0.0457	5.0706	0.5741	0.030
NOAA 46050	-0.1190	8.9863	1.5655	0.052	-0.1038	8.9429	1.6407	0.039
ERA 46050	-0.0251	7.1700	0.7646	0.047	0.0554	7.1705	0.7268	0.051
RON Alghero	-0.5089	7.4373	1.4405	0.112	-0.4992	7.4498	1.3588	0.043
ERA Alghero	0.0746	5.555	0.6298	0.069	-0.0874	5.5719	0.6003	0.061

631

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

632 **Table 3:** PWM and ML parameter estimators fitting GPD

Data	Threshold μ	No. of exceedence	PWM method			MLE method		
			σ	ξ	RMSE	σ	ξ	RMSE
ERA IN-1	2.5	153	0.4429	0.0415	0.028	0.4489	0.0286	0.026
ERA IN-2	1.5	160	0.2515	0.1438	0.045	0.2471	0.1588	0.052
ERA IN-3	2.5	107	0.3350	-0.0485	0.036	0.3149	0.0143	0.025
ERA IN-4	3.0	154	0.5428	-0.0651	0.035	0.5200	-0.0206	0.025
NOAA 44005	5.0	227	1.3147	-0.1677	0.055	1.3396	-0.1892	0.063
ERA 44005	4.0	190	0.7335	-0.1159	0.201	0.6938	-0.0560	0.025
NOAA 46050	6.0	232	1.4608	-0.0200	0.019	1.5058	-0.0514	0.031
ERA 46050	5.0	203	1.5480	-0.3879	0.126	1.2886	-0.1654	0.066
RON Alghero	5.0	153	1.6541	-0.2957	0.100	1.6110	-0.2614	0.089
ERA Alghero	4.0	128	0.9342	-0.1474	0.053	0.9642	-0.1835	0.066

633

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

634 **Table 4:** Selection of optimum values of approximation parameters

Data	No. of points used for approximation N_T	n	δ
ERA IN-1	6	2	0.176
ERA IN-2	6	3	0.044
ERA IN-3	5	3	0.032
ERA IN-4	7	3	0.063
NOAA 44005	8	2	0.118
ERA 44005	7	1	0.197
NOAA 46050	5	2	0.026
ERA 46050	6	1	0.200
RON Alghero	8	2	0.100
ERA Alghero	7	2	0.105

635

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

636 **Table 5:** Base-height regression parameters k_1, k_2 calculated considering a storm sample with
637 actual durations greater or equal to 18 hours, probability distribution parameters u, w and h_l.

Data	u	w[m]	h_l[m]	k_1[h]	k_2[m^{-1}]
ERA IN-1	1.320	0.714	0.459	397.61	-0.251
ERA IN-2	0.773	0.142	0.481	255.73	-0.097
ERA IN-3	1.600	0.851	0.488	348.02	-0.086
ERA IN-4	1.504	1.099	0.498	397.6	-0.159
NDBC 44005	1.121	1.150	0.409	76.125	0.0308
ERA 44005	1.141	0.884	0.461	114.05	-0.071
NDBC 46050	1.333	1.945	0.480	154.9	-0.101
ERA 46050	1.625	2.321	0.000	106.94	-0.055
RON Alghero	1.155	1.299	0.000	318.37	-0.235
ERA Alghero	1.227	1.157	0.000	135.53	-0.035

638

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

639 **Table 6:** 30 year return value estimates (m)

Data	Measured maximum	GEV		GPD		P-App	ETS
		PWM	MLE	PWM	MLE		
ERA IN-1	4.91	4.8	4.8	4.9	4.8	4.6	4.7
ERA IN-2	3.59	3.5	4.1	3.3	3.4	3.6	3.3
ERA IN-3	4.83	4.0	4.0	3.9	4.1	4.6	4.3
ERA IN-4	6.17	5.6	5.5	5.1	5.5	5.9	5.9
NOAA 44005	10.10	9.6	9.5	9.5	9.4	10.6	9.9
ERA 44005	8.27	7.3	7.2	6.5	7.0	7.9	8.0
NOAA 46050	14.05	13.7	13.4	12.4	12.3	14.1	12.8
ERA 46050	10.93	9.9	9.9	8.0	9.0	10.2	9.5
RON Alghero	9.88	9.8	9.7	9.4	9.5	9.2	10.3
ERA Alghero	7.51	7.5	7.4	6.6	6.9	7.6	7.4

640

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

641 **Table 7:** 100 year return value estimates (m)

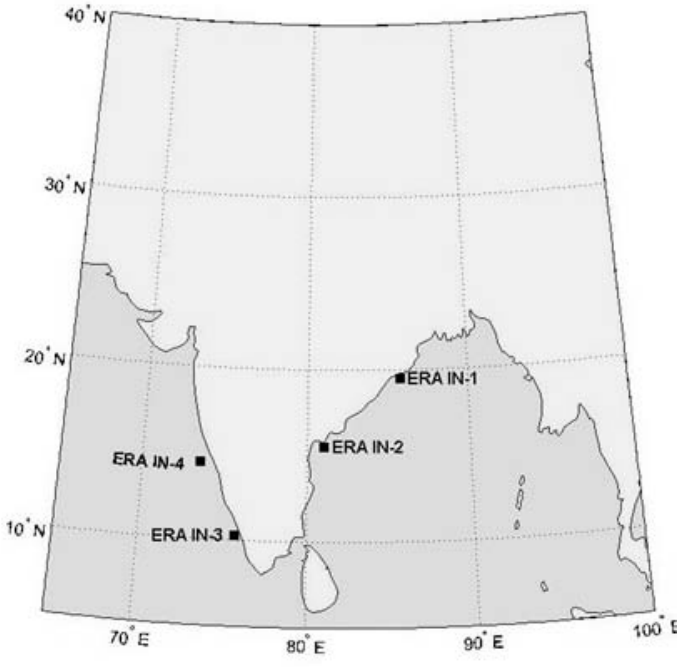
Data	Measured maximum	GEV		GPD		P-App	ETS
		PWM	MLE	PWM	MLE		
ERA IN-1	4.91	5.5	5.5	5.65	5.5	4.8	5.1
ERA IN-2	3.59	4.3	5.6	4.0	4.1	4.0	3.6
ERA IN-3	4.83	4.5	4.5	4.2	4.4	4.7	4.4
ERA IN-4	6.17	6.5	6.6	5.6	6.0	6.4	6.1
NOAA 44005	10.10	10.3	10.1	10.1	10.0	11.4	10.7
ERA 44005	8.27	8.3	8.0	7.2	7.9	9.0	8.4
NOAA 46050	14.05	15.1	14.6	14.2	14.1	15.2	13.8
ERA 46050	10.93	10.9	11.0	8.9	9.8	11.3	11.1
RON Alghero	9.88	10.1	10.0	9.9	10	9.7	12.5
ERA Alghero	7.51	8.5	8.0	7.7	8.0	8.2	8.7

642

643

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

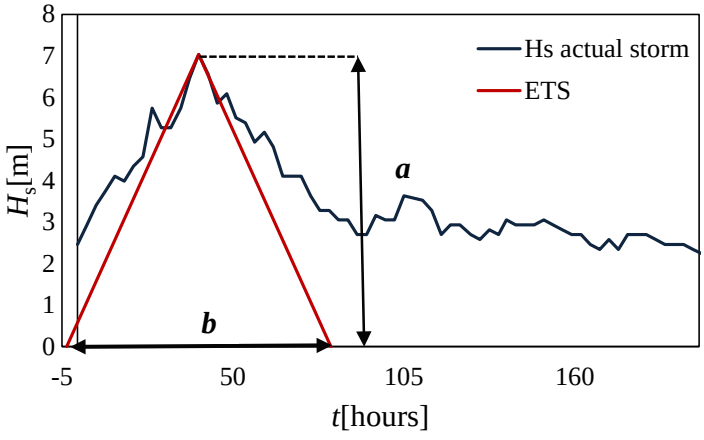


644

645 **Figure 1:** Locations along the Indian coast

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions

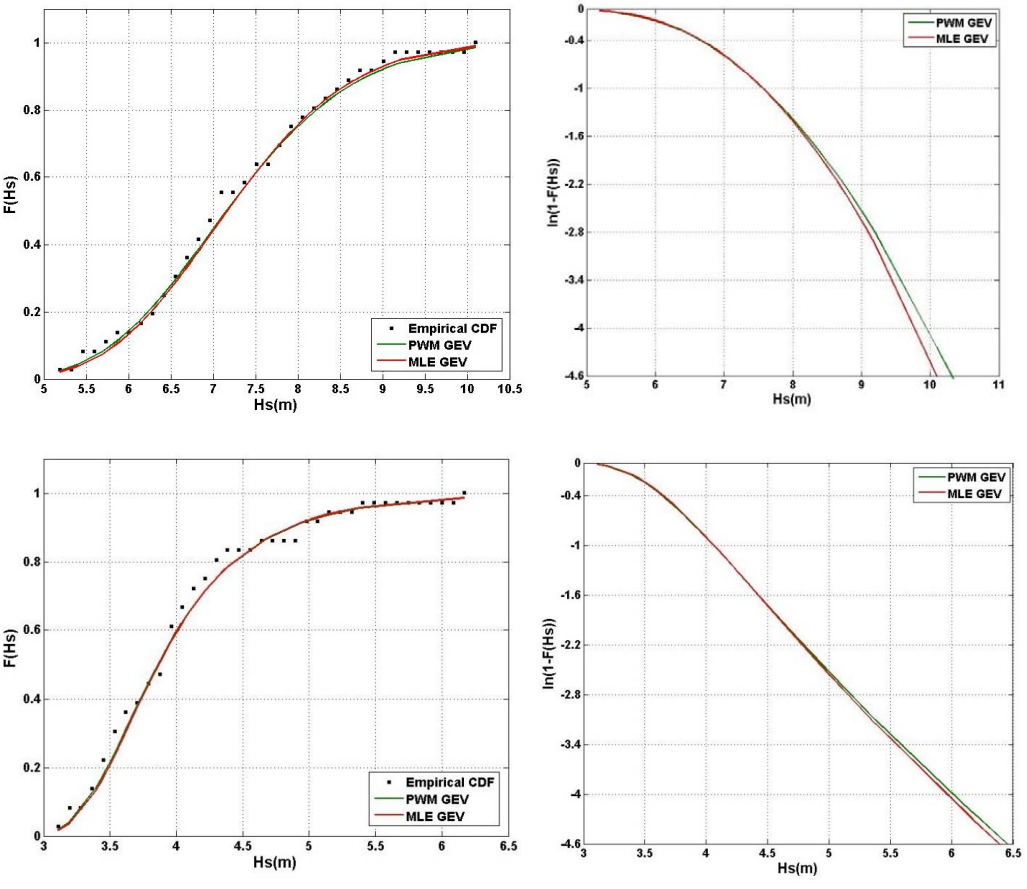


646

647 **Figure 2:** Typical representation of actual storm and associated ETS.

648

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

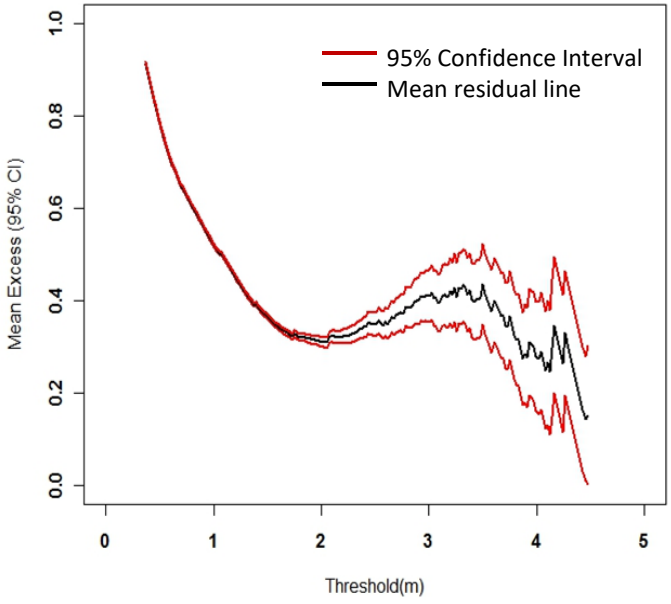


(a) **(b)**

Figure 3:(a) Comparison of GEV model CDF to the empirical CDF for NOAA44005 and ERA IN-4 **(b)** Variation of tail GEV model CDF in logarithmic coordinates for NOAA44005 and ERA IN-4

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions



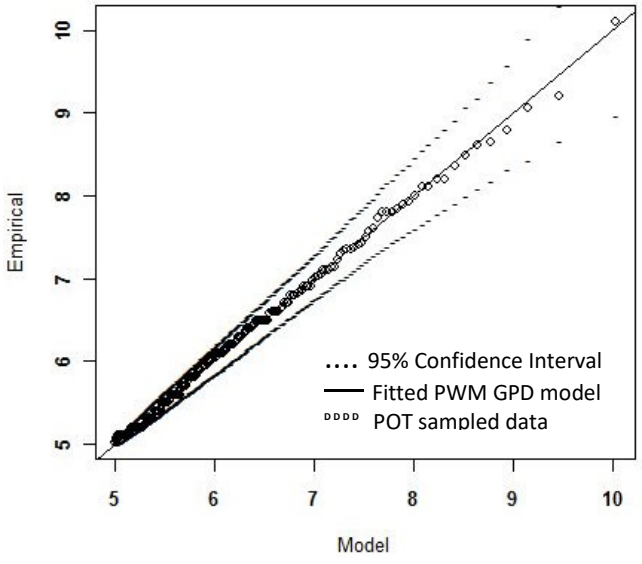
655

656 **Figure 4:** Mean Residual plot of ERA IN-1 with 95% confidence limits

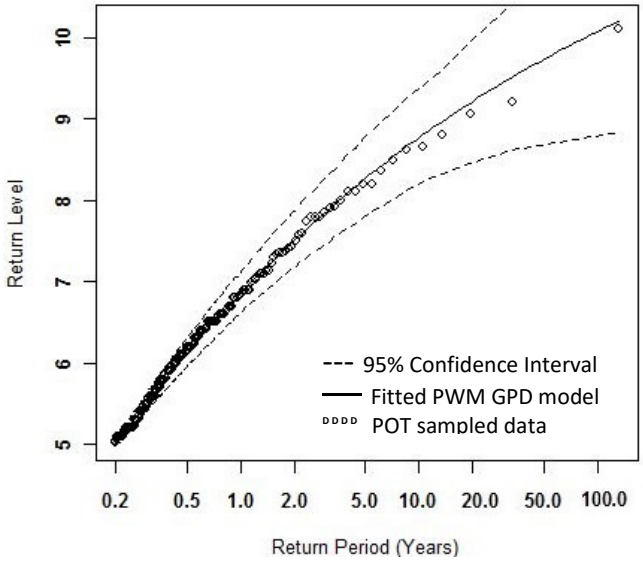
657

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions



(a)

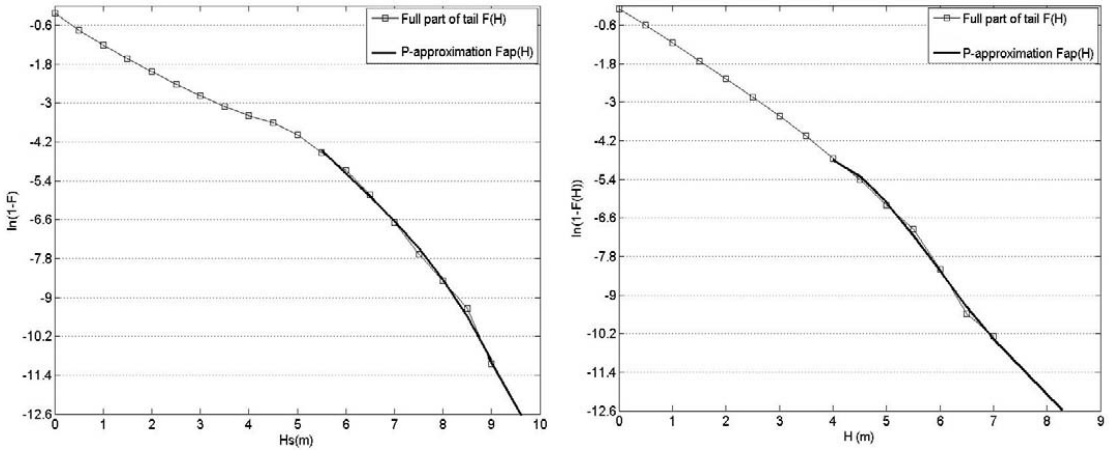


(b)

Figure 5 :(a) Quantile-Quantile plots of GPD model for NOAA44005 data **(b)** Return level plots of GPD model for NOAA44005 data.

Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-333, 2016
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Published: 24 October 2016

Natural Hazards and Earth System Sciences
Discussions



664

665 **Figure 6**: P-approximation for series of wave heights *Hs* at Alghero buoy station for Buoy
666 and ERA-interim datasets