

## **STATISTICAL ANALYSIS ON EXTREME WAVE HEIGHT**

N.V. Teena<sup>1</sup>, V. Sanil Kumar<sup>1</sup>, K.Sudheesh<sup>1</sup>, R. Sajeev<sup>2</sup>

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

<sup>2</sup>Department of Physical Oceanography, Cochin University of Science and Technology, Kochi, India - 682 016 India

### **ABSTRACT**

The classical extreme value theory based on generalized extreme value (GEV) distribution and generalized Pareto distribution (GPD) are applied to the wave height estimate based on wave hindcast data covering a period of 31 years for a location in the eastern Arabian Sea. Practical concern such as the threshold selection and model validation is discussed. Estimates of wave height having different return periods are compared with estimates obtained from different distributions. On comparing the distributions fitted to the GEV with annual maximum approach and GPD with peaks over threshold approach have indicated that both GEV and GPD models gave similar or comparable wave height for the study area since there is no multiple storm event in a year. Influence of seasonality on wave height having different return period is also studied.

Key words: Design wave height, peaks over threshold, generalized extreme value, generalized Pareto distribution, return period

## 1. INTRODUCTION

Ocean waves are of fundamental interest for engineering design and operation in coastal and offshore industry (Massel 1978). Characteristics of the ocean waves are also required for the design of ships and harbours and for prevention of coastal disasters. So it is important to consider the extreme wave for the safe design of ocean structures. Extreme wave analysis is the first step in coastal and offshore structure design for selecting design wave heights (Goda et al. 2010). In order to design a marine structure or a vessel, the designer must have the knowledge of extreme waves with return periods of 50 or 100 years. Estimation of design values contribute to the level of protection and the scale of investment. Thus understanding the extreme conditions of wave is of interest not only scientifically to the research community, but also economically to the government and industry. Extreme value analysis for waves is discussed in detailed by Goda (1992), Goda et al. (1993) and Mathiesen et al. (1994). Calculation of design wave height corresponding to a return period were done by many researchers (Petruaskas and Aagaard 1970; Ochi 1978; Carter and Challenor 1981; Isaacson and MacKenzie 1981; Muir and Shaarawi 1986; Goda 1992; Vledder et al. 1993; Dong and Ki-Cheon 2006). Ferreira and Guedes Soares (2000) suggested that the prediction of extreme values of significant wave height should rely on methods based on extreme value theory which makes use of the largest observation in the sample and whose models are justified by general theories. Coles (2001) has provided the statistical details of extreme value prediction based on the annual maximum (AM) and Peaks Over Threshold (POT) data points. Mendez et al. (2006) developed a time-dependent POT model for extreme significant wave height which considers the parameters of the distribution as functions of time. Menendez et al. (2009) studied the influence of seasonality in the estimation of wave height using time dependent GEV model. Edward et al. (2010) made use of discrete seasonal and directional models for the estimation of extreme wave conditions. Recently, the use of double threshold, with objective tools for determining high threshold for the estimation of extreme wave height was presented and justified (Franck and Luc 2011). Even though several studies are carried out for the improvement in estimating the extreme wave height by different statistical methods, the dilemma in extreme wave height analysis lies on whether to use GEV or GPD series. Hence a study is conducted to compare the significant wave height estimate for different return period using GEV and GPD series. Similar work has been carried out by Palutikof et al. (1999) on wind speed, Mkhandi et al. (2005) on flood frequency, Lin (2010) on wind gust and Emanuel and Jagger (2010) on hurricane data. Due to increase in the offshore oil and gas activities along the coast of India, several coastal and offshore facilities are planned. Hence, the relevance of study is to

understand the variability in the extreme wave height estimates using different distributions. The west coast of India is influenced by monsoon and hence there is a seasonal variations in the wave activity (Kumar et al., 2006). Hence the influence of seasonality on the wave height with different return period is also examined.

## 2. METHODOLOGY

The estimates for design wave heights are based on two closely related families of extreme value distributions; i) Generalized Extreme Value (GEV) family (Von Mises 1936) and ii) Generalized Pareto (GP) family (Coles 2001); where the type of data extraction determines the family to be applied.

### Generalized Extreme Value (GEV)

Classical extreme value theory explains how a sufficiently long sequences of independent and identically distributed random variables, the maxima of samples of size  $n$ , can be fitted to one of three basic families. These three families were combined into a single distribution by Von Mises (1936) and by Jenkinson (1955) now universally known as the generalized extreme value (GEV) distribution. This has the cumulative distribution function (CDF) as:

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

Where  $k$  is the shape parameter which determines the type of distribution and also reflects the fatness of tails of the distribution (the higher value of this parameter, the fatter tails);  $\alpha$  is the scale parameter which gives the degree of spread along abscissa and  $\beta$  is the location parameter which locates the position of density function along abscissa; then, if  $k = 0$ ; Fisher-Tippett Type I distribution or Gumbel distribution, if  $k > 0$ ; Fisher-Tippett Type II distribution or Frechet distribution, and if  $k < 0$ ; Fisher-Tippett Type III distribution or Weibull distribution.

The quantile value  $X_T$ , is the wave height which is exceeded, on average, once every return period of  $T$  years, which is given for GEV as (Palutikof et al. 1999):

$$X_T = \begin{cases} \beta + \frac{\alpha}{k} \left\{ 1 - \left[ -\ln \left( 1 - \frac{1}{T} \right) \right]^k \right\} & k \neq 0 \\ \beta - \alpha \ln \left[ -\ln \left( 1 - \frac{1}{T} \right) \right] & k = 0 \end{cases} \quad (2)$$

### Generalized Pareto distribution (GPD)

A major concern of traditional extreme value theory is that it only considers a single maximum within each epoch. This approach ignores other extreme events that may have occurred in each epoch. An alternative approach, known as the Peaks Over Threshold (POT) approach, considers all values greater than a given threshold value. POT approach enables the analyst to use all the data exceeding a sufficiently high threshold, compared to the classical extreme value analysis which typically uses annual extreme values. Given a threshold  $u$ , the distribution of excess values of  $x$  over  $u$  is defined by:

$$F_u(y) = Pr\{x - u \leq y | x > u\} = \frac{F(x) - F(u)}{1 - F(u)} \quad (3)$$

which represents the probability that the value of  $x$  exceeds  $u$  by at most an amount  $y$ , where  $y = x - u$ .

Balkema and de Haan (1974) and Pickands (1975) showed that for a sufficiently high threshold,  $u$ , the distribution function of the excess,  $F_u(y)$ , converges to the generalized Pareto distribution (GPD) which has a CDF as:

$$G(x; k, \alpha, \beta) = \begin{cases} 1 - \left( 1 - k \frac{x - \beta}{\alpha} \right)^{1/k} & k \neq 0 \\ 1 - e^{-(x - \beta)/\alpha} & k = 0 \end{cases} \quad (4)$$

when  $k = 0$ , the GPD corresponds to an exponential distribution (medium size tail); when  $k < 0$ , it takes the form of the ordinary Pareto distribution (long tailed); and if  $k > 0$ , it is known as a Pareto II type distribution (short tailed).

## Threshold selection

The choice of threshold is an important practical problem, which is mainly based on a compromise between bias and variance. The threshold must be high enough for the excess over the threshold to converge to the GPD, meantime the sample size should be large enough to ensure that there is enough data points left for satisfactory determination of the GPD parameters (Abild et al. 1992).

An important property of the GPD is that if  $k > -1$ , then the conditional mean exceedance (CME) over a threshold,  $u$ , is a linear function of  $u$ :

$$E[x - u \mid x > u] = \frac{\alpha - ku}{1+k} \quad (5)$$

The conditional mean exceedance (CME) graphs, also known as mean residual life graphs, plots the mean excess over threshold as a function of threshold (Davison 1984). For a GPD distributed spectrum, the CME graph should plot as a straight line. The linearity of the CME plot can thus be used as an indicator of the appropriateness of the GPD model.

The sample mean excess (SME) is an empirical estimate of the CME defined as

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n N_{i\{X_i > u\}}} \quad (6)$$

i.e. the sum of the excesses over the threshold  $u$  divided by the number of data points which exceed  $u$  and which can be used to chose as an appropriate threshold by selecting the lowest value above which the graph is approximately a straight line.

The quantile estimation for Generalized Pareto Distribution (GPD) for threshold value  $u$ , was given by Abild et al. (1992) for probability weighted moment (PWM), method of moment (MOM) and maximum likelihood (ML) method where  $k$  and  $\alpha$  are the fitted GPD parameters to  $x-u$ , where  $x > u$ ,

$$X_T = \begin{cases} k + \frac{\alpha}{k} [1 - (\lambda T)^{-k}] & k \neq 0 \\ k + \alpha \ln(\lambda T) & k = 0 \end{cases} \quad (7)$$

Here,  $\lambda$  is  $N_u/N$  where  $N_u$  is the total number of exceedances over the selected threshold  $u$ , and  $N$  is the number of years of record.

The routines in WAFO (2000) developed by faculty of Engineering, Mathematical Statistics, Lund University, Sweden, were used for fitting different statistical distribution. WAFO is a MATLAB (<http://www.mathworks.com>) toolbox for analysis of random waves and loads, which has a sub-toolbox of extreme value analysis.

### Data used

The simulated significant wave height ( $H_s$ ) based on MIKE-21 SW, a Spectral Wave model developed by DHI Water & Environment, Denmark (DHI 2009) was used in the study. Details of the model setup used in the study are available in Aboobacker et al. (2011). The significant wave height was obtained for every 1-hour interval at 15 m water depth off Honnavar, central west coast of India. The data set contains a total of 277,000 records representing sea states at 1 hour interval covering 31 years. The estimated wave height based on model was found to compare well with the buoy measurements at Honnavar during 2009 with a correlation coefficient ( $r$ ) value of 0.8 (**Fig. 2**). Information on storm events were taken from Joint Typhoon warning centre ([http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/best\\_tracks/ioindex.html](http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/best_tracks/ioindex.html)) and Indian Daily Weather Reports of India Meteorological Department. For non-cyclone period, reanalysis wind data from NCEP/NCAR database (Kalnay et al. 1996) has been used (<http://www.cdc.noaa.gov>) to arrive at the maximum wind speed. For cyclonic period, the wind speeds from the NCEP/NCAR data was replaced with the cyclone data. The yearly maximum  $H_s$  for 31 year period show that the maximum significant wave height of 5.53 m was estimated in 1996 (**Fig. 3**) and was during a storm event. During the same event,  $H_s$  of 5.69 m (Kumar 2006) was measured by a Datawell directional waverider buoy moored at 23 m water depth off Goa,

which is 125 km north of the present study location. This indicates that the model estimated  $H_s$  values are within an error of 3%.

### 3. RESULTS

#### 3.1 GEV METHOD

For generalized extreme value analysis, Annual Maximum (AM), Monthly Maximum (MM) and Weekly Maximum (WM) were extracted from the 1-hour interval wave data. The parameter estimation is done by PWM and ML method for these datasets. The PWM method yields a convenient and powerful test for GEV distribution (Hosking et al. 1985). Also it works for a wider range of parameter values than ML method (Hosking 1990). Besides if  $k$  lies between  $-1$  and  $-0.5$ , the maximum likelihood estimators are generally obtainable, but do not have the standard asymptotic properties and if  $k < -1$ , maximum likelihood estimators are unlikely to be obtainable. So here GEV parameters cannot be estimated accurately by ML method.

In order to check how different distributions fit to the data, the statistical parameters such as coefficient of determination ( $R^2$ ) and root mean square error (RMSE) were estimated (**Table 1**). The coefficient of determination will range from 0 to 1 and values close to 1 indicate better fit. RMSE is used as a measure of the differences between values predicted by a model or an estimator and the values actually observed. Lower the value of RMSE indicate better fit. The cumulative distribution and the quantile estimation indicate (**Table 1**) that GEV (AM) distribution fitted the data for the present location better than GPD (POT).

For the west coast of India, the significant wave height will be less than 1.5 m during the non monsoon months (Kumar and Anand 2004). Hence the return value plot for monthly maximum will be steeper than annual maximum. Consequently, it will underestimate the initial range of return periods and further leading nearer to the expected values. Hence the RMSE is higher for distribution based on monthly maximum and weekly maximum (**Table 1**) compared to AM and the  $R^2$  value is lower. So the wave height estimated for different return period based on monthly maximum dataset will not provide reliable value for the present location. Using the weekly maximum dataset, the estimated wave height having 10 year return period is 2.8 m (**Table 2**) which is less than the minimum value of annual maximum data

i.e., 3.2 m. So the wave heights estimated for different return period based on weekly maximum is less and will not be reliable. So we prefer to use only annual maximum dataset for further study.

The fitted GEV model uses parameters estimated using the PWM for AM and are plotted against the wave height as a CDF plot (**Fig. 4a**). The sorted data (ordered smallest to largest where  $x(1)$  is the smallest data value and  $x(N)$  is the largest data value) are plotted against the empirical cumulative probability distribution value. In GEV, probability weighted moments method performs favorable for negative shape parameter. The estimated shape parameter is negative indicating that the parent distribution is Type III distribution (Weibull) which is also the most favored distributions in coastal and ocean engineering (Goda et al. 2010).

An important issue when considering any model of a complex physical process is the reliability and appropriateness of the chosen model. Graphical methods are commonly used in model validation. A CDF plot of the fitted GEV/GPD and the empirical distribution derived from the data show the degree of these distribution fitting towards the data. The Quantile-Quantile (QQ) plot is a convenient visual diagnostic tool to examine whether a sample comes from a specific distribution. Specifically, the quantiles of an empirical distribution is plotted against the quantiles of a hypothesized distribution and in our case, the fitted GEV/GPD. The quantile estimates (**Fig. 4b**) utilized for distribution fit and assessment is the QQ plot which indicates rather a good fit. Altogether the fitness of the model is quite good with higher  $R^2$  value for GEV. Thus the estimation of design wave height for a return period provided by GEV holds good.

### 3.2 POT APPROACH

Although the POT method fundamentally assists with selecting an independent identically distributed data set, it is a good practice to ensure that this condition is satisfied. So the dataset for POT were extracted by assigning a minimum separation time to ensure the independence of extremes. Using a separation of 48 hours it extracts nearly independent POT extremes from the data series (Cook 1985; Gusella 1991). In POT method, the selection of a suitable  $H_s$  threshold value is key in achieving a robust data set. An approach was done to choose the threshold as the smallest value at which the GPD fitted the peak excesses reasonably. So here, Sample Mean Excess (SME) plot was plotted for excesses over the threshold using equation (6) for the selection of appropriate threshold. In the sample mean excess



plot (**Fig. 5a**), it can be seen that the graph appears to have two slopes with a major transition at a threshold of about 3.5 m. However, the parameters of Generalised Pareto distribution,  $k$  and  $\alpha$  vary dramatically above a threshold of 3.5 m indicating that this may be due to a reduced sample size resulting in greater scatter. In addition, the distribution parameters are approximately constant above a threshold of 3 m (ignoring the data above 3.5 m) and hence 3 m is the threshold used for the analysis. CME plot (**Fig. 5b**) clearly shows the linear graph which indicates that the selected data converges to GPD model.

As PWM estimators is utilized where computability was an issue (Hosking 1990) and similarly MOM approach does not lead to good estimators due to bias in higher order. So the MOM estimators are only used as a check on other methods or when other methods failed (Rice 1988). Maximum Likelihood (ML) 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 (Rice 1988). So the approach chosen here was to utilize a variety of techniques like PWM, MOM and ML for exploratory fitting for the probability model and then choose the best fit parameters based on a least square fit.

For parameter estimation, the CDF plots of exceedances and the fitted GPD using exceedances over threshold are done (**Figs. 6a, 6b and 6c**) and which indicates reasonably good fit for PWM, MOM and ML methods (**Table 1**). The shape parameter reveals that these corresponds with an exponential distribution in GPD. The model validation for all the methods (PWM, MOM and ML) includes the QQ plots (**Figs. 6d, 6e and 6f**) which also suggest a good fit. The comparison indicate that the quantiles lies close to the exact match line except few points. The statistical parameters indicate (**Table 1**) that the data fits well to the GPD distribution.

The fitted GEV by PWM and GPD by PWM, MOM and ML methods are used to estimate wave height quantiles for return periods of 10, 50 and 100 yr using equations (2) and (7) respectively (**Table 2**). Among the different dataset in GEV, annual maximum shows the highest value of 6 m followed by monthly maximum. The monthly and weekly maximum dataset estimates lower design values compared to annual maximum, as for these, the effect of extreme wave height during storm is reduced by the total number of observations in each datasets.

The wave height estimated for 100 year return period based on different methods of both GEV and GPD show that the variation in wave height upto 9% (**Table 2 and Fig. 7**). The wave height estimated for 100 yr return period is 6 m based on PWM method for GEV followed by 5.8 m using PWM method in GPD for the region. Even though the initial quantiles were closer, further on the return value plot of GEV becomes steeper than GPD and thus yielding slightly higher (3%) wave height estimate by the GEV.

In case of GPD, among PWM, MOM and ML methods, PWM has estimated  $H_s$  of 5.8 m followed equally (5.5 m) by MOM and ML. Thus PWM were more reliable than the ML estimates similar results were found by Hosking and Wallis (1987). All the methods of GPD with different values of shape parameter have a comparable performance because the sample size is large. Similar inferences have been drawn by Moharram et al. (1993).

### 3.3 COMPARISON OF GEV AND GPD METHODS

For comparison between GEV (AM) and GPD (POT) by PWM method, the empirical and theoretical distributions fitted to the AM and POT were plotted on the same scale. It is observed that the initially upto 7 years return period, the wave height estimates by GEV was less than GPD and later the values estimated following GPD are higher than that based on GEV. Both GEV (AM) and GPD (POT) gave comparable (within 3%) estimate for wave height (**Fig. 8**). This also has been proven in a theoretical way by Langbein (1947) and Chow et al. (1988). In the study area, the occurrence of multiple storm event is absent and hence the wave height based on AM and POT gave comparable estimates. Hence for the study area either technique could be employed for the estimation of wave height for different return periods. As the AM series considers only the highest values in the year, this might lead to the slight overestimation of AM based approach in comparison with the POT based analysis in areas where multiple storms occur.

### 3.4 INFLUENCE OF SEASONALITY

Waves along the west coast of India is influenced by summer (southwest) monsoon (Kumar et al. 2012). Hence an attempt was made to provide further insight into the seasonal influence on the wave height having different return period by using GEV method considering seasonal maximum data. On comparing the estimated wave height based on season data (**Fig. 9**), it was found that the wave height estimated based on the monsoon season are higher than the non monsoon months since the extreme

wave heights occur during the monsoon season. The annual maximum wave height along the west coast of India occur during the southwest monsoon period and hence the wave height estimated using annual maximum and the monsoon maximum are same. The 100 year return period wave height based on monsoon data is 1.55 times the value based on pre-monsoon data and is 2 times the value based on post-monsoon data.

#### 4. CONCLUSION

A particular type of distribution cannot be adopted as a common extreme wave distribution because of different meteorological and geographical condition at a particular location. The study shows that as the sample size was large, the wave height estimate using GPD based on all the methods (PWM, MOM and ML) have a comparable performance. Among the GPD, the estimate based on PWM proves to be more reliable than that based on ML. GEV (AM) distribution fitted the data better than GPD (POT).

On comparing the wave height values for different return periods estimated based on different distributions, it is found that the values estimated based on GEV (AM) was slightly larger than those from the GPD (POT) and this agrees with the report of Lin (2010). The wave height estimated for 100 year return period based on different methods considering GEV and GPD distribution show that the variation in wave height was within 9%. GEV (AM) and GPD (POT) distributions gave similar or comparable (within 3%) estimate for wave height for west coast of India since the occurrence of multiple storm event was absent in this region. In locations where multiple storms can occur, POT approach is favoured as the datasets includes all the significant storm events. The 100 year return period wave height based on monsoon data is 1.55 times the value based on pre-monsoon data and is 2 times the value based on post-monsoon data.

#### ACKNOWLEDGEMENT

The present study was part of the project on shoreline management plans for Karnataka coast. We thank Integrated Coastal and Marine Area Management Project Directorate (ICMAM PD), Ministry of Earth Sciences, New Delhi for funding the project. Director, National Institute of Oceanography, Goa and Director, ICMAM PD, Chennai provided encouragement to carry out the study. This is NIO contribution xxxx.

## REFERENCES

- Abild J, Andersen EY and Rosbjerg D (1992) The climate of extreme winds at the Great Belt. Denmark J Wind Eng Ind Aerodyn. 41–44: 521–532.
- Aboobacker VM, Rashmi R, Vethamony P, Menon HB (2011) On the dominance of pre-existing swells over wind seas along the west coast of India. Cont Shelf Res 31, 1701–1712.
- Balkema A and de Haan L (1974) Residual life time at great age. Ann Prob. 2:792–804.
- Carter DJT, and Challenor PG (1981) Estimating Return Values of Environmental Parameters. Quart Jour Roy Meteorol Soc Vol 107:259-266.
- Chow VT, Maidment DR and Mays LW (1988) Applied hydrology. McGraw-Hill, New York.
- Coles S (2001) An Introduction to Statistical Modeling of Extreme Values. Springer series in statistics, Springer-Verlag 208 p.
- Cook NJ (1985) The Designer’s Guide to Wind Loading of Building Structures Part 1: Background. Damage Survey, Wind Data and Structural Classification Building Research Establishment, Garston and Butterworths London, 371 pp.
- Davison AC (1984) Modelling excesses over high thresholds with an application. In Statistical Extremes and Applications J Tiago de Oliveira, Editor pp 461–482.
- DHI (2009) MIKE 21 User Manual, DHI Water & Environment, Danish Hydraulic Institute, Denmark.
- Dong Y Lee and Ki-Cheon J (2006) Estimation of Design Wave Height for the Waters around the Korean Peninsula”, Ocean Science Journal Vol. 41 No. 4: 245-254.
- Edward BLM, Peter GC and AbuBakr SB (2010) On the use of discrete seasonal and directional models for the estimation of extreme wave conditions. Ocean Engineering 37: 425–442.
- Emanuel K and Jagger T (2010) On Estimating Hurricane Return Periods. Journal Of Applied Meteorology And Climatology, Vol 49:837-843.
- Ferreira JA and Guedes SC (2000) Modelling distributions of significant wave height. Coastal Engineering 40: 361-374.
- Franck M and Luc H (2011) A multi-distribution approach to POT methods for determining extreme wave heights. Coastal Engineering 58:385–394.
- Goda Y (1992) Uncertainty of design parameters from viewpoint of extreme statistics. Journal of Offshore Mechanics and Arctic Engineering ASME Vol.114:76-82.
- Goda Y, Hawkes P, Mansard E, Martin MJ, Mathiesen E, Peltier E, Thompson E, and Van Vledder G (1993) Intercomparison of extremal wave analysis methods using numerically simulated data. Proc. 2nd Int. Symp On Ocean Wave Measurement and Analysis ASCE New Orleans pp 963-977.

- Goda Y, Kudaka M and Kawai H (2010) Incorporation of Weibull distribution in L-moments method for regional frequency of peaks-over-threshold wave heights. Proceedings of 32<sup>nd</sup> International Conference on Coastal Engineering, ASCE in press.
- Gusella V (1991) Estimation of extreme winds from short term records. *J Struct Eng* 117: 375–390.
- Hosking JRM (1990) L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics. *J Royal Statistical Society Ser B, London*, Vol 52:105-124.
- Hosking JRM, Wallis JR and Wood EF (1985) Estimation of the Generalized Extreme Value Distribution by the Method of Probability Weighted Moments. *Technometrics* Vol. 27: 251-261.
- Hosking JRM and Wallis JR (1987) Parameter and quantile estimation for the generalised Pareto distribution. *Technometrics* Vol. 29: 239-349.
- Isaacson MQ and MacKenzie NG (1981) Long-term Distributions of Ocean Waves: a Review. *J Waterway, Port, Coastal and Ocean Division ASCE* Vol 107(WW2): 93-109.
- Jenkinson AF (1955) The Frequency Distribution of the Annual Maximum (or Minimum) of Meteorological Elements. *Quarterly Journal of the Royal Meteorological Society*, 81, 158-171. (1969) *Statistics of Extremes*. Technical Note 98, World Meteorological Office, Geneva.
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell S, Saha G, White J, Woollen Y, Zhu A, Leetmaa B, Reynolds M, Chelliah W, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, and Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* 77 (3): 437-471.
- Kumar VS, Anand NM (2004) Variations in wave direction estimated using first and second order Fourier coefficients. *Ocean Engineering* 31: 2105–2119.
- Kumar VS (2006) Variation of wave directional spread parameters along the Indian coast. *Applied Ocean Research* 28:93-102.
- Kumar VS, Pathak KC, Pednekar P, Raju NSN, Gowthaman R, (2006) Coastal processes along the Indian coastline, *Current Science* 91; 530–536.
- Kumar VS, Johnson G, Dora GU, Philip SC, Singh J, Pednekar P, (2012) Variations in nearshore waves along Karnataka, west coast of India, *Journal of Earth Systems Science* 121: 393-403.
- Langbein WB (1947) Annual floods and the partial duration flood series. *Transactions of the American Geophysical Union*, vol. 30 No. 6: 879-881.
- Lin XG (2010) Statistical Modelling of Severe Wind Gust. Risk Modelling Project, Minerals and Geohazards Division, GeoScience Australia, Canberra, Australia.
- Massel RS (1978) Needs for Oceanographic Data to Determine Loads on Coastal and Offshore Structures. *IEEE Journal of Oceanic Engineering* vol oe-3 no. 4.

- Mathiesen M, Hawkes P, Martin MJ, Thompson E, Goda Y, Mansard E, Peltier E and Van Vledder G (1994) Recommended practice for extreme wave analysis. *Journal of Hydraulic Research IAHR* 32 (6): 803-814.
- Mendez FJ, Menendez M, Luceno A, and Losada, IJ (2006) Estimation Of The Long-Term Variability Of Extreme Significant Wave Height Using A Time-Dependent Peaks Over Threshold (Pot) Model. *Journal Of Geophysical Research* Volume 111 Issue C7.
- Menendez M, Mendez JF , Izaguirre C, Luceno A and Losada I (2009) The influence of seasonality on estimating return values of significant wave height. *Coastal Engineering* 56: 211–219.
- Mkhandi S, Opere AO, and Willems P (2005) Comparison between annual maximum and peaks over threshold models for flood frequency prediction. *International Conference of UNESCO Flanders FIT FRIEND/Nile Project – Towards a better cooperation, Sharm-El-Sheikh, Egypt, CD-ROM Proceedings* 16 p.
- Moharram SH, Gosain AK, and Kapoor PN (1993) A comparative study for the estimators of the Generalized Pareto distribution. *Journal of Hydrology* 150: 169-185.
- Muir LR and El-Shaarawi AH (1986) On the Calculation of Extreme Wave Heights: a Review. *Ocean Engineering*, Vol 13, No 1: 93-118.
- Ochi MK (1978) On Long-term Statistics for Ocean and Coastal Waves. *Proc 16th Int Coastal Engg ASCE* pp 59-75.
- Palutikof JP, Brabson BB, Lister DH and Adcock ST (1999). A review of methods to calculate extreme wind speeds. *Meteorological Applications* 6(1): 119–132.
- Petruaskas C and Aagaard PM (1970) Extrapolation of Historical Storm Data for Estimating Design Wave Heights. 2nd OTC Paper No.1190 pp 1410-1420.
- Pickands J (1975) Statistical inference using extreme order statistics. *Ann Stat* 3: 119–131.
- Rice JA (1988) *Mathematical Statistics and Data Analysis*. Wadsworth, Belmont, CA.
- Vledder GV, Goda Y, Hawkes P, Mansard E, Martin MH, Mathiesen M, Peltier E and Thompson E (1993) Case Studies of Extreme Wave Analysis: a Comparative Analysis. *Proc 2nd Int Symp Ocean Wave Measurement and Analysis ASCE* pp 978-992.
- Von Mises R (1936) La distribution de la plus grande de n valeurs. Reprinted in *Selected Papers Volume II American Mathematical Society, Providence R I, 1954, pp 271-294*.
- WAFO (2000) – A MATLAB toolbox for analysis of random waves and loads, Lund University, Sweden, homepage <http://www.maths.lth.se/matstat/wafo/,2000>.

**Table 1:** Statistical results of data and fitted distribution for cumulative distribution and quantile estimates

Distribution	Method	Datasets	cumulative distribution		Quantile estimate	
			R-square	RMSE	R-square	RMSE
GEV	PWM	Annual maximum	0.989	0.031	0.983	0.071
GEV	PWM	Monthly maximum	0.982	0.041	0.931	0.288
GEV	PWM	Weekly maximum	0.982	0.040	0.927	0.255
GPD	PWM	POT	0.995	0.018	0.965	0.063
GPD	MOM	POT	0.995	0.018	0.961	0.065
GPD	ML	POT	0.995	0.018	0.936	0.078

**Table 2:** Comparison of wave height for 10, 50 and 100 yr return period by different methods

Distribution	Method	Datasets	Wave height	Wave height	Wave height
			estimated for 10 yr return period (m)	estimated for 50 yr return period (m)	estimated for 100 yr return period (m)
GEV	PWM	Annual maximum	4.8	5.6	6.0
		Monthly maximum	3.4	5.0	5.9
		Weekly maximum	2.8	4.3	5.1
GPD	PWM	POT	4.4	5.5	5.8
	MOM	POT	4.4	5.2	5.5
	ML	POT	4.4	5.2	5.5

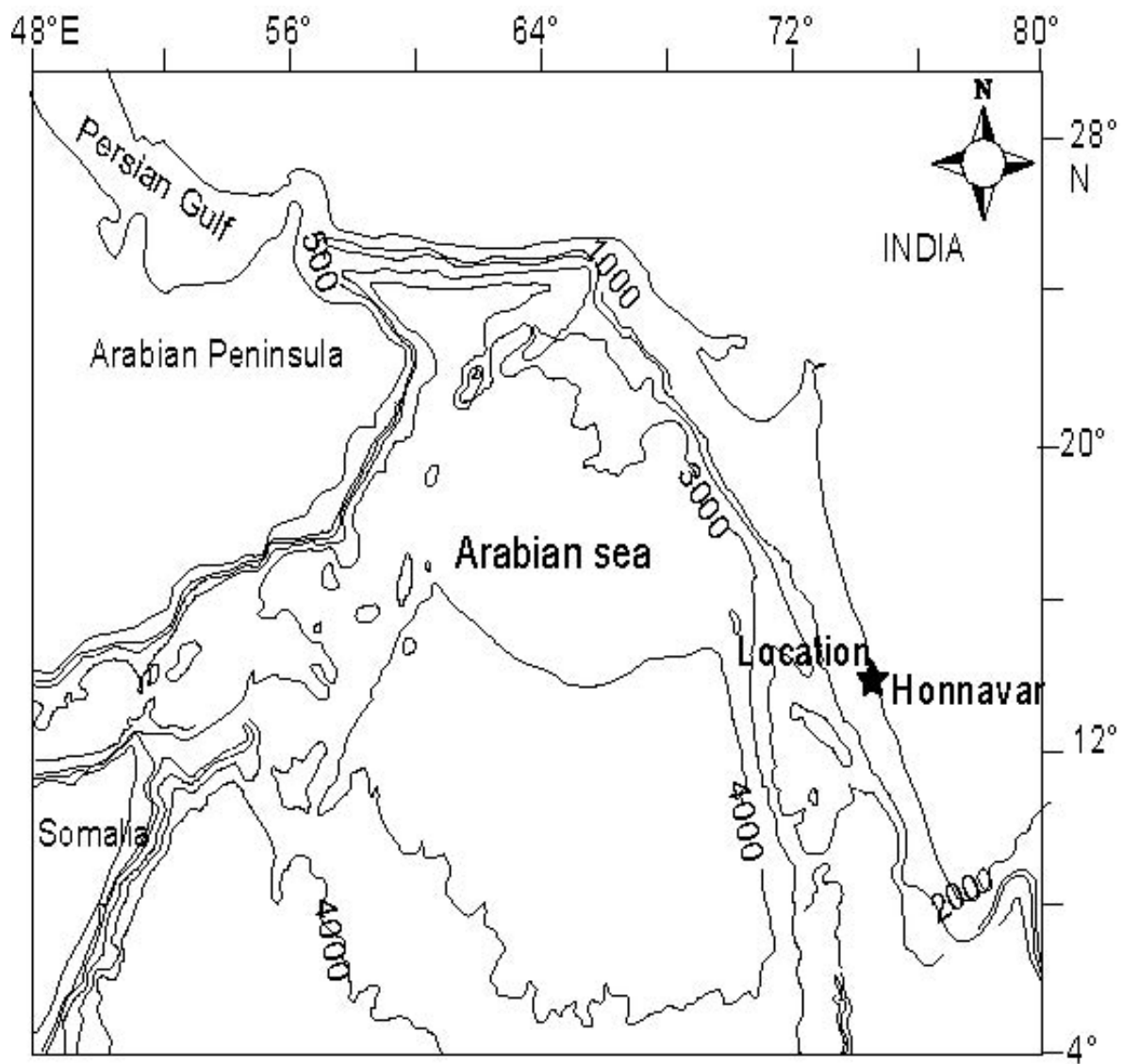
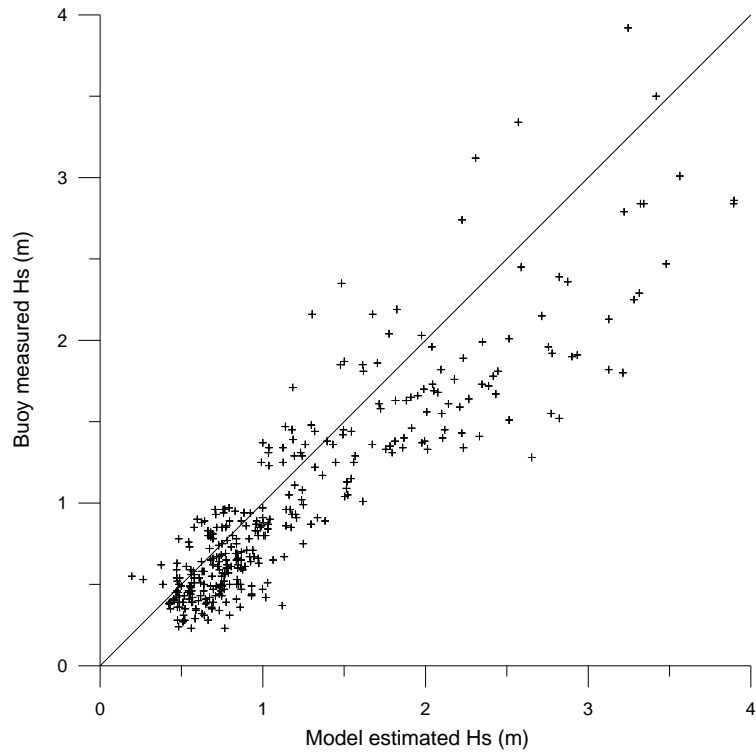
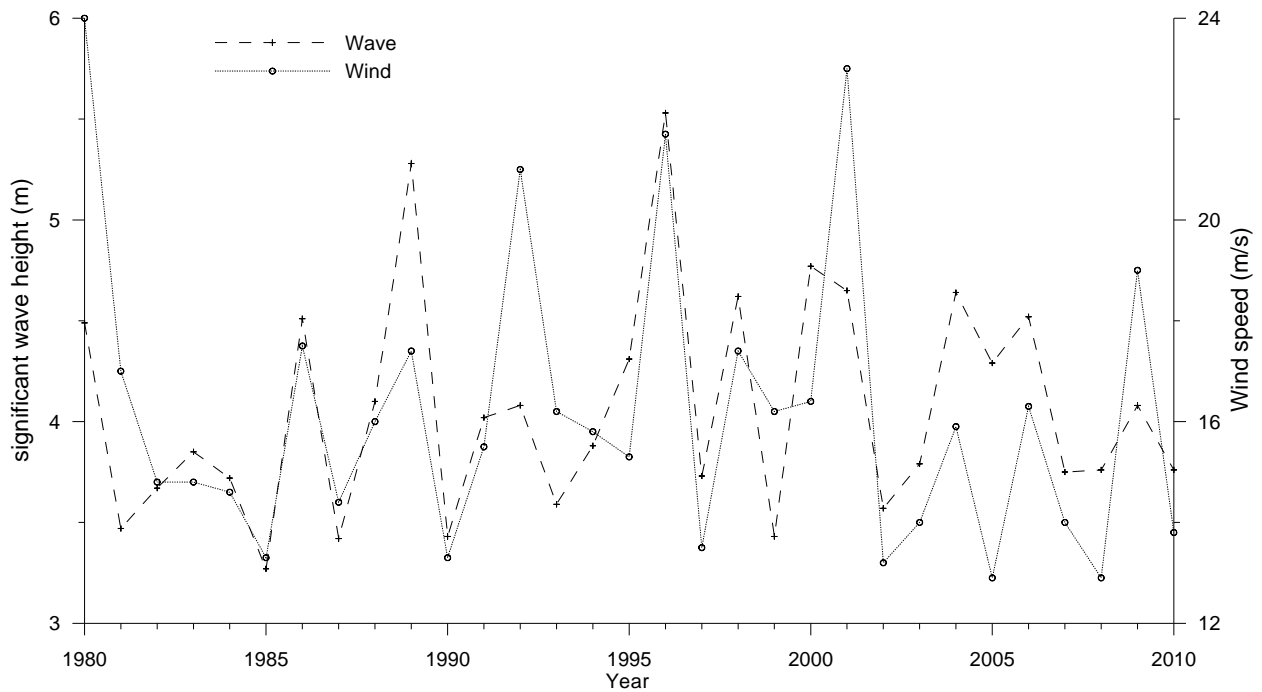


Fig. 1: The map showing the study location

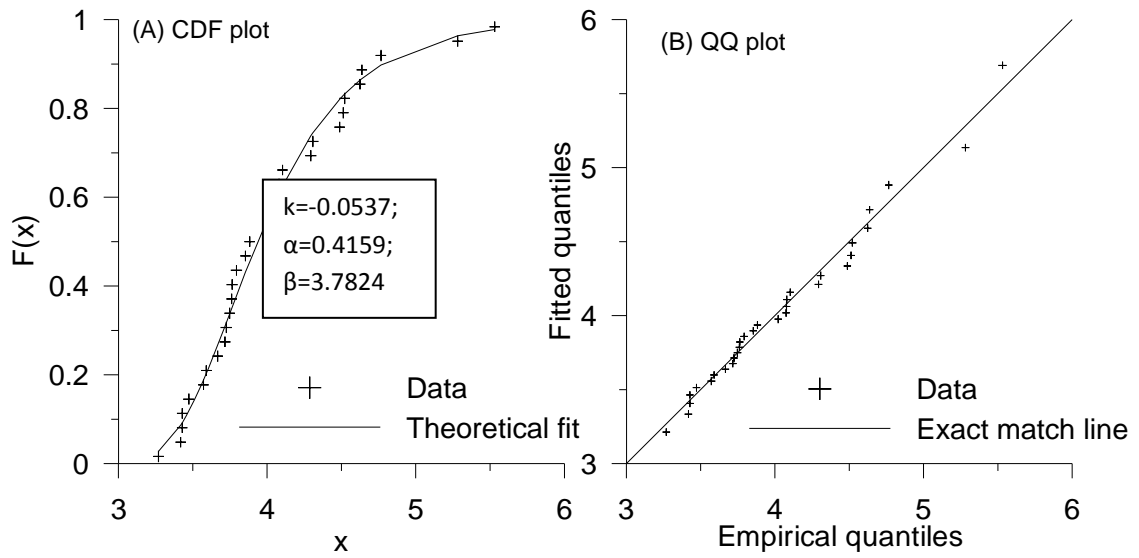




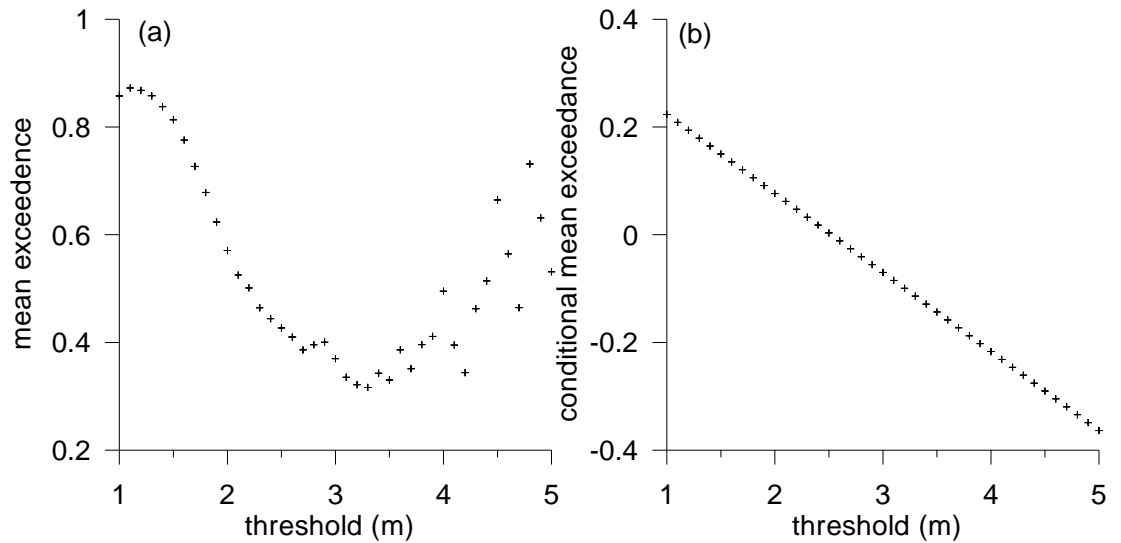
**Fig. 2:** Comparison between wave height measured by buoy and that estimated by the model



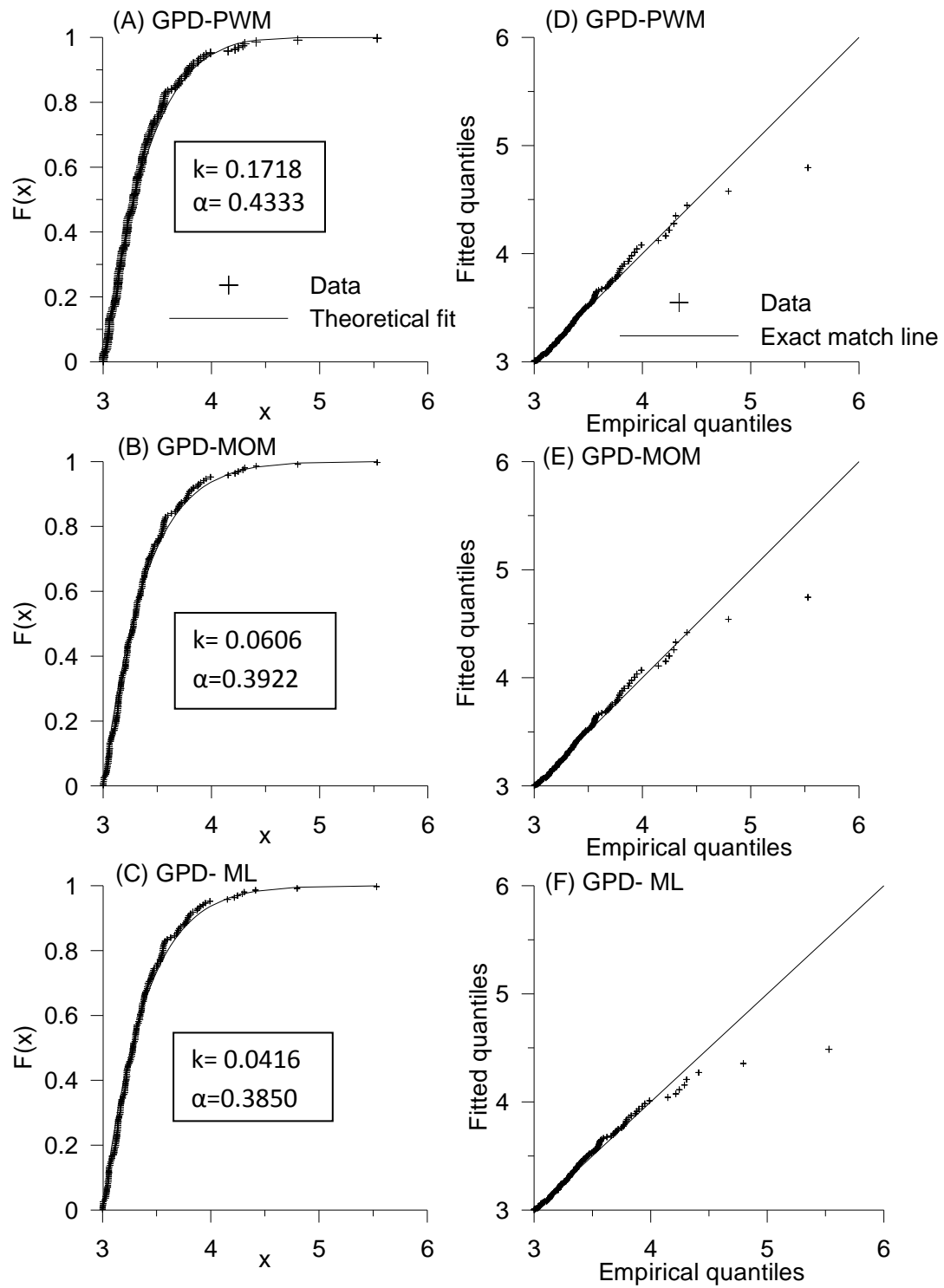
**Fig. 3:** Annual maximum significant wave height and maximum wind speed in different years



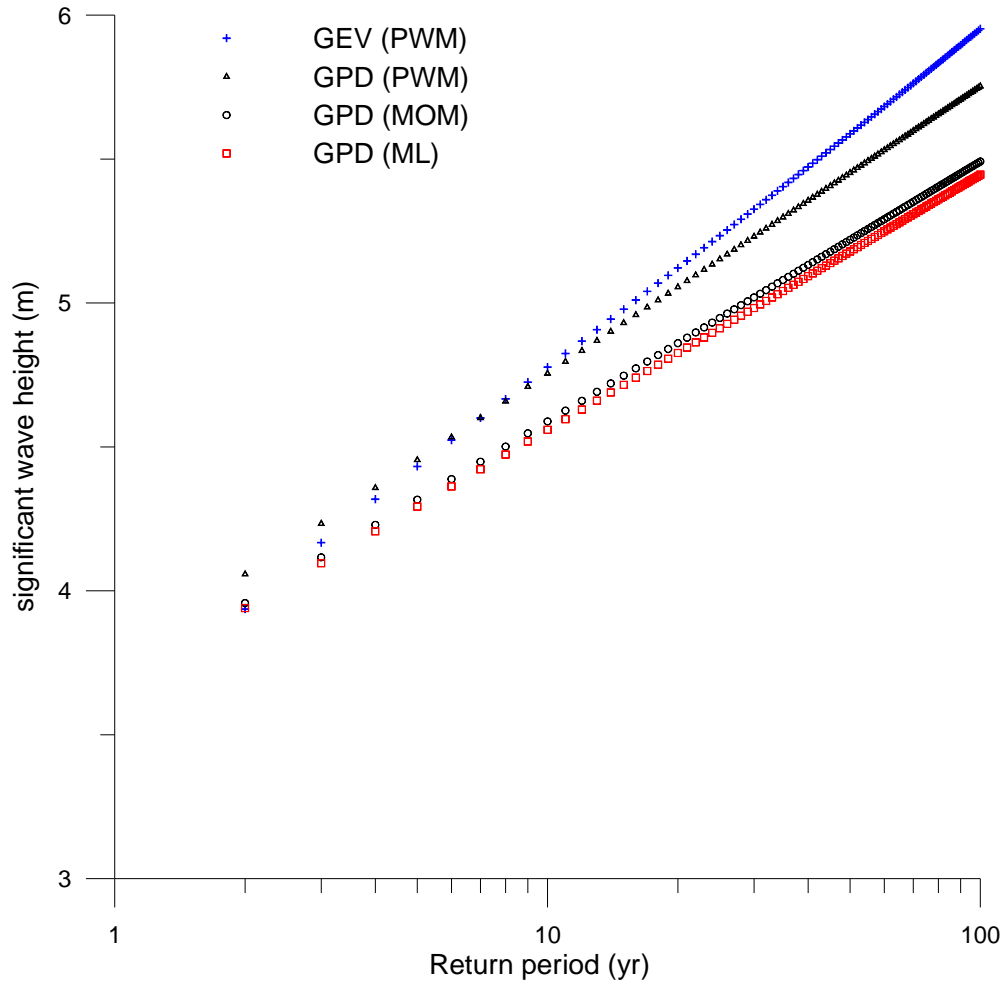
**Fig. 4:** a) CDF plot with fitted GEV using PWM method b) QQ plot for the fitted GEV.



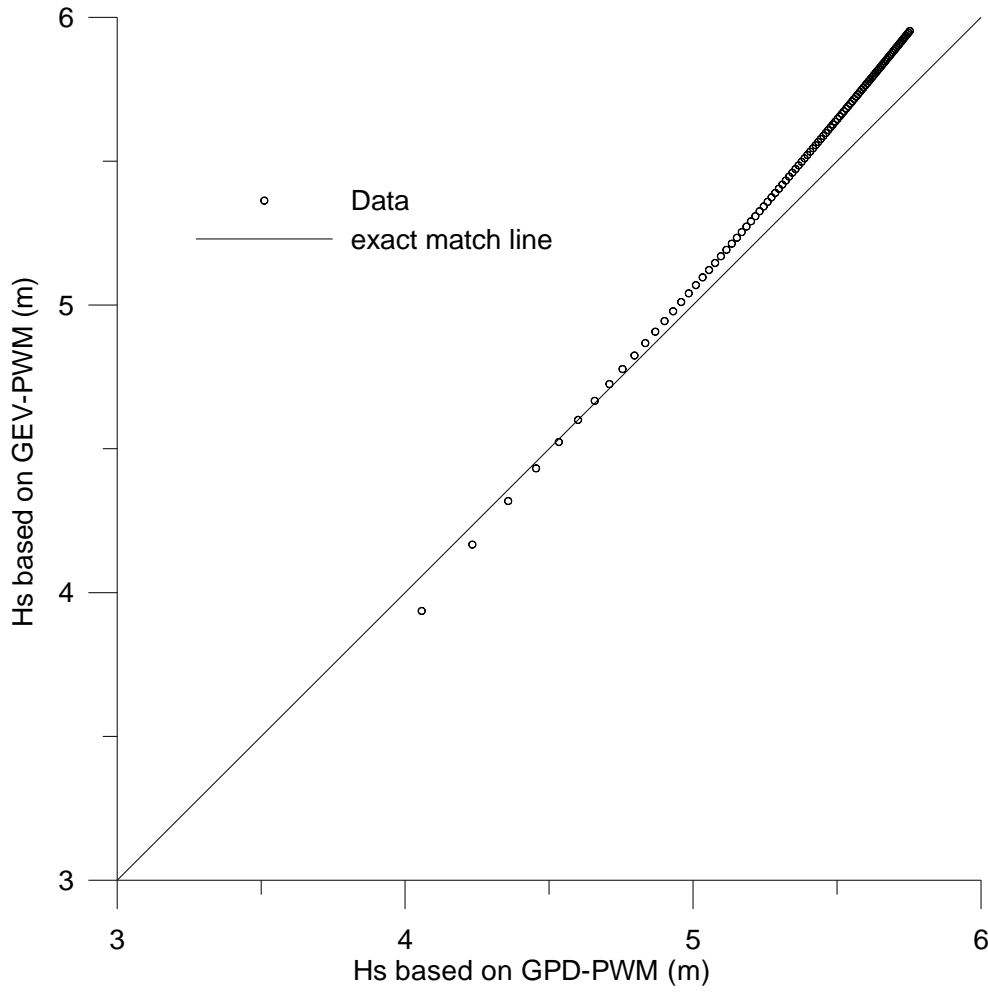
**Fig. 5:** a) Sample Mean Excess (SME) plot for appropriate threshold selection b) Conditional Mean Exceedances (CME) plot for the verification of GPD



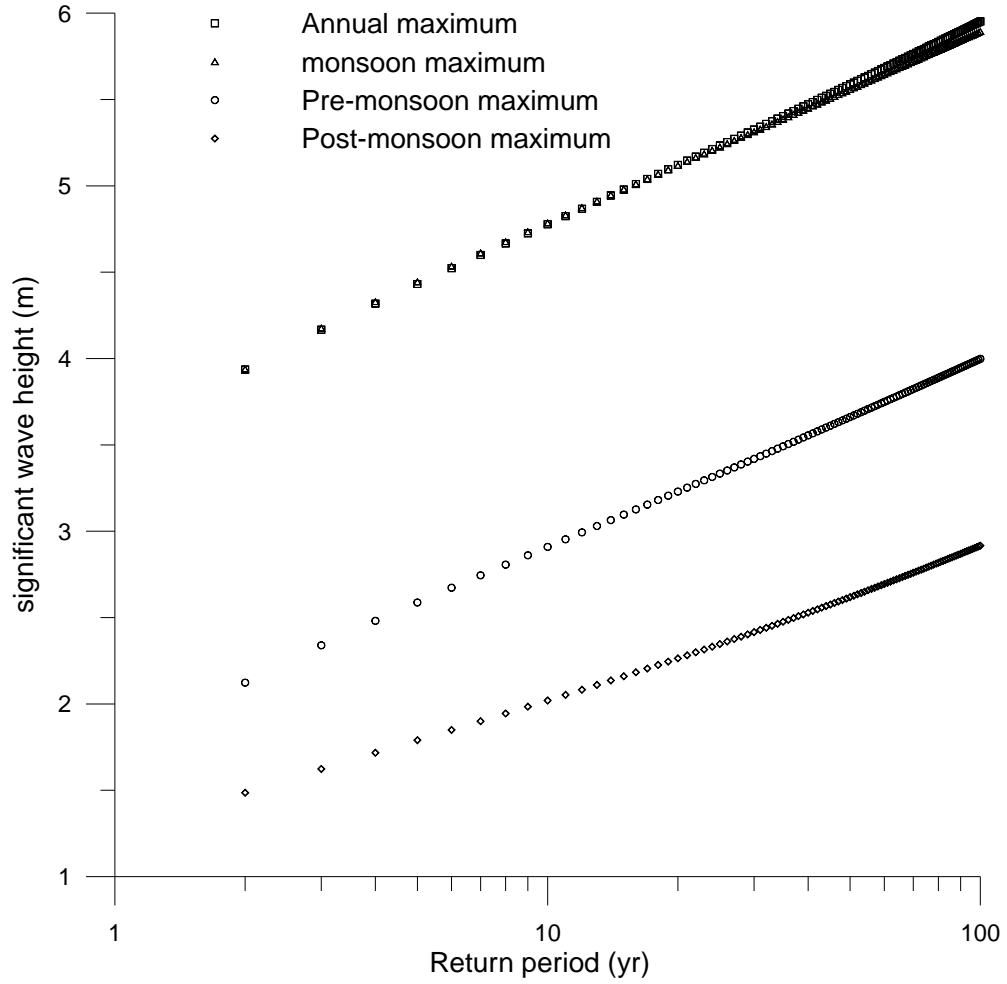
**Fig. 6:** CDF plot for GPD by (A) PWM, (B) MOM and (C) ML method and QQ plot for fitted GPD by (D) PWM, (E) MOM and (F) ML method



**Fig. 7:** Comparison between wave height estimated using different statistical methods for different return periods



**Fig. 8:** Comparison of wave height estimated using GEV and those estimated using GPD for different return periods



**Fig. 9:** Wave height having different return period based on annual maximum and maximum value during monsoon, post-monsoon and pre-monsoon using GEV