Estimating Return Values of Significant Sea Wave Heights in Colombo, Sri Lanka

S.H.Shajitha[1], and K.Perera[2]

[1]Postgraduate Institute of Science,
University of Peradeniya, Sri Lanka
[2]Department of Engineering Mathematics,
Faculty of Engineering, University of Peradeniya, Sri Lanka

Corresponding author's email: shajithash@gmail.com

Abstract

In coastal science and offshore engineering estimating the return values of significant extreme sea wave heights plays a vital role. This paper discusses two methods, Block Maxima (BM) approach and Peak-Over- Threshold (POT) method to estimate the 50 and 100 year return levels of extreme sea wave heights in Colombo. Generalized Extreme Value (GEV) distribution and Generalized Pareto (GP) distribution were fitted to the BM data and POT data respectively. Parameters of GEV and GPD were estimated with Maximum Likelihood estimation method. Gumbel distribution fitted well with the BM approach and GP distribution fitted well with the POT approach.BM method and POT method gives much more comparable results on estimating return values.

Keywords: return level, block maxima, threshold value, generalized extreme value distribution, generalized pareto distribution.

Introduction

The N-year return level of sea wave height is a statistic of vital important often used in the coastal and offshore engineering. Also the return values are very useful to describe the extremes of other environmental parameters. Commonly the return values are estimated by fitting the Generalized Extreme Value (GEV) distribution to blocked maxima (for instance: annual maxima) or by fitting the Generalized Pareto (GP) distribution to exceedances above a specific threshold value, (Davidson and Smith (1990), Coles (2001).

Many numbers of studies have been carried out to find the return levels of sea wave heights in different locations all over the world. Return values of significant wave height in the North-East Atlantic, the Norwegian Sea and the North Sea were computed by Aarnes et al. (2012). They have estimated the 100 year return level of sea wave height using three approaches: Block Maxima, r-Largest Order Statistics and Peak-Over-Threshold. For each method the generalized extreme value (GEV) distribution, the joint GEV distribution and the Generalized Pareto (GP) distribution were fitted to the data respectively. A comparative study on significant wave height in northern North Sea carried out by Soars and Scotto (2004) in Portugal. To estimate return values of significant wave height, they have considered GEV and limiting joint GEV distributions respectively to the BM approach and r-LOS where "annual wave heights" were taken as blocks. From their study results, they have examined that r-LOS method gives more reliable return levels than the BM method.

Two convenient analytical methods: Block Maxima (BM) approach and Peak over Threshold (POT) method are used in this study to estimate the 50 and 100 year return values of sea wave height. Seven years of 3-hourly Sea wave heights for the years 2002 to 2007 in Colombo district were used for the study. To estimate the 50 and 100 year return values of sea wave heights, Generalized Extreme Value (GEV) distribution and Generalized Pareto (GP) distribution were fitted to the data to each methods, BM and POT respectively. Maximum likelihood parameter estimation method was used to estimate the parameters of GEV and GP.

BM Approach

Let X_{1n}, X_{2n}……$X_{12,n}$ are series of blocked in to m sequences of length n and X_{in}'s are independent identically distributed random variables with common distribution function F(x).That is:$F(x)=Pr(X_{in} \leq x)$ .And let $M_n=\max(X_{1n},X_{2n}\ldots\ldots X_{12,n})$ denote the maximum of n^{th}sample. Then $Pr\{M_n \leq x\}=F(x)^n$. For non-trivial limit results, using normalization find $a_n>0$, b_n such that:

$$Pr\{(Mn- b_n)/ a_n \leq x\}= F(a_nx + b_n)^n \rightarrow G(x) \quad (1)$$

Choosing the block size is quiet important because, a too small value of n leads to a bias estimation on parameters and large value of n makes fewer number block maxima leads to large estimation on variance. Therefore, the block size has to be chosen between bias and the amount of variance, more commonly the block size is chosen as a time period of one year, where n is the

number of observations occurred in a year resulting in a series of annual maxima. Generally this Block Maxima approach is closely related with the GEV distribution.
G(x) is the Generalized Extreme Value distribution (GEV), defined by:

$$G\left(x;\xi,\sigma,\mu\right)=\exp\left\{-\left(1+\xi\frac{(x-\mu)}{\sigma}\right)^{-1/\xi}\right\} \quad (2)$$

where $\sigma > 0$ –scale, ξ- shape and μ - location parameter.According to the value of ξ , H(x) can be divided into three standard types of distributions: When $\xi = 0$, (Type 1) It is the Gumbel distribution, When $\xi > 0$ (Type II) It is the Frechet Distribution and when $\xi < 0$ (Type III) It is the Weibull Distribution.

POT Approach
The classical extreme value approach, called "Block Maxima", was strongly criticized because the estimation of the distribution based on extracted maximum event of each blocks, which ignores the other extreme events in the blocks.An alternative to the BM method is the "Peaks-Over-Threshold (POT)" method. In this approach, all the events exceeding a high threshold value are considered. Threshold selection is a bias variance trade-off. A threshold which is too small, leads to bias because of model asymptotic being invalid. A threshold which is too large leads to large variance because of few data points. Therefore one method is to construct the Mean Residual Life plot (MRL) for the mean excesses for some ranges of threshold and to fit GPD (Generalized Pareto Distribution) fits for some ranges of thresholds.

With an assumption of the daily data to be independent with common distribution function F, given a high threshold value u and looking at all exceedance of u, the distribution of excess value is given by:

$$F_u(y)=\Pr\{X\le u+y|X>u\}=\frac{F(u+y)-F(u)}{1-F(u)}, y>0 \quad (3)$$

For the sufficiently large threshold the distribution of exceedances may be approximated by GPD, Balkema and de Hann (1974), Pickands (1975), implying that, letting $u\to\infty$ leads to an approximate family of distributions given by,$G(y)=1-(1+\frac{\xi(y-u)}{\sigma})^{-1/\xi}$is the Generalized Pareto family. Also when $\xi\to 0$ which is converges to an Exponential family.

Return Level

Considering extreme values of a random variable, the return level of an extreme event, defined as the value, z_p, such that there is a probability of p that z_p is exceeded in any given year, or alternatively, the level that is expected to be exceeded on average once every 1/p years (1/p is often referred to as the return period). For example, if the 100-year return level for rainfall at a given location is found to be 280 mm, then the probability of rainfall exceeding 280 mm in any given year is 1/100 = 0.01.The return level is derived from the distribution GEV or GPD by setting the cumulative distribution function equal to the desired probability/quantile, 1−p; and then solving for the return level. For example, for the GEV distribution, the return level, z_p, is given by the following equation.

$$Z_p=\begin{cases}\mu-\frac{\sigma}{\xi}\left[1-\{-\log(1-p)\}^{-\xi}\right], & for\ \xi\neq 0\\ \mu-\sigma log\{-\log(1-p)\}, & for\ \xi=0\end{cases} \quad (4)$$

Data and Methodology

Three hourly sea wave heights (in meter) data for the years2002 to 2007 for the Colombo District were obtained from the Department of Meteorology, Colombo. For the entire analysis the statistical software R, especially "in2extRemes"packages in R was used. To fit the distribution and to estimate the 50 and 100 year return levels the Extreme Value theory was applied for the data set. Initially maximum wave height of each months were extracted from the data set and GEV distribution was fitted to the monthly maximum data. By using the Confidence Interval (CI) for the shape parameter, the best fitting distribution was identified. Secondly a suitable threshold value was identified in all 3-hourly data by using MRL plot and GPD fits for some ranges of thresholds. Then GP distribution was fitted to the data exceeding the specified threshold value. With the identified distributions the return levels for the return periods and their 95% confidence band were estimated.

Results and Discussion

Parameter Estimates of GEV Distribution Fitted to the Monthly Maximum Wave Height

Parameter estimates and their standard error within the parenthesis of the GEV distribution using maximum likelihood estimation method fitted to the monthly maxima of 3-hourly data are: μ: 1.3 (0.06), σ: 0.37 (0.04) ξ: 0.01 (0.1).Also, it can be concluded that the data fits well with the Gumbel distribution, (95 % CI for ξ is:(-0.1402, 0.2408)).

Figure 1 shows the two goodness-of fit plots, quantile plot for empirical data and quantile plot for model simulated data along with a density estimate plotand return-level plot for the monthly maximum sea wave height. Since much of the data line up on the diagonal of the empirical quantile and model simulated quantile plots, it can be concluded that the Gumbel distribution fitted well with the data. The return level plot plots the return values against the return period and includes 95% confidence interval and in which all the data points fall within the confidence band. The 50 and 100 year return values obtained from BM approach is 2.747 (2.3966, 3.0977) m and 3.01 (2.6351, 3.4291) m respectively.

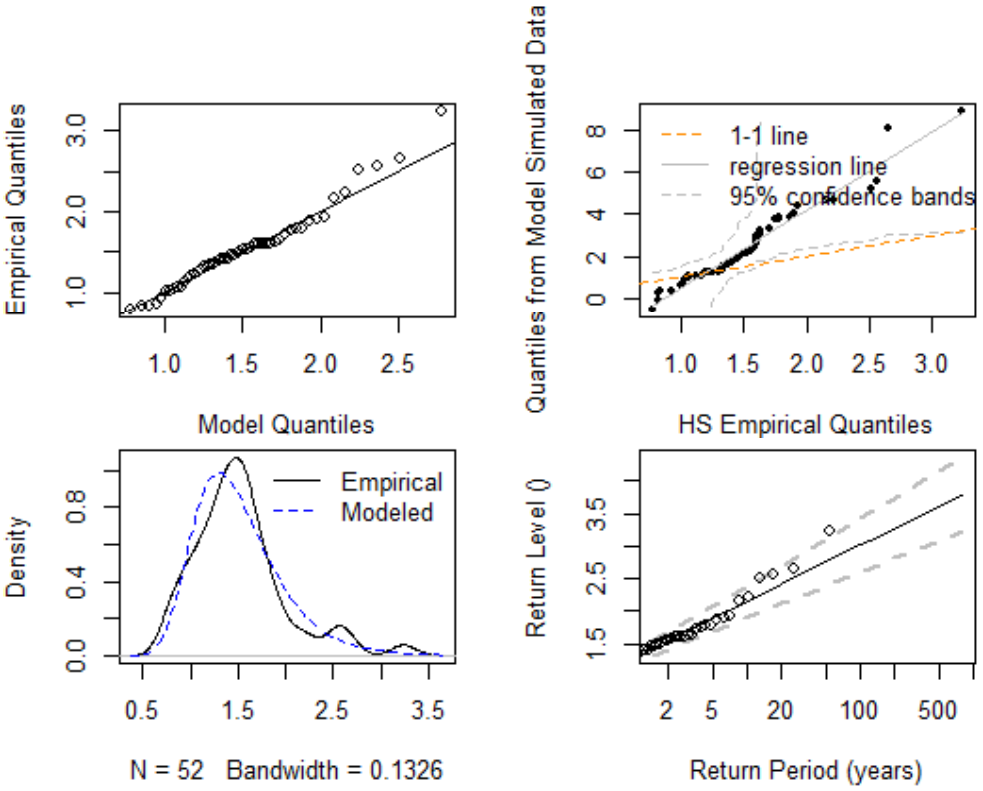

Figure 3: Gumbel diagnostic plot of the monthly maximum wave heights

Parameter Estimates of GP Distribution Fitted to the 3-hourly Wave Heights

A suitable threshold value has to be chosen in this approach. One method for choosing threshold value is with mean residual Life plot. The mean exceedances above u (threshold value) should be linear, so the idea is to looking for linearity in a plot of the empirical mean residual life plot (Stuart Coles, 2001, Matthew J. Pocernich, 2002).But it is difficult to identify the threshold value only with the MRL plot. Therefore an alternative approach is to plot the GPD fits for some ranges of thresholds and look for parameter stability. Fig 2, shows the MRL plot for the 3-hourly wave heights, According to that, the threshold greater than 1.1 m looks reasonable. However According to the Fig 3, the threshold 1.33 m looks high.

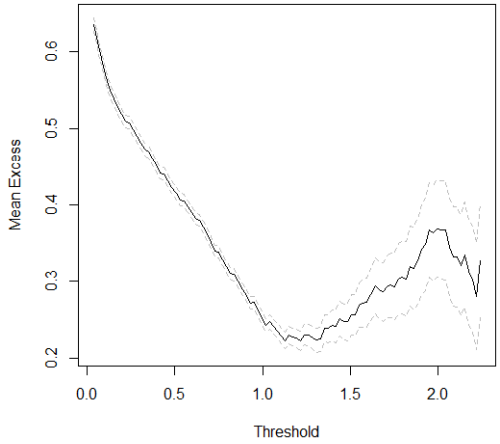

Figure 4: MRL plot for 3-hourly wave height

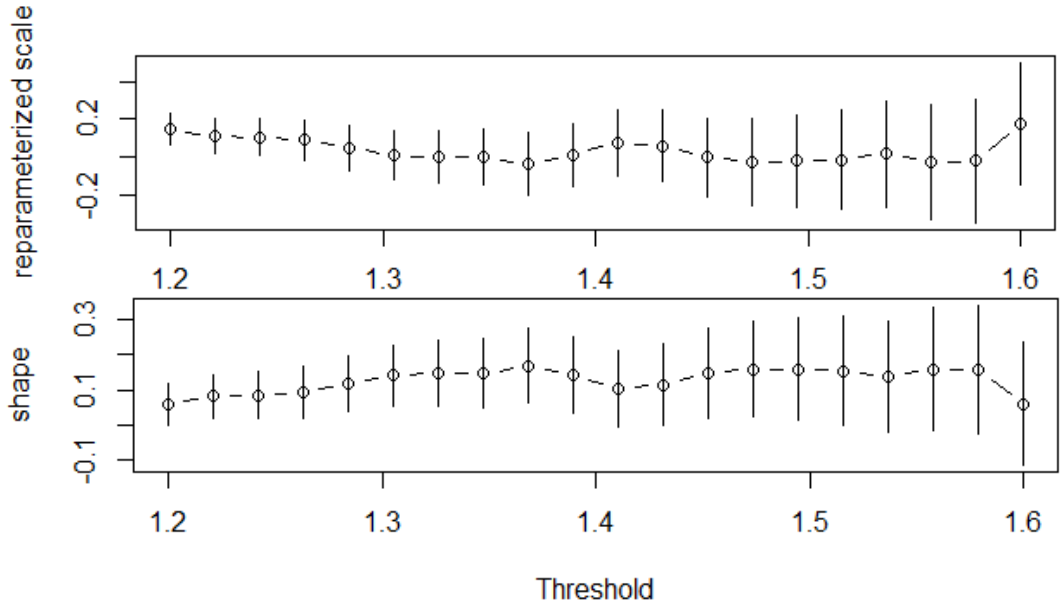

Figure 5: GPD fits of 20 thresholds from 1.2 to 1.6 m for 3-hourly wave heights

Maximum Likelihood Parameter estimates of GPD distribution fitted to the 3-hourly sea wave height above 1.33 m with the corresponding standard errors are σ=0.21 (0.095) and ξ=-0.12 (0.01). Also, it can be concluded that the data fits well with the GP distribution, (95 % CI for ξ is:(0.16, 0.17))Figure 4 shows the two goodness-of fit plots, quantile plot for empirical data and quantile plot for model simulated data along with a density estimate plot and return-level plot for the peak over threshold data. Since much of the data line up on the diagonal of the goodness of fit plots we can say that which fits the data perfectly. The return level plot plots the return period against the return level and also includes 95% confidence interval and in which all the data points fall within the confidence band. The 50 and 100 year return values obtained from POT approach is 3.541 (2.991, 4.358) m and 3.895 (3.2502, 4.7662)m respectively.

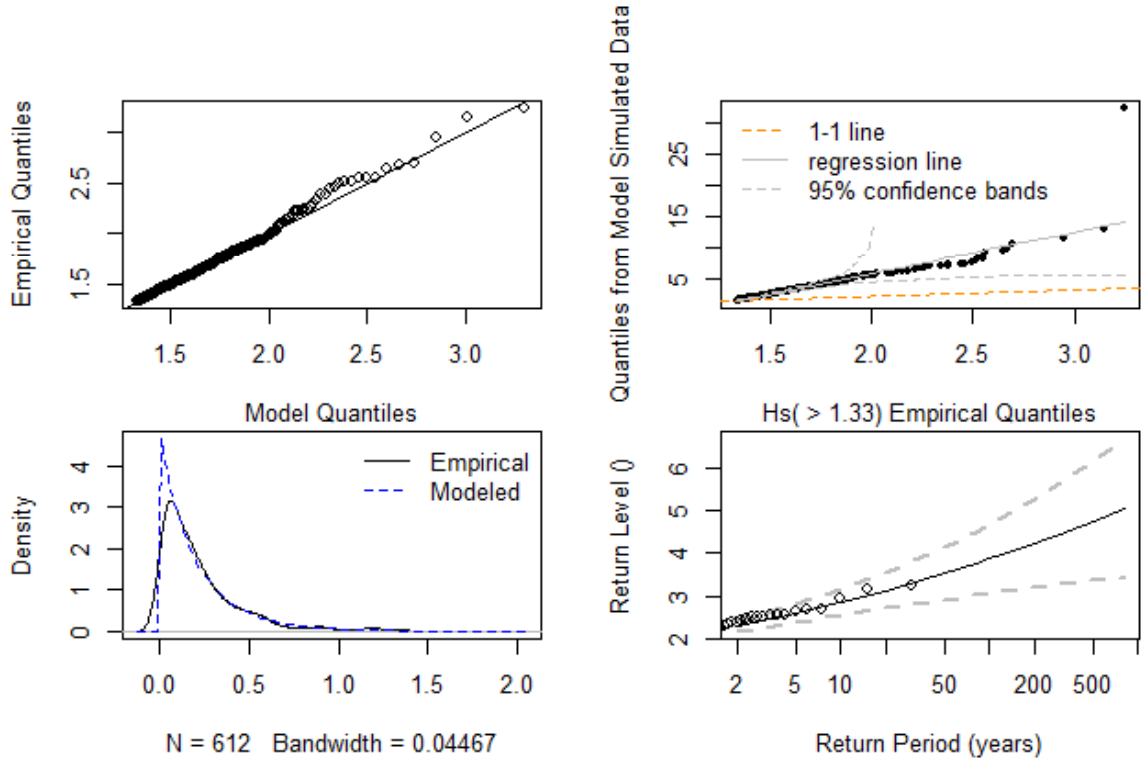

Figure 6: Diagnostic plot for the POT Wave Height

Conclusions

Three hourly sea wave height data in Colombo, Sri Lanka for the years from 2002-2007 were used in this study .To estimate the 50 and 100 year return levels of sea wave height, two techniques: BM approach and POT approach were applied. In the BM approach, block size was considered as month and which fits well with the Gumbel distribution .In POT approach specified threshold value was chosen as 1.33 m using MRL plot and GPD fits for some ranges of thresholds. Wave heights exceeding 1.33 m fit well with the GPD. Estimated 50 and 100 year return levels obtained from BM approach is 2.747 (2.3966, 3.0977) m respectively and 3.01 (2.6351, 3.4291) m and from the POT method is 3.541 (2.991, 4.358) m and 3.895 (3.2502, 4.7662)m respectively.

References

Aarnes, OJ, Breivik, Q & Reistad, M 2012, 'Wave Extremes in the Northeast Atlantic', *Journal of Climate*, vol. *25*(5).

Balkema, A & de Haan, L 1974, 'Residual Life Time at Great Age', *Annals of Probability*, vol. 2, pp. 792–804.

Matthew, J & Pocernich 2002, 'Application of Extreme Value Theory and Threshold Models to Hydrological Events', A Master thesis submitted to the University of Colorado at Denver.

Pickands, J 1975, 'Statistical Inference Using Extreme Order Statistics', *Annals of Statistics*, vol. 3, pp. 119–131.

Stuart Coles 2001, 'An Introduction to Statistical Modeling of Extreme Values', Springer, New York, NY,.

Soares, CG & Scotto, MG 2004, 'Application of the r Largest Order Statistics for Long-Term Predictions of Significant Wave height', *Coastal Eng*, vol. 51, pp. 287–294.

Davison, AC & Smith, RL 1990, 'Models For Exceedances over High Thresholds', *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 393-442.

Ledermann, E, Lloyd, E, Vajda, S & Alexander,C. (Eds.), 'Handbook of Applicable Mathematics', Wiley, pp. 437–471.