Comparison Between Maximum Annual and Peak Over Threshold Methods for the Determination of Extreme Waves in Moroccan Atlantic Coast

H Bakali*, N Serhir and I Aouiche
Er-SHEMSS, Laboratory of Civil Engineering, Hydraulics, Environment and Climate, Hassania School of Public Works, Morocco
E-mail: albakalihosny@gmail.com

Abstract. In this article, we present a study that focuses on forecasting the one hundred-years return period waves height using two methods; i) Peak Over Threshold (POT) and ii) Annual Maxima (MA). The analysis is carried out at three points at the Atlantic coast of northern Morocco. The results obtained by these two methods indicate a difference of less than 10% and have shown that to ensure safe dimensioning of maritime structures, the application of the two statistical methods is necessary. This is explained by the drawbacks encountered in both methods. In the MA method, the inconvenience is linked to the choice of non-extreme events when no significant storm is recorded during one year, furthermore, the non-consideration of important events recorded in other year. In the second method (POT), the difficulty consists in the uncertainty of the calculations by the POT method due to the mistrust associated with the choice of the censoring threshold.

1. Introduction
The design of maritime and coast structures is generally carried out according to 100-years wave storm. The determination of the extreme heights is based on statistical forecasting by models fitted to the data of the maximum waves measured on site or generated by numerical models from wind fields [1]. Generally, wave data are processed by one of two methods:

Annual Maxima method [2]: is derived from the general BLOCK MAXIMA method. It consists in decomposing the data period into regular periods and considering the maximum value for each period. In the particular case when this period is equal to one year, the method is designated by the method of Annual Maxima.

Peak Over Threshold (POT) method [3]: is based on selecting extreme values of a data above a threshold value.

The POT method has the advantage of studying all waves exceeding a censorship threshold [2]. However, this method has the disadvantage of results dependency on the chosen threshold. An additional verification by the annual maxima method is sometimes a way for the verification of the POT method results [4].

In this study, we will compare the results of extreme waves forecasted with the MA and POT methods at three points of the Moroccan Atlantic coast. In the POT method, many possible thresholds will be examined. Our purpose is to determine the safest method for coastal structures design.
The wave data is the sample of significant wave heights (Hs), for the period from 1958 to 2019, they are procured from the SIMAR-44 (Puertos del estados) database in front of three Moroccan coastal cities: KENITRA, MOHAMMEDIA, and SAFI.

2. Methods

2.1. Study area
The study area is located between the cities of Kenitra and Safi in the Atlantic coast of the northern morocco as shown in Figure 1. Major cities and ports of morocco are located in this littoral band: Mohammedia, Casablanca, Eljadida, Safi… Most of infrastructures in this zone incurred minor to considerable damages due to the storm of the 6th and 7th January 2014 [5]. In order to avoid possible underestimation of extreme waves, the determination of a correct value of wave height design is crucial to obtain safe structures.

\[ \hat{e}_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)_{|X_i > u|}}{\sum_{i=1}^{n}1_{|X_i > u|}} \]  

(1)

Xi: the events studied
u : Threshold value
The linearity of the ME function indicate a possible threshold. The threshold used should not be very high in order to obtain an average number of waves events per year between 2 and 5 [7].
2.3. Method for selecting maximum wave heights for the POT method
In order to verify the independence of the maximum values of the storms studied, the minimum interval retained between two maximum values of two successive storms is 2 days [8].

2.4. Method for determining the empirical non-exceedance probability
The compromise formula recommended by Cunnane [9] for the determination of the empirical non-exceedance probabilities is (2):

$$P_k = \frac{k - 0.4}{n + 0.2}$$

$P_k$: Empirical probabilities of non-exceedance of the k rank event in the increasing order of the dataset
$n$: total number of events of the dataset

2.5. Software used
We performed the statistical adjustments using the HYFRAN-PLUS software. This software allows the graphical comparison between different statistical models as well as the comparison between statistical distributions by goodness-of-fit tests [10]

2.6. Method for determining the most suitable statistical models
The two methods of a standard approach of an extreme value analysis (EVA) are:
- **GEV**: The Generalized Extreme Value is a theoretical model for the distribution of BLOCK MAXIMA sample [5]. GEV is a three parameters model: $\mu$, $\sigma$, and $\xi$. Variation of the shape parameter $\xi$ gives three types of distributions:
  - $\xi = 0$ : Gumbel distribution
  - $\xi < 0$ : Weibull distribution
  - $\xi > 0$ : Fréchet distribution
- **GPD**: In the particular case when the number per year of exceedances is a random Poisson process, the GPD (General Pareto Distribution) is a model of exceedances over threshold [5]. In the particular case of the shape parameter $\xi = 0$, the model obtained is the exponential distribution. General models (GEV and GPD) shouldn’t be used automatically for natural phenomena such as ocean waves [11]. Indeed, it is not recommended to carry out the standard EVA for events that are the results of non-linear processes [12].

Authors of [10] recommend a multi-distribution approach in the particular case when the adjustment on the distribution tail is crucial for determining the forecasted extreme values (such the 100 years return period wave). In this study, the probability distribution models tested are: General Extreme Value (GEV), General Pareto Distribution (GPD), Gumbel (EV1), Weibull (W), Gamma (G2), Inverse Gamma (IG), Log Norma (LN), Pearson type III (P3), Exponential (EXP).

In order to choose the most adequate statistical models, we used the graphical method which is efficient for a statistical analysis [13], while the goodness-of-fit tests $\chi^2$ [14], AIC [15], and BIC [16] tests are used for a refining purpose and confirming the graphic choice.

2.7. Method of determining model parameters
The estimation of the theoretical model parameters is made by one of the three methods:
- Maximum likelihood method (MV)
- Method of moments (MM)
• Weighted moments method (MMP)

3. Results

3.1. Determination of Thresholds

The Figure 2 presents the graphical determination of thresholds for the three studied points.

The curves of the ME function offer several choices of thresholds, we retain all possible thresholds to have an average number per year of events as indicated by [6].

The ME function curve of the data in Mohammed’s coast offer a unique possible choice of threshold, we add another value \( u = 4.10 \) for a comparative purpose.

The thresholds used and the average annual number of events studied for the three coast points are summarized in Table 1:

![Figure 2](image-url)
Table 1. Values of thresholds and annual average per year extreme waves above thresholds.

| Censorship threshold value | Annual average per year number of events considered (Na) |
|---------------------------|-------------------------------------------------------|
| Coast of KENITRA          | 4.00 m                                                | 5.06 |
|                           | 4.60 m                                                | 2.78 |
| Coast of Mohammedia       | 4.10 m                                                | 7.21 |
|                           | 4.60 m                                                | 3.64 |
| Coast of Safi             | 4.40 m                                                | 4.00 |
|                           | 4.80 m                                                | 2.64 |

The analysis of thresholds obtained shows homogeneity of results. Indeed, there are two possible thresholds choices for the three samples:

- Low threshold between 4 m and 4.4 m, tied to \( N_a \) from 4 to 5 events per year,
- High threshold between 4.6 m and 4.8 m, tied to \( N_a \) from 2 to 3 events per year,
- Values of low and high thresholds in the points of Kenitra and Mohammedia are practically equal; this observation means that thresholds may be a regional character.

Selecting a unique choice of threshold is an unresolved issue [17]. Hence, to have a global comparison of POT and MA methods, we will check all possible thresholds indicated in Table 1.

We present in Figures 3, 4, and 5 the results of the graphical fittings of the models examined for the trends of extreme waves in the three studied points at coasts of Kenitra, Mohammedia, and Safi.

Selected models are indicated with the addition of a square to the identifier of the curve.

The results of the goodness-of-fit tests for data analysis in the coast of KENITRA are presented in the Table 2.
Table 2. Results of the fitting tests for wave data analysis in the coast of KENITRA.

| POT Method | Model | EV1 | GEV | W | G2 | IG | LN | P3 | EXP | GPD |
|------------|-------|-----|-----|---|----|----|----|----|-----|-----|
|            | Method | MV  | MV  | MM | MMP| MV | MV | MV | MM  | MV  |
| KH2 test   | u=4.0 | 82.3| 30.1| 64.3|57.7| 255.6| 137.2| 104.4| 125.1| 24.0|
| AIC test   |       | 696 | 654 | 687 |667 | 905 | 773 | 735 | 752  | 605 |
| BIC test   |       | 704 | 665 | 698 |678 | 912 | 781 | 742 | 760  | 612 |
|            | Method | MV  | MV  | MM | MMP| MV | MV | MV | MM  | MV  |
| KH2 test   | u=4.6 | 43.3| 10.9| 31.2|11.8| 119.5| 86.1 | 47.0| 56.9 | 8.7 |
| AIC test   |       | 370 | 345 | 364 |351 | 509 | 424 | 401 | 412  | N/D |
| BIC test   |       | 377 | 354 | 374 |361 | 515 | 430 | 408 | 418  | N/D |
| MA method  |       | 7.0 | 5.4 | 6.1 |5.4 | 9.3 | 6.7 | 3.5 | 6.7  | 50.6 |
|            | Method | MV  | MV  | MM | MMP| MV | MV | MV | MM  | MV  |
| KH2 test   |       | 192 | 192 | 193 |192 | 206 | 192 | 190 | 190  | N/D |
| AIC test   |       | 192 | 192 | 193 |192 | 206 | 192 | 190 | 190  | N/D |
| BIC test   |       | 195 | 198 | 199 |198 | 210 | 196 | 194 | 194  | N/D |

Analysis of goodness-of-fit tests and results of the 100 years return period forecasted wave (or 100y-Hs) highlights the following points:

- **Sample over threshold u=4.0 m**: we retain the Exponential model for predicting extreme quantiles, this choice is confirmed by the (AIC, BIC) tests. Other models shouldn’t be chosen because they lead to an underestimation of the design wave height, GPD model gives a low estimation 100y-Hs = 8.03 m.

- **Sample over threshold u=4.6m**: GPD model gives the best results of goodness-of-fit tests values, the value obtained for 100y-Hs is 8.48 m this value shouldn’t be retained because it have been exceeded by a total of three waves in the last 60 years in the studied point. Hence we will retain the Exponential model.

- **Sample of maximum annual waves**: Gamma and Log normal models underestimate the extreme values. Hence, we retain the Gumbel model, this choice is confirmed by the (AIC, BIC) tests.
The results of the fitting tests for data analysis in the coast of MOHAMMEDIA are presented in the Table 3

Table 3. Results of the fitting tests for wave data analysis in the coast of MOHAMMEDIA.

| Model       | EV1 | GEV | G2  | IG   | LN  | P3  | EXP | GPD |
|-------------|-----|-----|-----|------|-----|-----|-----|-----|
| Method      | MV  | MV  | MM  | MM   | MV  | MV  | MM  | MM  |
| KHI2 test   | 91,6| 113,5| 32,3| 96,0 | 57,8| 233,8| 156,0| 153,1| 15,0| 15,0| 20,3|
| AIC test    | 852 | 866 | 776 | 832  | 795 | 986 | 929 | 955 |
| POT Method  | POT Method u=4.1 |
| BIC test    | 861 | 875 | 788 | 844  | 808 | 994 | 937 | 963 |
| KHI2 test   | 54,4| 53,7| 25,8| 43,1 | 45,9| 99,0| 102,9| 102,7| 12,9| 9,1 | 10,66|
| AIC test    | 441 | 448 | 401 | 431  | 413 | 511 | 483 | 496 |
| POT Method  | POT Method u=4.6 |
| BIC test    | 447 | 455 | 412 | 442  | 423 | 517 | 490 | 503 |
| KHI2 test   | 9,9 | 9,6 | 7,7 | 8,7  | 7,7 | 5,4 | 6,4 | 7,0 |
| MA Method   | AIC test | 184 | 184 | 186 | 186  | 185 | 185 | 183 |
| BIC test    | 188 | 188 | 191 | 192  | 191 | 189 | 187 | 188 |

Analysis of goodness-of-fit tests and results of the 100 years return period forecasted wave (or 100y-Hs) highlights the following points:

- **Sample over threshold u=4.1m and u=4.6m**: the values of the AIC and BIC tests are practically the same for the two best models (Exponential and GPD), considering the KHI test we retain the Exponential model, this model gives a better approximation of the extreme quantiles than the GPD model.

- **Sample of maximum annual waves**: Gumbel model gives the best fitting for extreme values for MA method, other models shouldn’t be selected because of their underestimation of the wave design height.
Figure 5. Results of graphic adjustments by POT and MA methods in the point in Safi’s coast for POT method with threshold u=4.4 m (a), for POT method with threshold u=4.8 m (b), and MA method (c).

The results of the fitting tests for data analysis in the coast of SAFI are presented in the Table 4.

### Table 4. Results of the fitting tests for wave data analysis in the coast of SAFI.

| Model Method | EV1 KHI2 test | GEV AIC test | W BIC test | G2 | LN | P3 | EXP | GPD |
|--------------|---------------|--------------|------------|----|----|----|-----|-----|
| POT Method u=4.4 MMP | 48.7 19.2 33.2 | 429 436 514 | 29.3 172.8 70.1 | 172.8 62.2 13.9 | MM MV MV MM MV MM MM |
| POT Method u=4.8 MMP | 46.7 29.4 33.2 | 425 514 614 | 37.4 117.7 70.1 | 94.7 67.0 12.1 | MM MV MV MM MV |
| POT Method u=4.8 MMP | 285 256 275 | 263 398 | 263 263 263 | 398 398 398 | 263 263 263 |
| POT Method u=4.8 MMP | 231 231 231 | 231 231 231 | 231 231 231 | 231 231 231 | 231 231 231 |
| POT Method u=4.8 MMP | 162 162 162 | 162 162 162 | 162 162 162 | 162 162 162 | 162 162 162 |
| POT Method u=4.8 MMP | 166 166 166 | 166 166 166 | 166 166 166 | 166 166 166 | 166 166 166 |
| POT Method u=4.8 MMP | 395 395 395 | 395 395 395 | 395 395 395 | 395 395 395 | 395 395 395 |
| POT Method u=4.8 MMP | 43,5 - | 43,5 - | 43,5 - | 43,5 - | 43,5 - |
| POT Method u=4.8 MMP | 224 224 224 | 224 224 224 | 224 224 224 | 224 224 224 | 224 224 224 |
| POT Method u=4.8 MMP | 231 231 231 | 231 231 231 | 231 231 231 | 231 231 231 | 231 231 231 |
| POT Method u=4.8 MMP | 202 202 202 | 202 202 202 | 202 202 202 | 202 202 202 | 202 202 202 |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |
| POT Method u=4.8 MMP | - - - | - - - | - - - | - - - | - - - |

Analysis of fitting tests and results of the 100 years return period forecasted wave (or 100y-Hs) highlights the following points:

- **Sample over threshold u=4.4m**: GEV-MMP is the best fitted model to the sample selected with threshold u=4.4m, other models lead to over-estimations of 100y-Hs (GEV-MV, GEV-MM, EXP) or underestimation of 100y-Hs (EXP, GPD).

- **Sample over threshold u=4.8m**: the values of the AIC and BIC test are practically the same for the two best models (Exponential and GPD), based on the KHI test we retain the Exponential model, this model gives a better approximation of the extreme quantiles than the GPD model.

- **Sample of maximum annual waves**: Based on the fitting tests we retain the Gamma model.

Results of the 95% confidence interval (C.I.) for the values of the 100y-Hs are presented in Figure 6.
Figure 6. Graphical synthesis of the mean values of the 100y-Hs C.I. determined by annual maxima method (MA) and POT method for high and low threshold.

The analysis of the results highlights the following points:

- Width of 95% confidence intervals vary from 1 to 2 meters, this uncertainty can lead to significant increase in coastal structures costs if choice is made for the highest value of the C.Is.
- For all studied points, the three C.Is for 100y-Hs determined by MA and POT methods overlaps. Hence, the difference between the three methods may be not statistically different [18]
- The graphical adjustments indicate that the Exponential distribution is the most fitted the POT method. Most suitable statistical models for MA method are Gumbel and Gamma.
- Variation of threshold does not imply a large variation in the results of the POT method, this observation is valid for the condition of the average number of events per year $N_a$ between 2 and 5
- The gap between the results obtained by the two methods POT and MA is less than 10%.
- The analysis of the wave data indicates that the point in the coast of Mohammedia is characterized by the most important waves. However, the predicted 100-years return period wave height is not maximal in this point. This disparity is due to the difference in the distribution tail of data in each sample, the trends of theoretical models fitted to the data are consequently different.
- In the point of Mohammedia’s coast, the safest value to retain for structures design is the value obtained by the maxima annual method. In this particular point, the statistical determination of the wave high design by the two methods (MA and POT) is the best method for a safe value of wave design.

4. Conclusions

Following the use of the two methods POT and MA for the determination of long return period waves, we conclude that the use of the two methods is required for the verification of the results obtained to ensure safe dimensioning of maritime structures. The variation of thresholds in agreement with the conditions proposed by Mazas and Hamm [7] does not imply large variations of forecasting 100-years return period wave height. This result can be considered as a reference to set an adequate threshold at the level of the Mohammedia, Kenitra and Safi zones. On the scale of Morocco, a general study of regional thresholds is relevant to define the most appropriate threshold for each coastal zone.
The theoretical probability distribution for the data vary according to the method of data selection (MA or POT). In addition, it seems that the most suitable theoretical distribution is a regional character. Definition and methods for a regional frequency analysis are introduced by Hosking and Wallis [19]. The local analysis for the determination of the censorship threshold allowed the definition of the interval of the most suitable values.

The results of the two methods concur for a relatively long period of 62 years wave’s data. Examination of the forecasted extreme quantiles stability as a function of the data period indicated that the POT method presents non-homogeneity of the forecasting’s results for hydraulic phenomena over periods less than 50 years [20]. Hence, other studies are necessary to examine the validity of the conclusions of this paper for data series with shorter periods.

References

[1] The Waveswatch III Development Group. User manual and system documentation of WAVESWATCH III version 6.07. NOAA/NWS/NCEP/MMAB. College Park. MD. 2019.
[2] Ferreira A et De Haan L 2015 On the block maxima method in extreme value theory Annals of Statistics 43 276-98.
[3] Pickands J 1975 Statistical inference using extreme order statistics the annals of statistics 3 119-31
[4] Mathiesen M, Goda Y, Hawkes P J, Mansard E, Martín M J, Peltier E, Thompson E F and Van Vledder G 1995 Recommended practice for extreme wave analysis J. Hydraul. Res 32 803-14.
[5] Ministère de l’Equipment de transport et de la Logistique 2014 Impact de la houle exceptionnelle du 06-7 janvier 2014 sur les infrastructures portuaires
[6] Coles S 2001 An Introduction to Statistical Modeling of Extreme Values Springer.
[7] Mazas F and Hamm L 2011 A multi-distribution approach to POT methods for determining extreme wave heights Coastal Engineering 58 385-94
[8] Caires S and STERL A 2005 100-Year Return Value Estimates for Ocean Wind Speed and Significant Wave Height from the ERA-40 Data Journal of Climate 1032-48
[9] Cunnane C 1978 Unbiased plotting positions - A review Journal of hydrology 37 205-22
[10] El Adlouni S and Bobée B 2014 Analyse Fréquentielle avec le logiciel HYFRAN-PLUS
[11] Mazas F and Hamm L 2010 Théorie statistique du renouvellement pour la détermination des houles extrêmes. Partie 1 : le point sur les méthodes disponibles LA HOUILLE BLANCHE 4 92-102
[12] Bouleau N. 1991 Splendeurs et misères des lois des valeurs extrêmes Revue Risques 4 85-92
[13] Fienberg S E 1979 Graphical Methods in Statistics The American Statistician 33 165-78
[14] Pearson K 1900 On the criterion that a given system of deviations given system of deviations of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling Philosophical Magazine Series 5 157-75
[15] Hirotugu 1974 A New Look at the Statistical Model Identification IEEE Transactions on automatic control 19 716-23
[16] Schwarz G 1978 Estimating the dimension of a model The Annals of Statistics 6 461-64
[17] Solari S, Eguen M, Jose Polo M and Losada M A 2017 Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value Water Resources Research 53 2833-49
[18] Sze Huey Tan and Say Beng Tan 2010 The Correct Interpretation of Confidence Intervals Statistics 19 276-78
[19] Hosking J and Wallis J 1997 Regional Frequency Analysis: An Approach Based on L-Moments Cambridge: Cambridge University
[20] Fischer S and Schumann A H 2014 Comparison between classical annual maxima and peak over threshold approach concerning robustness Sonderforschungsbereich (SFB) 823.