

Single Site Extreme Wave Analysis in the Pacific Ocean Comparing Stationary and Non-Stationary GEV Models

ABSTRACT

The adequate knowledge of the weather behavior is very important for the design and management of socioeconomical, environmental and sustainability human interests in the coasts and oceans. In the present study an extreme value analysis of maximum significant waves recorded at a buoy located in the Pacific Ocean is carried out. The analysis is carried out from two perspectives, by considering a Generalized Extreme Value (GEV) model with stationary distribution (i.e., the time variations are not accounted for), and by considering a non-stationary GEV model, which incorporates the monthly seasonality of maximum observed values in time increments; the maximum significant wave behavior is parameterized using harmonic functions for the distribution measures. Both approaches are compared for a single buoy. In the study a seasonality effect is found, which was also present at the Gulf of Mexico in previous studies, and which cannot be captured by a stationary model.

Keywords: [GEV model, stationary, non-stationary, seasonality, projections]

1. INTRODUCTION

The use of the extreme value theory has been widely extended in recent years to meteorological and oceanographic variables. In a recent study, [1] use the Gumbel probability density function (PDF) together with maximum annual values to assess the extreme maximum annual significant wave height for projection purposes. [2] used non-stationary GEV models for investigating period trends in extreme waves. [3] developed design values for significant wave height accounting for direction and seasonality by means of a non-stationary model. [4] compared the behavior of tropical cyclone data (extreme values), obtained through simulations, versus those recorded at buoys located in the Meridional Chinese Sea and the Oriental Chinese Sea, and proposed an approach combining the set of data.

Philip and Ewans [5] refer to the Bayesian inference for modelling extreme values of characteristic marine environments and for designing maritime structures. Calderón-Vega et al. [6] investigated the seasonality of extreme waves in the Gulf of Mexico; it was found that, for the Gulf of Mexico, there is indeed seasonality effects on the extreme values; moreover, two clear peaks were found which are associated to hurricanes and cold fronts in distinctive seasons of the year. The study was carried out for several buoy sites in the Gulf of Mexico, but it was not performed for buoys in the Pacific. The use of stationary models is very common when carrying out analyses of extremes, however, as already mentioned and found in a previous study [6], seasonality may play a role in the probabilistic characterization of

meteo-oceanographic variables, and it may also have an impact on design, management and reliability-oriented tasks; furthermore, since meteo-oceanographic variables are directly linked to climate change detection, vulnerability, future projection and sustainability of coasts and coastal infrastructure [7, 8], an impact due to seasonality effects would be expected on all these issues for extreme waves. Although not pursued in the present study, the relation between extreme wave heights and sustainability (using stationary versus non-stationary models) could be explored in future projects; the present study can be considered as a first step in that direction.

The main objective of this study is to select a single buoy site to compare stationary versus non-stationary GEV models for significant wave height projections and to inspect possible seasonality effects in the Pacific Coast for this meteo-oceanographic variable.

2. DATA USED

Significant wave data recorded at 51004 buoy located at the Pacific Ocean are analyzed; the buoy is operated by the National Data Buoy Center (NDBC, www.ndbc.noaa.gov).

Figure 1 shows the buoy location and general information. The 51004 buoy geographical coordinate are 17.602N 152.395W, and is located to the south of the Hawaii islands. It comprises a time window from 1984 to 2017 including all years, except 2010. The significant wave height (H_s) is obtained as the mean of the highest third for wave heights during a given sample time period, which is 20 minutes for the used NDBC data. The NDBC H_s have been preliminarily processed for uniformizing the data series, so that they can be employed in the probabilistic model.



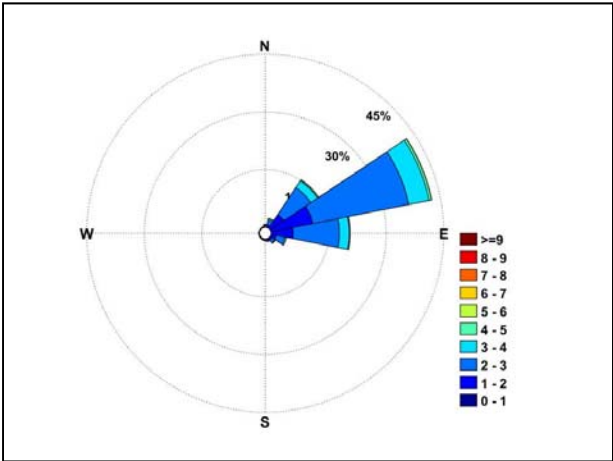
Fig. 1. NOAA 51004 Buoy

3. RESULTS AND DISCUSSION

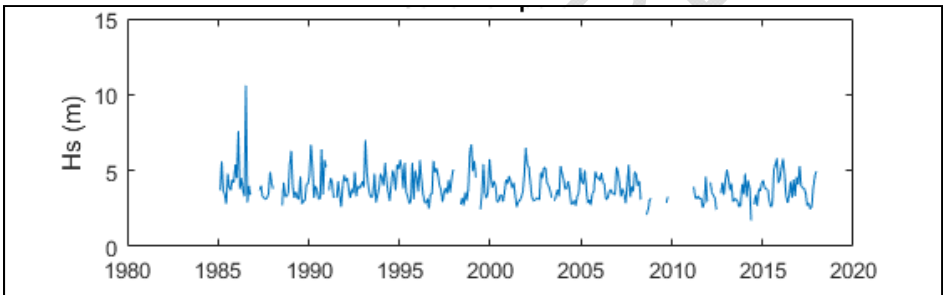
3.1. Preliminary Analysis

A preliminary analysis of the data is carried out to broadly inspect the behavior of the variable under study at the selected site. In Figure 2a (upper part) a main wave incidence to

the NEE is observed. The maximum recorded H_s is equal to 10.60 m; see Figure 2b (lower part). The most frequently observed significant wave height fluctuates around 2 m.



(a)



(b)

Fig. 2. Basic Statistics of the buoy under study. (a) Wave rose and (b) Time series data

3.2. Extreme Value Model

3.1.1. GEV Model

The model is based in the Generalized Extreme Value (GEV) distribution e.g., [9], which encompass, in a single mathematical expression, the three well-known families of extreme distributions, i.e., the Gumbel, Weibull and Frechet Distributions, and is given by the following equations:

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\psi}\right)\right]_+^{-1/\xi}\right\} \rightarrow \xi \neq 0$$
(1)

$$G(x) = \exp\left\{-\exp\left[-\left(\frac{x - \mu}{\psi}\right)\right]\right\} \rightarrow \xi = 0$$

The PDF is defined by means of:

$$g(x) = \frac{1}{\psi} \left[1 + \xi\left(\frac{x - \mu}{\psi}\right)\right]_+^{-(1+1/\xi)} \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\psi}\right)\right]_+^{-1/\xi}\right\} \text{ si } \xi \neq 0$$
(2)

$$g(x) = \frac{1}{\psi} \exp\left(-\frac{x - \mu}{\psi}\right) \exp\left[-\exp\left(-\frac{x - \mu}{\psi}\right)\right] \text{ si } \xi = 0$$

where $[a]_+$ implies the $\max(a, 0)$, $-\infty < \mu < \infty$ is the location parameter, $\psi > 0$ is the scale parameter and ξ the shape parameter.

The location parameter represents the mean values of the random variable, x , and defines the value with a non-exceedance probability $\exp(-1)$. The scale parameter represents the sample dispersion. Finally, the shape parameter defines the distribution family as a function of the tail behavior; when ξ exists, $G(x)$ includes two distribution families.

To develop the model, the lack of data should be accounted for, since this issue affects the parameters estimation stability for the extreme value distribution. Therefore, since it is known that data for some years may be absent (e.g., due to buoys maintenance), a minimum of data per time unit is considered to define the maximum values, by adopting the criterion of rejecting maximum monthly events with data blank spaces of up to 40% e.g., [6]

3.1.2. Non-stationary GEV Model

For modelling H_s the GEV is used for block maxima of a time interval. In this study, the monthly maximum values are used for the extreme value samples. Note that throughout this study, the methodology used and referenced by [6] is followed.

The seasonality is defined as the cyclic changes during the year. Usually, this cycle is linked to established climatic patrons; consequently, it repeats itself for many years and behaves in a sinusoidal way, as shown in Figure 3. In Figure 3 the marine weather variation is exhibited; in the winter the wave heights are larger than in the spring.

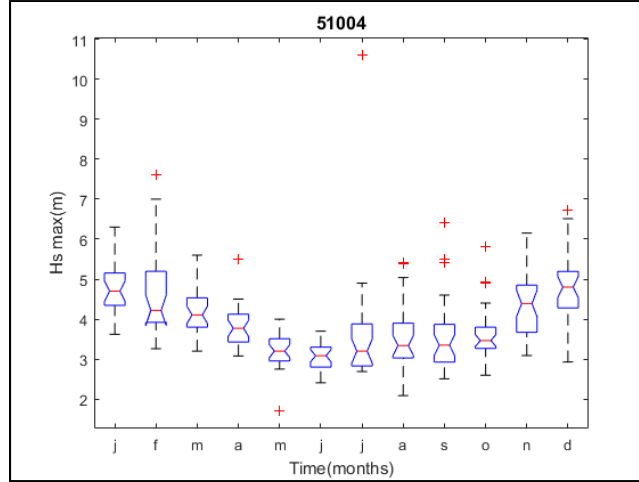


Fig. 3. Seasonality of monthly maxima

For the assessment of time-dependency an extension of traditional stationary models in extreme value theory is used. In this case the successive monthly maxima are considered as independent random variables, disregarding the need of the homogeneity hypothesis for the consecutive months, because they do not have identical distributions. It is assumed that the maximum monthly value Z_t of the observed Hs in month t follows a GEV distribution, where the location $\mu(t) > 0$, scale $\psi(t) > 0$ and shape $\xi(t)$ parameters are time-dependent.

The cumulative distribution function (CDF) of Z_t is given by:

$$F_t(z) = \begin{cases} \left\{ \exp \left\{ - \left[1 + \xi(t) \left(\frac{z - \mu(t)}{\psi(t)} \right) \right]_+^{-1/\xi(t)} \right\} \right. \\ \left. \exp \left\{ - \exp \left[- \left(\frac{z - \mu(t)}{\psi(t)} \right) \right] \right\} \right\} \quad (3) \end{cases}$$

where $[a]_+ = \max [a, 0]$.

The non-stationarity of the models is introduced in the GEV parameters as:

$$\theta = \begin{cases} \mu(t) = \beta_0 + \beta_1 \cos(2\pi t) + \beta_2 \sin(2\pi t) + \beta_3 \cos(4\pi t) + \beta_4 \sin(4\pi t) \\ \psi(t) = \alpha_0 + \alpha_1 \cos(2\pi t) + \alpha_2 \sin(2\pi t) + \alpha_3 \cos(4\pi t) + \alpha_4 \sin(4\pi t) \\ \xi(t) = \gamma_0 + \gamma_1 \cos(2\pi t) + \gamma_2 \sin(2\pi t) + \gamma_3 \cos(4\pi t) + \gamma_4 \sin(4\pi t) \end{cases} \quad (4)$$

where $\psi(t > 0)$

β_0, α_0 y γ_0 are mean values; β_i, α_i y γ_i (for $i=1,2$) are the harmonic amplitudes; β_i, α_i y γ_i (for $i=3,4$) are the subharmonic amplitudes and t is in years.

In Figure 4 the results from the variability analysis within a year of the maximum-likelihood estimators for the location, μ , scale, ψ , and shape, ξ , parameters along the year are depicted. The circles represent the values computed with the stationary GEV model, month to month, and the line is the resulting function from the regression fitting with two harmonics.

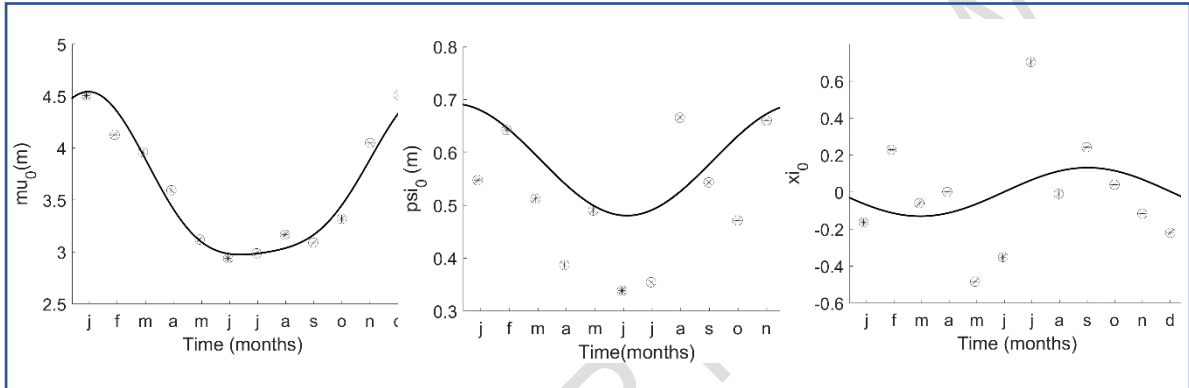


Fig. 4. Estimators for distribution parameters month to month by regression fitting

The location parameter exhibits maximum values in winter season, which indicates, in general terms, larger wave heights in November, December, January and February.

3.1.3. Non-stationary Model Fitting

To estimate the regression coefficients and the probability distribution parameters, the method of maximum likelihood (MML) is employed.

This method is based in the search for point estimators (θ) for the parameters of a X function, so that the probability of observing the sample data $\{x_1, \dots, x_n\}$ is maximum. The likelihood function $L(x; \theta)$ is the joint density function associated to all initial extreme values, from a previously selected distribution and, in general terms, is given by

$$L(x; \theta) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (5)$$

The location $\mu(t) > 0$, scale $\psi(t) > 0$ and shape $\xi(t)$ parameters are expressed in terms of harmonic functions, whose amplitudes are regression parameters to be mathematically estimated [10]

For our study the specific maximum likelihood function is given by

$$l(\theta | t_i, x_i) = -\sum_{i=1}^m \left\{ \log \psi(t_i + \left(1 + \frac{1}{\xi(t_i)}\right)) \log \left[1 + \xi(t_i) \left(\frac{x_i - \mu(t_i)}{\psi(t_i)} \right) \right]_+ + \left[1 + \xi(t_i) \left(\frac{x_i - \mu(t_i)}{\psi(t_i)} \right) \right]_+^{-1/\xi(t_i)} \right\} \quad (6)$$

3.1.4. Automatic Selection, Confidence Intervals and Return Period Values

To find the best model, a selective search methodology named Stepwise is carried out; the Stepwise is in turn based on combining two approaches denoted “Forward Selection” and “Backward Elimination”. Any probabilistic model should ideally incorporate some measure of uncertainty to evaluate the randomness or error from the selected model; for the present study standard errors and confidence intervals are used to deal with the uncertainty. Finally, with regard to the values associated to a return period of interest, the following equations are adopted,

$$X_q(t, \theta) = X_q(\mu(t), \psi(t), \xi(t)) = \begin{cases} \mu(t) - \frac{\psi(t)}{\xi(t)} \left\{ 1 - [-\log(1-q)]^{-\xi(t)} \right\} & \text{if } \xi(t) \neq 0 \\ \mu(t) - \psi(t) \log[-\log(1-q)] & \text{if } \xi(t) = 0 \end{cases} \quad (7)$$

where q is the exceedance probability defined from $G_t(x) = 1 - q$ and the estimated quantile, and $X_q(t, \theta)$ is the time-dependent value associated to the return period $R = 1/q$.

A more detailed description can be found in the study by [6] and the references cited in that paper.

4. DISCUSSION

4.1 Stationary Model

For the stationary model the location, scale and shape parameters, which are invariant in time, are obtained. Table 1 lists the computed values of the distribution parameters, as well as the maximum recorded significant wave height, $H_{s_{\max}}$, and the projected value for H_s associated to a return period of 30 years.

Table 1. Parameters for the stationary model, $H_{s_{\max}}$ and H_s and $Tr=30$ years

$H_{s_{\max}}$ (m)	10.60
H_s (m) for $Tr=30$ yr	5.60
μ	360.34
ψ	58.59

From the values reported in Table 1, it can be inferred that the stationary model is dominated by a Frechet type (heavy tail) distribution behavior. Note that in Table 1, $H_{s_{\max}} = 10.60$ m is an atypical value recorded on 23 July, 1986 during hurricane Estelle (<http://www.prh.noaa.gov/cphc/summaries/1986.php>); it can be clearly observed in Figure 5.

In Figure 5 the instantaneous quantiles for a return period of 30 years for the location and scale parameters are shown, the depiction corresponds to the stationary GEV over the monthly extreme values distributed month to month along the year. In Figure 5, dots indicate the monthly maximum H_s ; the red line represents the time-dependent quantile associated to a 30-year return period; the black, green and yellow lines depict the location, scale and shape parameters, respectively.

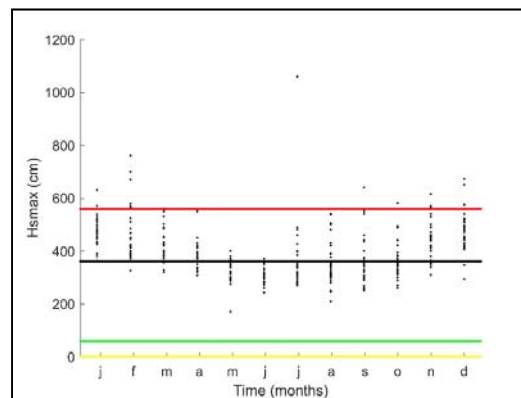
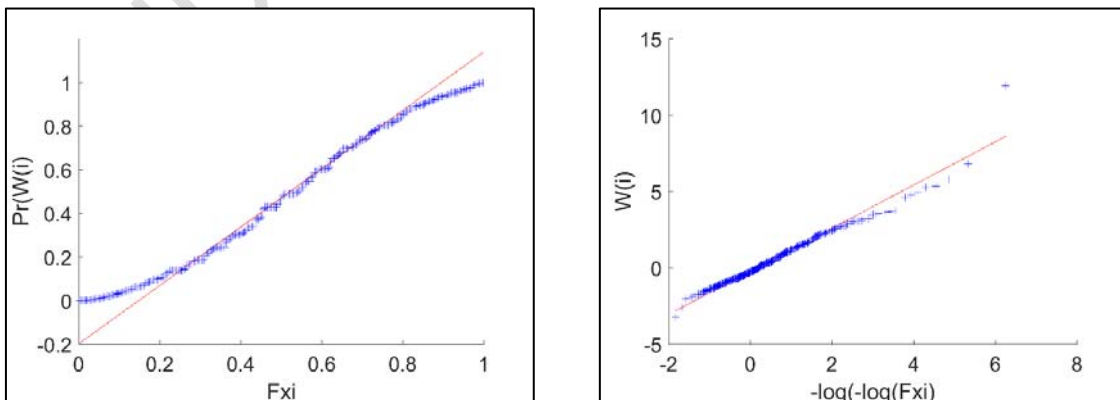


Fig. 5. Estimated quantile values for $Tr=30$ years, monthly maximums and location, scale and shape parameters

In addition to the previous figure, QQ and probability Plots are shown in Figure 6, where it can be observed that for buoy 51004 the probability paper implies an adequate fit for the data; nonetheless, the QQ plot underestimates the quantiles of the extreme values.



(a)

(b)

Fig. 6. (a) probability and (b) quantile plots for the stationary GEV model

4.2 Non Stationary Model

For the non-stationary model, it was found that values for the annual and semiannual cycles for the location and scale parameters are present, as well as the annual cycle for the shape parameter. The zero positive gamma value indicates that data exhibit a heavy-tailed Frechet type distribution, which is expected, if it is considered that the buoy maximum value is related to a hurricane activity, which is quite large as compared to the regular maximum recorded waves. Note that an emphasis on the tail behavior is useful for engineering purposes, and the use of the GEV model is preferred as an adequate and objective criterion for the extremes [11]. Using the computed numerical values in the non-stationary GEV distribution equations, the best fitted parameters for the distribution are obtained. In Figure 7 the red line shows the time-dependent quantile associated to a 30-year return period; the black line depicts the location parameter, the green line corresponds to the scale parameter and the yellow line corresponds to the shape parameter. The monthly maxima for H_s are shown by the black dots. An adequate fit for the studied case can be observed in Figure 7. Note that the maximum H_s associated to hurricane Estelle (mentioned previously), does not significantly deviate the trend of the values associated to a 30 year return period. Nevertheless, it may lead to a smoother curve as compared with the trend for buoys in the Gulf of Mexico [6], where two peaks associated to cold fronts are hurricanes are shown; Figure 7 also seems to show that kind of behavior, if the black dots (monthly maxima) are observed for February and September, but those peaks may be somehow attenuated due to the atypical maximum H_s in July. Nevertheless, it is noteworthy that a seasonality effect does exist also in the Pacific Coast and, although the results in the present study are limited to one site only, it could be preliminarily concluded that seasonal trends are existent in both, the Gulf of Mexico and the Pacific Ocean. Consequently, further studies to compare the trends for more buoys in the Pacific are strongly recommended.

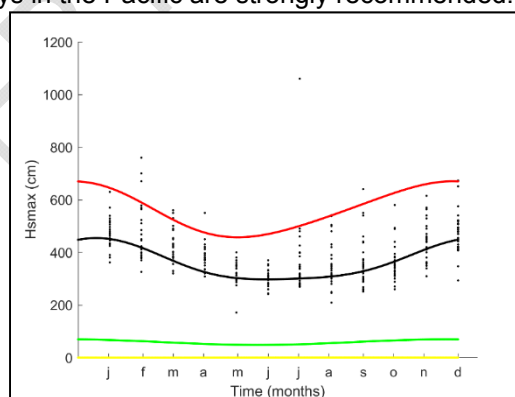


Fig. 7. Fitting for the monthly maximum for the non-stationary GEV model

To represent the adequacy of the fitting, the empirical probabilities obtained with the model and with the quantiles are shown in Figure 8; the probability associated to the Gumbel distribution from the $W(t)$ statistic is shown.

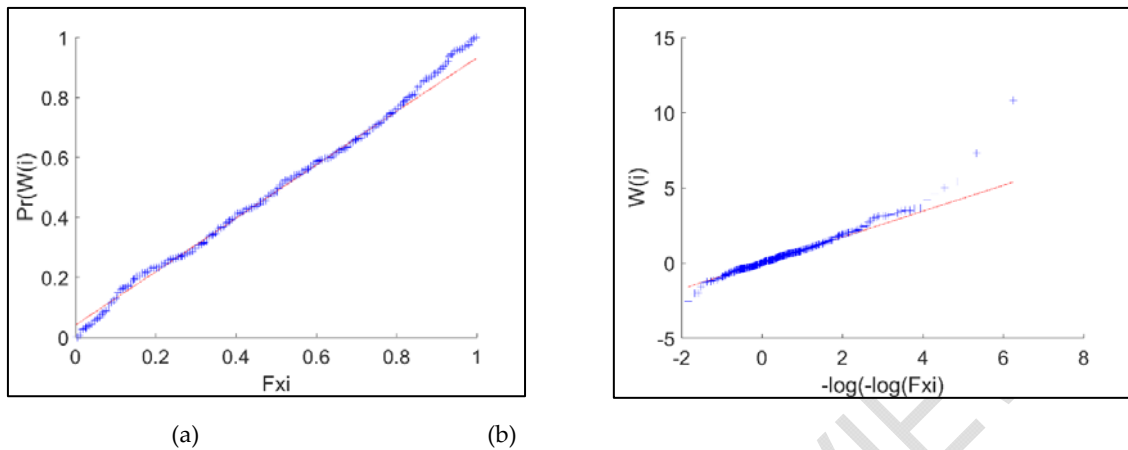


Fig. 8. Diagnostic plots (a and b) for the non-stationary GEV model

4.3 Comparison of Stationary versus Non-Stationary GEV Models

In this section a comparison of the two employed models for the extreme value analysis is carried out. The basic stationary GEV model and the non-stationary GEV model, which, unlike the former, captures seasonal variation for monthly maxima data series, are contrasted.

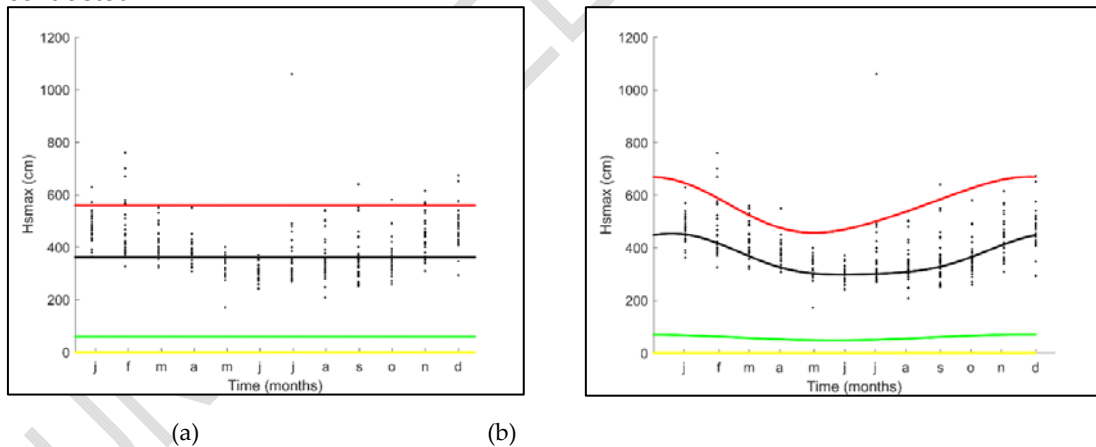


Fig. 9. Quantiles associated to 30-year return period for (a) the stationary and (b) non-stationary GEV models.

In Figure 9 the most adequate fitting of the non-stationary GEV model to the data can be clearly observed, since the curve follows the seasonal trends; conversely, it can be seen that the stationary GEV model is unable to capture the seasonality of the significant wave heights. Figure 10 shows also the best fitted results month-to-month, but for a 100-year return period. In Figure 10 the depicted quantile for a 100-year return period shows that the non-stationary model includes the extreme values and represents more adequately the

seasonality of the used data (red line), as compared to the stationary model (green line), except by the extreme value in July (see black dots for monthly Hs).

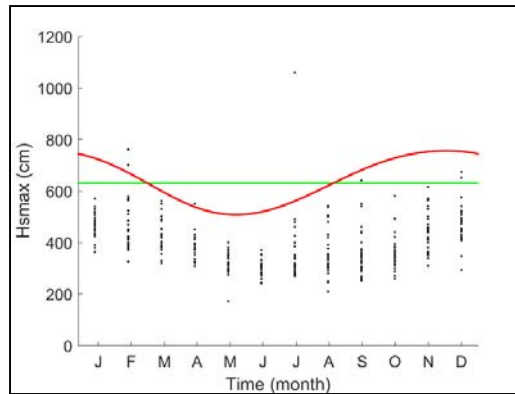


Fig. 10. Quantiles comparison for 100-year return period and measured Hs

Figure 11 shows 95% confidence intervals, where it can be observed that the bounds are more reduced for the non-stationary model and better correlated to the month-to-month data.

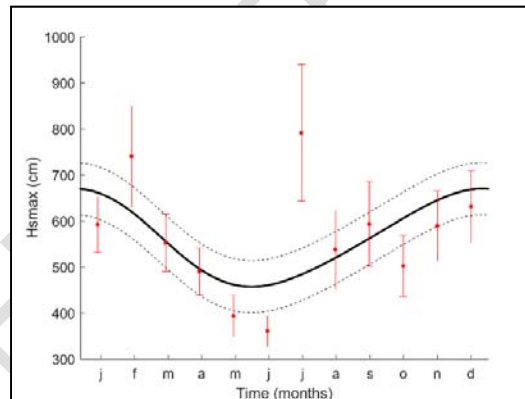


Fig. 11. 30-year return period quantile and 95% confidence interval for the non-stationary GEV model and the month-to-month stationary GEV.

Results in Figures 10 and 11 show the advantage of using the non-stationary GEV model, since the computed seasonal trend tend to capture the maximum values and fits better the data series; the expected wave heights for $T_R=100$ years computed for the non-stationary case are larger than those computed with the stationary one for the winter season, and the opposite occurs for the summer season; this is important, since an impact in design, management, reliability related tasks and sustainability related issues is expected. The reason of the smaller values with the stationary GEV model for some months, is that it does not effectively account for scarce and extreme wave heights associated to hurricanes. In contrast, the non-stationary GEV model can capture these extreme values (also included in the integrated quantile) by means of the sinusoidal trend.

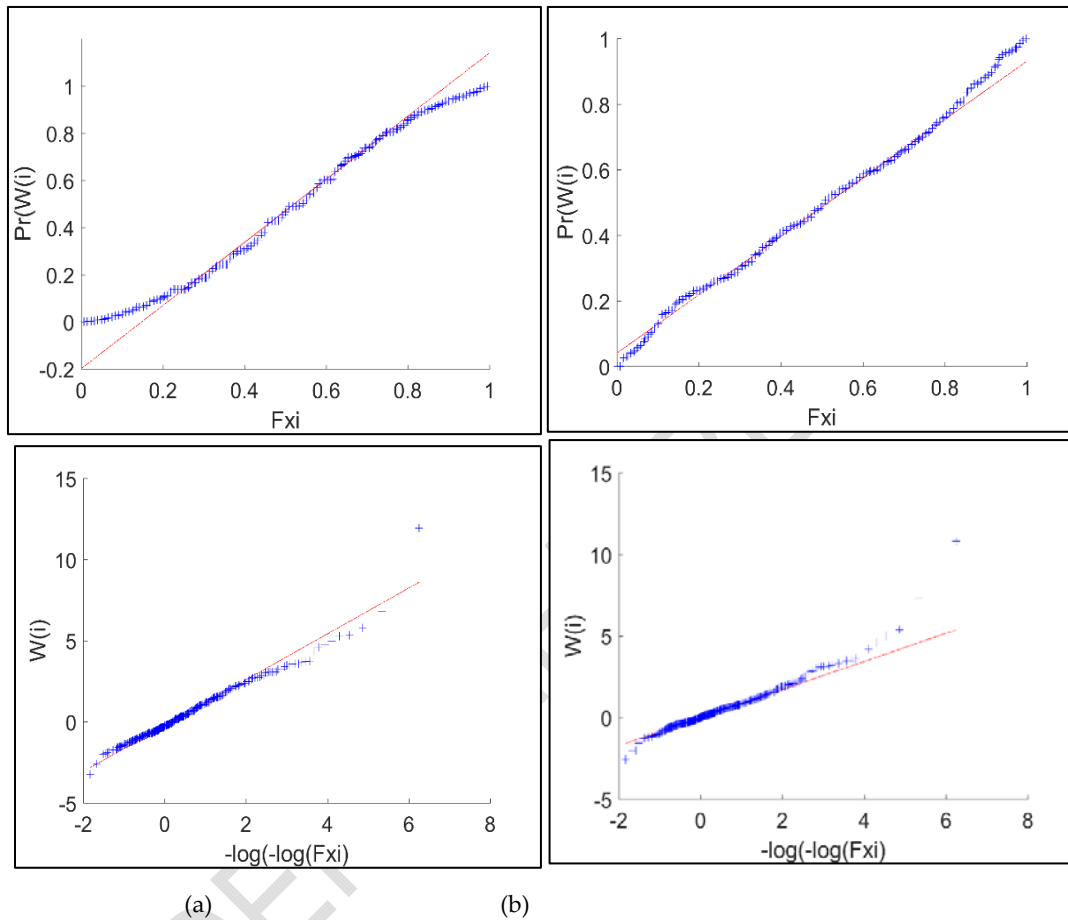


Fig. 12. Goodness-of-fit plots for stationary (a) and non-stationary (b) models

A final Figure in this study, Figure 12, shows that the best behavior (in terms of representing the selected meteo-oceanographic variable) for the used empirical data is obtained with the non-stationary GEV distribution, at least for selected buoy, as observed by the goodness-of-fit for both models in Figure 12 (left side column for the stationary model and right side column for the non-stationary model, respectively). It is pointed out once more, that this is due to the fact that the non-stationary model try to capture the majority of extreme values through a sinusoidal computing scheme; also, the existent extreme wave values caused by hurricanes, which significantly alter the seasonal periodic variation, are more adequately incorporated by the non-stationary GEV model in general terms; moreover, the uncertainty induced by hurricane activity, is compensated by the non-stationary GEV model for the monthly data series.

5. CONCLUSIONS

In this study a methodology for fitting empirical meteo-oceanographic data to extreme value models for projections purposes is presented. The selected meteo-oceanographic variable is significant wave height. The GEV model is used from two perspectives: stationarity and non-stationarity. A comparison is carried out between the two approaches for a single site. For the non-stationary GEV model the seasonality is model by means of sinusoidal curves for a single buoy in the Pacific Ocean.

The non-stationary GEV model predicts adequately the monthly maximum significant wave heights and reduce the uncertainty of the estimated quantiles; it also projects in a better way the wave heights associated to given return periods, since the methodology includes not only the point maxima, but also the peripheric values along the year. From the comparison of the stationary and non-stationary GEV models, it is concluded that the later leads to a better representation of the meteo-oceanographic variable, and also leads to larger projected return period values than the former for winter season. This is important, since implications in design of maritime structures, managements tasks, reliability and risk assessments, as well as sustainability related issues, are expected depending on the selected model.

It is also concluded that, at least for the selected buoy and its recorded data, a seasonality trend is found, as it was the case for the Gulf of Mexico [6], although some differences are found in the trends. Future studies for more buoys and wider regions are desirable, to characterized extreme wave activity, aimed at improving the design and management of coasts and ports by including seasonality. It is suggested to investigate the impact of selecting different models (i.e., stationary versus non-stationary models for extreme waves) in the sustainability of coasts and coast infrastructure. Further research for other meteo-oceanographic variables is also recommended.

REFERENCES

1. Shao, Z.; Liang, B.; Li, H.; Lee, D. Study of sampling methods for assessment of extreme significant wave heights in the South China Sea. *Ocean Eng* **2018**, *168*, 173-184. DOI: [10.1016/j.oceaneng.2018.09.015](https://doi.org/10.1016/j.oceaneng.2018.09.015).
2. Vanem, E. Non-stationary extreme value models to account for trends and shifts in the extreme wave climate due to climate change. *Appl Ocean Res* **2015**, *52*, 201-211. DOI: [10.1016/j.apor.2015.06.010](https://doi.org/10.1016/j.apor.2015.06.010).
3. Feld, G.; Randell, D.; Wu, Y.; Ewans, K.; Jonathan, P. Estimation of storm peak and intrastorm directional-seasonal design conditions in the North Sea. *Journal of Offshore Mechanics and Arctic Engineering* 2015. 137(2), 021102, DOI: [10.1115/1.4029639](https://doi.org/10.1115/1.4029639).
4. Shao, Z.; Liang, B.; Li, H.; Wu, G.; Wu, Z. Blended wind fields for wave modeling of tropical cyclones in the South China Sea and East China Sea. *Appl Ocean Res* **2018**, *71*, 20-33. DOI: [10.1016/j.apor.2017.11.012](https://doi.org/10.1016/j.apor.2017.11.012).
5. Johathan, P.; Ewans, K. Statistical modelling of extreme ocean environments for marine design: A review. *Ocean Eng* **2013**, *91-109* DOI: [10.1016/j.oceaneng.2013.01.004](https://doi.org/10.1016/j.oceaneng.2013.01.004)
6. Calderón-Vega, F; Vázquez-Hernández, A.O.; García-Soto, A.D. Analysis of extreme waves with seasonal variation in the Gulf of Mexico using a time-dependent GEV model. *Ocean Eng* **2013**, *73*, 68-82. DOI: [10.1016/j.oceaneng.2013.08.007](https://doi.org/10.1016/j.oceaneng.2013.08.007).

7. Sanchez-Arcilla A.; Sierra, J.P.; Gracia, V.; Garcia, M.; Jimenez, J.A.; Mösso, C.; Valdemoro, H. Coastal sustainability for uncertain futures: a Spanish Mediterranean case from the RISES-AM project. A: International Association for Hydraulic Research Congress. E-proceedings of the 36th IAHR World Congress. L'Haia, **2015** pp. 1-8.
8. Sanchez-arcilla, A.; Sierra, J.P.; Mösso, C.; Gracia, V.; Garcia, M. Progress on climate change detection and projections over Spain since the findings of the IPCC AR5. International Symposium CLIMATE-ES **2015**; pp. 146.
9. Coles, S.G. *An Introduction to Statistical Modelling of Extreme Values*, 1st ed.; Springer-Verlag: London, UK, **2001**; pp. 1-208.
10. Menéndez, M.; Méndez, F.J.; Izaguirre, C.; Luceño, A.; Losada, I.J. The influence of seasonality on estimating return values of significant wave height. *Ocean Eng* **2009** 56, pp. 211-219. DOI: 10.1016/j.coastaleng.2008.07.004.
11. Panchang, V.; Kwon Jeong Ch.; Demirebilek, Z. Analyses of extreme wave heights in the Gulf of Mexico for offshore engineering applications. *J Offshore Mech Arct* **2013**, 135(3), pp. 1-15.