Statistical assessment of annual patterns in coastal extreme wave conditions

J. L. Vega[1], J. González[2], G. Rodríguez[1]
[1]Departamento de Física, Universidad de Las Palmas de Gran Canaria, Spain
[2]Plataforma Oceánica de Canarias, Spain

Abstract

The annual cycle in extreme values of significant wave height is examined by transforming the timing of the storm peaks in a circular variable, with the aim of taking advantage of the many tests devised to explore uniformity on the circle. The use of four different but complementary uniformity tests makes possible a robust assessment of the annual cycle statistical significance. Seasonality of storms in a long time series of significant wave heights, measured in a coastal zone, is examined. The presence of a seasonal pattern is statistically beyond doubt.
Keywords: wave storms, seasonality, circular statistics, circular uniformity.

1 Introduction

Many geophysical processes may exhibit periodic cycles along time, such as diurnal, weekly, monthly, and annual variations. In particular, geophysical variables may exhibit marked periodic behaviour on annual time scales, mainly induced by the solar cycle and commonly known as seasonal variations.

Information on the temporal behaviour of severe wave conditions, and more specifically on the mean annual pattern of these extreme conditions, is required in many applications such as, for example, coastal zone management, coastal morphodynamics, and coastal engineering infrastructure operations.

Furthermore, confidence on the existence of a seasonal periodicity in long-term time series of significant wave height, H_S, the most common parameter used to characterize the sea state severity, is very important when using time series models to represent the stochastic evolution of such parameter (e.g.,

Athanassoulis and Stefanakos [1]; Guedes Soares *et al.* [2]). Additionally, statistical methods of extreme value analysis used for the derivation of design events, by extrapolating the data outside the range of observations, commonly assume that data are independent and identically distributed. The assumption of independence is generally well satisfied by using an adequate data selection procedure which ensures the independence of representative values of successive storms. Nevertheless, the presence of seasonal variability in the selected extreme values violates the assumption of data identically distributed.

It seems natural that if processes underlying wave climate show a seasonal variation their extremes do as well. However, such cycles in extremes have not received much attention, as the conventional method of extreme wave height analysis, the commonly named annual maxima technique, does not require their explicit modelling. The increasing adoption of the peaks over threshold approach, which demands the inclusion of annual cycles, has stimulated the interest in considering the seasonal effects in the study of extreme wave events. Thus, many studies (e.g., Morton *et al.* [3]; Méndez *et al.* [4]; Jonathan and Ewans [5]; Mackay *et al.* [6]) have evidenced the importance of accounting for seasonality, as a covariate when estimating extreme wave conditions by means of the peaks over threshold method, or removing its effects, by using some appropriated approach, such as analysing the data in discrete seasons (e.g. Morton *et al.* [3]), to allow the assumption of data identically distributed to be met.

The identification of the annual cycle in a long-term series of H_S is relatively easy, such as is observed in Figure 1, where the average year constructed from a time series of H_S three-hourly sampled is shown. Contrarily, the presence of this cycle in extreme values is commonly doubtful and not easily assessable.

This paper aims to evaluate in a consistent way the statistical significance of possible seasonal variations present in series of extreme values derived by considering storms separated by several days and declustering values corresponding to the same storm. For this, the time of occurrence of individual values of H_S representing each storm is considered as a circular variable, so that it can be properly described in terms of directional statistical tools.

The paper is structured as follows. Principal characteristics of the significant wave height time series used in the study, as well as the statistical methodology used to assess the existence of seasonal variations in the timing of wave storms along the year are introduced in section 2. The results of applying the statistical approaches are discussed in section 3. Conclusions are summarised in section 4.

2 Data and methodology

2.1 Data set

The experimental data set used in the study is a time series of three-hourly values of significant wave heights derived from wave measurements made in the north coast of Gran Canaria island, at a place of coordinates (28° 8.5′N, 15° 27.5′W) and a water depth of about 40 m. Measurements cover a period of fifteen years,

from 1987 to 2001. Each year includes 365 days. The records of 29 February are omitted in leap years. The mean year obtained by averaging the fifteen values measured at the same time at each of the fifteen years is shown in Figure 1 and reveals, on average, a more or less clear annual cycle.

The time of each storm event has been extracted from the H_S series by selecting the value and timing associated to the peak sea state during a storm, defined as a period of relative severe wave conditions that satisfy the requirement of independence. The approach used to extract the peaks representing the storms was suggested by Simiu and Heckert [7] and considers that peaks of different storms are separated at least by 48 hours. This methodology has been applied by González *et al.* [8] to characterize the intensity of storms in the same series by using the peaks over threshold approach.
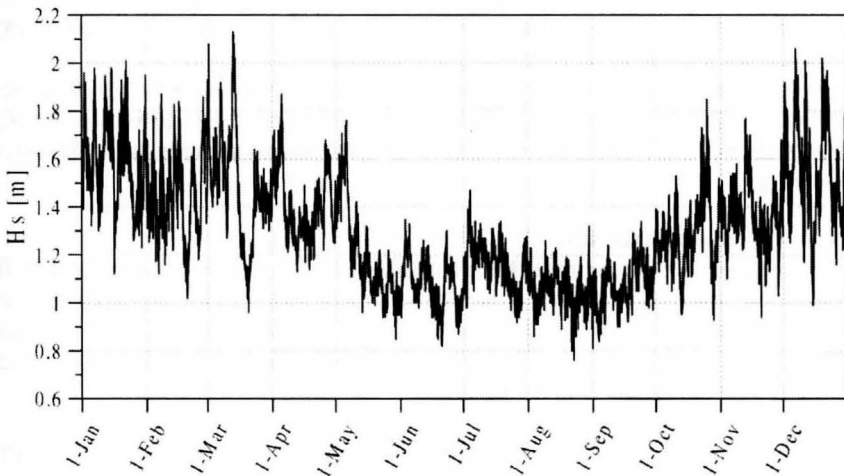

Figure 1:   Three-hourly mean value of the H_S time series recorded in the period (1987–2001).

2.2 Statistical assessment of seasonality

In many practical situations the examined variable is a direction in a two-dimensional plane. In such a case, the sample space consists of points on a unit circle and conventional statistics is not applicable. Circular statistics is the special branch of statistics developed for the proper analysis of this kind of random variable, in which probability distributions are characterised by their cyclic nature. That is, circular statistics includes methods to study random variables that have a cyclic behaviour (Mardia and Jupp [9]). In this context, it is interesting to note that any circular temporal measure can be translated into angles. Thus, for example, it is possible to consider the day of the year at which a storm occurs as a circular random variable. For this, the day of the year, d, must be converted to a angular value, θ, in radians, by

$$\theta = \frac{2\pi}{365} d \tag{1}$$

A very common question in circular statistics is whether a data sample is, or not, distributed uniformly around the circle. This means that the uniform distribution is usually considered as the null hypothesis. Then, to know if wave storms at a given coastal region are uniformly distributed through the year (null hypothesis), or there is one or several time periods during which storms are more frequent (alternative hypothesis), is necessary to know if the time of occurrence of wave storms along the year follows a uniform distribution

$$f(\theta) = \frac{1}{2\pi} \qquad 0 \le \theta \le 2\pi \tag{2}$$

There are multiple tests to answer this question, which differ in their efficiency to detect certain departures from uniformity. Note that uniformity, also referred as randomness or isotropy, represents the situation in which probability is spread out uniformly on the circumference of a circle.

Four commonly used tests to assess the uniformity of circular variables are applied in the present study. The fundamentals of these tests are briefly introduced bellow. A detailed description of these and other tests can be found in Fisher [10].

2.2.1 The chi-squared test
The chi-square test is the most frequently used test to assess the goodness of fit of the empirical distribution of a data set to a theoretical model. To apply the chi-square test the circle is divided in twelve sectors, $k=12$, of 30° each one, and the observed, o_i, and expected, e_i, frequencies of storms in the i-th month computed. The Chi-square statistic is given by

$$T = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \tag{3}$$

Null hypothesis (uniformity) is rejected if T exceeds the corresponding critical value for k-1 degrees of freedom. Critical values for several confidence levels, α, are given in Fisher [10]. It is worth of mentioning that this test is robust for unimodal and multimodal samples.

2.2.2 The Rayleigh test
The Rayleigh test is based in the estimated value of the resultant vector length R, given by

$$R = \sqrt{C^2 + S^2} \tag{4}$$

where

$$C = \frac{1}{N} \sum_{i=1}^{k} n_i \cos \theta_i \tag{5}$$

and

$$S = \frac{1}{N} \sum_{i=1}^{k} n_i \sin \theta_i \tag{6}$$

and θ_i is the date at the centre of each monthly bin i, expressed in radians. The confidence level P associated with the mean resultant length R is given by

$$P = e^{-Z}\left[1 + \frac{2z - z^2}{4n} - \frac{24z - 132z^2 + 76z^3 - 9z^4}{288n^2}\right] \qquad (7)$$

The Rayleigh test is considered a powerful test only if is possible to assume that the population distribution does not have more than one mode. Furthermore, it is important to remark that, it assumes sampling from a von Mises distribution.

Note that a value of $R=1$ indicates that all storms occur on the same calendar day in all years. However, a value of $R=0$ that the probability of occurrence of storms is the same for any calendar day. The latter is exactly true only if the distribution does not have more than one mode.

2.2.3 Kuiper test

This test is an alternative to de chi-square one based on the cumulative distribution function. The basic idea of this test is that the observed and the theoretical distributions should closely resemble one another if the sample has been drawn from the assumed circular uniform distribution. The main step in the approach to assess uniformity is to compute the deviations between the uniform and empirical cumulative distributions. The following statistics are defined

$$\begin{aligned} D^+ &= max\{F_n(\theta) - F(\theta)\} \\ D^- &= max\{F(\theta) - F_n(\theta)\} \end{aligned} \qquad (8)$$

where $F_n(\theta)$ and $F(\theta)$ are the sample and the uniform cumulative distributions. The sum of D^+ and D^- values define the Kuiper test statistic

$$V_n = D_n^+ + D_n^- \qquad (9)$$

or even best, V

$$V = V_n\left(n^{1/2} + 0.155 + \frac{0.24}{n^{1/2}}\right) \qquad (10)$$

The uniformity hypothesis is rejected if the test statistic, V, exceeds the critical values tabulated in Fisher [10]. An important aspect to remark is that the Kuiper test is specially indicated in case of multimodal distributions.

2.2.4 The modified Kolmogorov-Smirnov test

Freedman [11] suggested a modification of the classical Kolmogorov-Smirnov test to examine seasonality in data. This non-parametric methodology removes some drawbacks existing with the conventional one. The hypothetical (uniform) cumulative distribution is a step function denoted by $F(t)=t/12$, where t is the rank of each month of the year. The sample cumulative distribution is also a step function denoted by $FN(t)=k/N$, where k is the number of events that have occurred during all months $\leq t$. The test statistic, T, is given by

$$V_N = max(F_N(t) - F(t)) + |min(F_N(t) - F(t))| \; ; \; 1 \leq t \leq 12 \qquad (11)$$

The distribution of T does not follow any specified distribution, but has been empirically evaluated by means of Monte Carlo simulations, and is tabulated in Freedman [11].

3 Results and discussion

Representation of circular data in polar coordinates is commonly useful. This requires the specification of the angle, θ, and a distance, r, and provides a method of uniquely defining the location of data points in the circle. This representation method has the advantage of clearly separating directional and distance (intensity) information. Polar plots of the storm events for six different thresholds of significant wave height are represented in Figure 2. The threshold varies from 2 meters until 4.5 meters with increments of 0.5 meters, from the left upper corner to right and down. The number of storms considered for each threshold is indicated in the second column of Table 1.

It can be observed that the non-uniformity of the time of occurrence of wave storms becomes more and more clear as the wave height threshold imposed to define the extreme events is increased. Thus, while for a threshold of 2 meters storms occur in any period of the year, for a threshold of 4 meters the stormy period reduces to autumn and winter.

Tables 1–4 include the results of the Chi-square, Rayleigh, Kuiper, and modified Kolmogorov-Smirnov test of uniformity. Each table includes a first column indicating the wave height threshold, a column with the corresponding test statistic value and the critical value for a confidence level of α=0.01. The last column indicates the acceptance or rejection of the null hypothesis of uniformity. All the tests used reject the uniformity, accept the seasonality, of the wave storm occurrence for any threshold.

It should be noted that the tests used to assess randomness are mutually complementary. Thus, while Chi-square test is robust for unimodal and multimodal samples, Rayleigh test is considered a powerful test if the population distribution is unimodal, Kuiper and modified Kolmogorov-Smirnov tests are specially indicated in case of multimodal distributions. Then the rejection of uniformity by all the tests demonstrate statistically significant trend of wave storms to cluster during a given period of the year. In other words, the use of various types of tests, parametric and non-parametric, adequate for unimodal and for multimodal distributions, evidences without doubts the existence of a cyclic annual pattern in the timing of wave storms for the studied zone.

4 Conclusions

Results derived from the application of four tests of uniformity clearly reveal the presence of an annual cycle in the time of occurrence of wave storms in the coastal area examined. This seasonal pattern is independent of the significant wave height threshold considered.
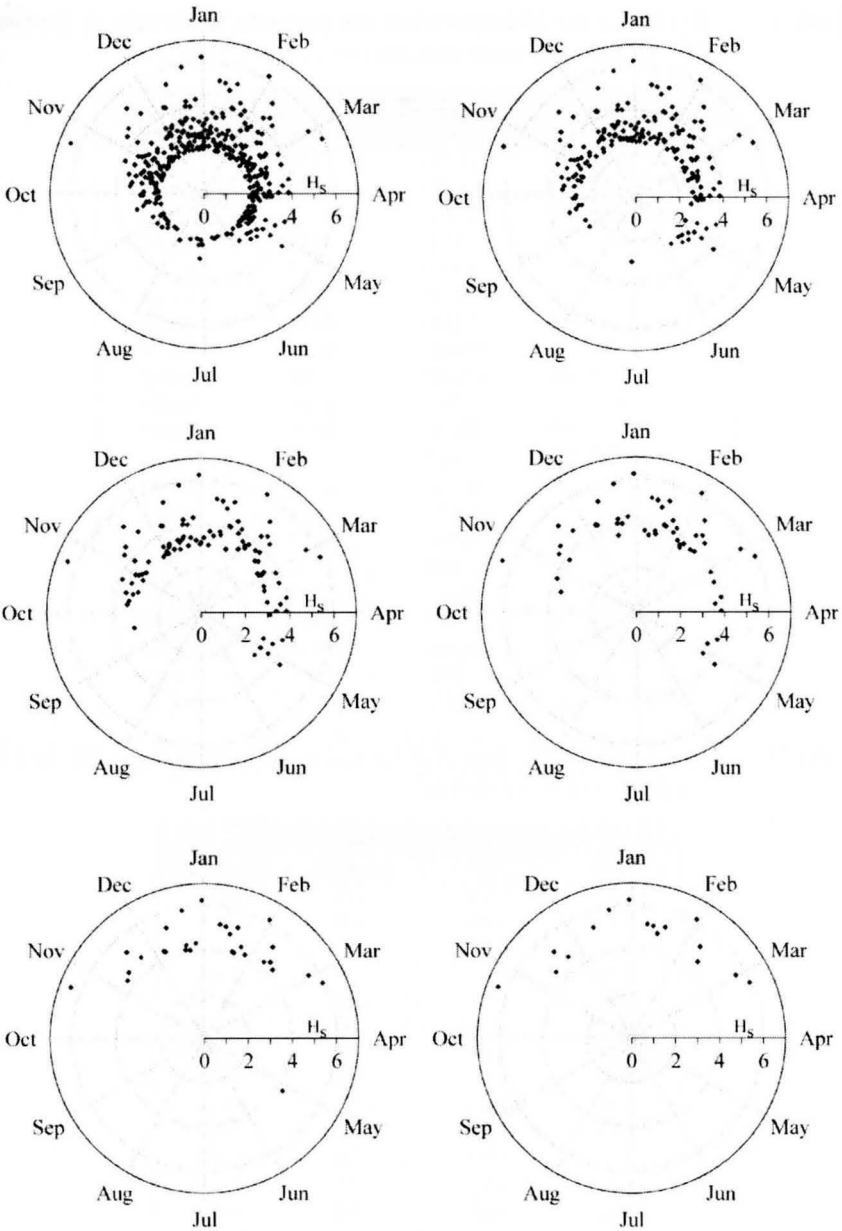
Figure 2: Polar plots of the time of occurrence in the year and the severity of wave storms for different threshold values. Upper left 2.0 m, upper right 2.5 m, middle left 3.0 m, middle right 3.5 m, lower left 4.0 m, and lower right 4.5 m.

Table 1:   Results of the Chi-square test for assessing uniformity in the time of occurrence of wave storms.

Chi-square test				
Hs (m)	N	T	Tc (α = 0,01)	Uniformity
> 2,0	355	117,6648	24,725	Rejected
> 2,1	309	133,0583	24,725	Rejected
> 2,2	270	138,2667	24,725	Rejected
> 2,3	238	130,6723	24,725	Rejected
> 2,4	210	125,6571	24,725	Rejected
> 2,5	186	120,7097	24,725	Rejected
> 2,6	168	116,7143	24,725	Rejected
> 2,7	153	112,4902	24,725	Rejected
> 2,8	132	100,9091	24,725	Rejected
> 2,9	116	99,3793	24,725	Rejected
> 3,0	104	96,0769	24,725	Rejected
> 3,1	85	91,0471	24,725	Rejected
> 3,2	77	88,3506	24,725	Rejected
> 3,3	74	84,2703	24,725	Rejected
> 3,4	64	75,8750	24,725	Rejected
> 3,5	55	64,3455	24,725	Rejected
> 3,6	47	62,0213	24,725	Rejected
> 3,7	45	56,6000	24,725	Rejected
> 3,8	38	45,3684	24,725	Rejected
> 3,9	31	56,0968	24,725	Rejected
> 4,0	30	52,4000	24,725	Rejected

Table 2:   Results of the Rayleigh test for assessing uniformity in the time of occurrence of wave storms.

Rayleigh Test			
Hs (m)	$\bar{R}$	Rc (α=0.01)	Uniformity
> 2,0	0,3989	0,4035	Rejected
> 2,1	0,4566	0,4618	Rejected
> 2,2	0,4973	0,5030	Rejected
> 2,3	0,5118	0,5177	Rejected
> 2,4	0,5307	0,5368	Rejected
> 2,5	0,5509	0,5573	Rejected
> 2,6	0,5646	0,5711	Rejected
> 2,7	0,5811	0,5878	Rejected
> 2,8	0,5928	0,5997	Rejected
> 2,9	0,6220	0,6291	Rejected
> 3,0	0,6431	0,6505	Rejected
> 3,1	0,6723	0,6800	Rejected
> 3,2	0,6983	0,7063	Rejected
> 3,3	0,6959	0,7039	Rejected
> 3,4	0,7073	0,7154	Rejected
> 3,5	0,7025	0,7106	Rejected
> 3,6	0,7411	0,7496	Rejected
> 3,7	0,7324	0,7408	Rejected
> 3,8	0,7133	0,7215	Rejected
> 3,9	0,7930	0,8022	Rejected
> 4,0	0,7868	0,7958	Rejected

Table 3:   Results of the Kuiper test for assessing uniformity in the time of occurrence of wave storms.

Kuiper Test			
HS (m)	V	Vc (α = 0,01)	Uniformity
> 2.0	5,2461	2.001	Rejected
> 2.1	5,5660	2.001	Rejected
> 2.2	5,6769	2.001	Rejected
> 2.3	5,5663	2.001	Rejected
> 2.4	5,4874	2.001	Rejected
> 2.5	5,4100	2.001	Rejected
> 2.6	5,2613	2.001	Rejected
> 2.7	5,1094	2.001	Rejected
> 2.8	4,7861	2.001	Rejected
> 2.9	4,7162	2.001	Rejected
> 3.0	5,5727	2.001	Rejected
> 3.1	4,3132	2.001	Rejected
> 3.2	4,2666	2.001	Rejected
> 3.3	4,1337	2.001	Rejected
> 3.4	4,0448	2.001	Rejected
> 3.5	5,6588	2.001	Rejected
> 3.6	3,7719	2.001	Rejected
> 3.7	3,6585	2.001	Rejected
> 3.8	3,3373	2.001	Rejected
> 3.9	3,5140	2.001	Rejected
> 4.0	3,4471	2.001	Rejected

Table 4:   Results of the modified Kolmogorov-Smirnov test for assessing uniformity in the time of occurrence of wave storms.

Modified Kolmogorov-Smirnov Test					
HS(m)	Max	Min	Vn	Vn*(n)^0.5	Uniformity
> 2.0	0,1455	0,1077	0,2533	4,7723	Rejected
> 2.1	0,1845	0,1092	0,2937	5,1626	Rejected
> 2.2	0,2000	0,1167	0,3167	5,2034	Rejected
> 2.3	0,2080	0,1155	0,3235	4,9912	Rejected
> 2.4	0,2071	0,1357	0,3429	4,9685	Rejected
> 2.5	0,2070	0,1478	0,3548	4,8394	Rejected
> 2.6	0,2024	0,1607	0,3631	4,7063	Rejected
> 2.7	0,2141	0,1552	0,3693	4,5678	Rejected
> 2.8	0,2348	0,1439	0,3788	4,3519	Rejected
> 2.9	0,2414	0,1638	0,4052	4,3638	Rejected
> 3.0	0,2308	0,1827	0,4135	4,2165	Rejected
> 3.1	0,2451	0,1863	0,4314	3,9771	Rejected
> 3.2	0,2825	0,1710	0,4535	3,9791	Rejected
> 3.3	0,2905	0,1577	0,4482	3,8555	Rejected
> 3.4	0,2969	0,1615	0,4583	3,6667	Rejected
> 3.5	0,3500	0,1061	0,4561	3,3822	Rejected
> 3.6	0,3670	0,1312	0,4982	3,4157	Rejected
> 3.7	0,3500	0,1444	0,4944	3,3168	Rejected
> 3.8	0,2763	0,2018	0,4781	2,9470	Rejected
> 3.9	0,3172	0,2527	0,5699	3,1730	Rejected
> 4.0	0,3000	0,2667	0,5667	3,1038	Rejected

The circular statistical tools results appropriate to reveal cyclic patterns in the extreme wave event sequences, allowing to accept or reject on a robust statistical basis if wave storms at a given coast are uniformly distributed through the year, or there is one or several time periods during which storms are more frequent. Complementarity of the statistical tests enhances the statistical significance of the results.

Acknowledgements

References

[1]   Athanassoulis, G. A. and Ch. N. Stefanakos, A non-stationary stochastic model for long-term time series of significant wave height, J. Geophysical Research. 100 (C8), pp. 16149-16162, 1995.

[2]   Guedes Soares, C., Ferreira, A.M., Cunha, C., Linear models of the time series of significant wave height in the Southwest Coast of Portugal, Coastal Engineering, 29, pp. 149–167, 1996.

[3]   Morton, I.D., Bowers, J., Mould, G., Estimating return period wave heights and wind speeds using a seasonal point process model. Coastal Engineering, 31, 305–326, 1997.

[4]   Méndez, F.J., Menéndez, M., Luceño, A., Losada, I.J. Estimation of the long-term variability of extreme significant wave height using a time-dependent POT model, J. Geophysical Research. 111, C07024. doi:10.1029/2005JC003344., 2006.

[5]   Jonathan, P., Ewans, K.C., Modelling the seasonality of extreme waves in the Gulf of Mexico. J Offshore Mechanics and Arctic Engineering, 133, 021104-1-9, 2011.

[6]   Mackay, B.L., Challenor, P.G., Bahaj, A.S., On the use of discrete seasonal and directional models for the estimation of extreme wave conditions, Ocean Engineering 37(5–6), 425–442, 2010.

[7]   Simiu, E. and N.A. Heckert, Extreme wind distribution tails: A peaks over threshold approach, J. Structural Engineering, ASCE, 122(5), 195-210, 1996.

[8]   González, J, J. Vega, and G. Rodríguez, Extreme value analysis of sea waves in Canary islands by means of the POT method, Proc. 30th Bienal de Física RSEF, 388-389, 2005.

[9]   Mardia, K. V., and P. E. Jupp, Directional Statistics, John Wiley, Hoboken, N.J., 2000.

[10] Fisher N.I., Statistical Analysis of Circular Data. Cambridge University Press. 1993.

[11] Freedman, I.S., The use of a Kolmogorov-Smirnov type statistic in testing hypotheses about seasonal variation. J. Epidemiology and Community Health, 33, 223-228., 1979.